Ghent University<br>Faculty of Bioscience Engineering<br>Department of Data Analysis and Mathematical Modelling

# Combining absolute and relative evaluations for determining sensory food quality: analysis and prediction 

Marc Sader

Thesis submitted in fulfilment of the requirements for the degree of Doctor (Ph.D.) of Applied Biological Sciences

Supervisor: $\quad$| Prof. dr. Bernard De Baets |
| :--- |
| Department of Data Analysis and Mathematical Modelling, |
|  |
| Ghent University, Belgium |

Examination committee: Prof. dr. ir. Frank Devlieghere (Chairman)

Prof. dr. ir. Wim Verbeke
Prof. dr. José Luis García-Lapresta
dr. ir. Mike Vanderroost
dr. ir. Michiel Stock

Dean:
Prof. dr. ir. Marc Van Meirvenne

Rector:
Prof. dr. ir. Rik Van de Walle

Marc Sader

# Combining Absolute and relative EVALUATIONS FOR DETERMINING SENSORY FOOD QUALITY: ANALYSIS AND PREDICTION 

Thesis submitted in fulfilment of the requirements for the degree of
Doctor (Ph.D) of Applied Biological Sciences

Dutch translation of the title: Combinatie van absolute en relatieve evaluaties voor het bepalen van sensorische voedselkwaliteit: analyse en voorspelling

Please refer to this work as follows:
Marc Sader (2019). Combining absolute and relative evaluations for determining sensory food quality: analysis and prediction, PhD Thesis, Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium.

This research has been supported by the project IWT-SBO-130036: CHECKPACK - Integrated optical sensors in food packaging to simultaneously detect early-spoilage and check package integrity.

ISBN 978-94-6357-229-3

The author and the supervisor give the authorization to consult and to copy parts of this work for personal use only. Every other use is subject to the copyright laws. Permission to reproduce any material contained in this work should be obtained from the author.

## Acknowledgements


#### Abstract

"In the sweetness of friendship let there be laughter and sharing of pleasures. For in the dew of little things, does the heart find its morning and is refreshed."


Kahlil Gibran
Along the way, I've had many wonderful influences that allowed me to complete this dissertation. Therefore, it is a humbling experience for me to acknowledge those people who have influenced and helped.

This journey would not have been possible if it weren't for Bernard, who gave me not just the opportunity to start my PhD in his department, but also for the chance to carry out this dissertation. I convey my sincere expression of thanks for the support and contribution in reviewing my writing and providing corrections to improve and finalize the research papers as well as this dissertation.

I am grateful to my Doctoral jury members for their careful reading of this dissertation and for their constructive comments and suggestions.

I would also like to thank everybody who was involved in CheckPack, in particular, Jan, for suggesting the predictive modelling idea and helping me develop it, and Lotta and Angelos, for giving me all the spoiled food samples to smell and for providing me with the necessary data.

There have been many friends who provided a stimulating and fun-filled environment in (and out of) the department and to whom I would like to show my gratitude. Firstly, to my office friends, Hilde, Michaël, Michiel, Gang and Peter, who always brightened my day. Secondly, to all my department friends with whom I did not only share cafeteria meals.

Some friends have shared unforgettable memories with me and have become close and dear to me. They hold a special place in my heart, and I owe them for being there and picking me up when I was down. To Guillermo, who from day one became an instant friend and, since then, has brought bright colours into my life. To Max, who has the biggest heart and has been endlessly generous. I am sincerely grateful for all the hope, joy and gifts that they have given me. To Aisling, who saw it all and was there for me in every moment. I appreciate the non-stop support and I am truly grateful for the liveliness, real conversations and endless laughter that has been brought into my life. To Raúl, who made an important contribution to this dissertation, without whom I don't think I could have made it. I am sincerely
thankful for the friendship and all the help and encouragement. To Michael, who knew how to lift my spirits high.

To my Lebanese friends, who were always supportive and generous. An meine deutschen FreundInnen, die mich immer mit Freundlichkeit empfangen haben. To my Belgian friends, who welcomed me and showed me different ways to grow.

There are not enough words to describe how thankful I am to my parents, who were my very first teachers and friends. I send them my warmest love and gratitude for providing me with endless amounts of love, support and encouragement. To my two brothers, the Yin and the Yang, the brothers of honour at my wedding and in my life. I thank them for having my back. To my whole family for encouraging me. An meine deutsche Familie für ihre Unterstützung.

Finally, to Carlotta, my life partner, without whom I would have been lost in the journey of life. I am deeply grateful for your patience and unconditional support and encouragement. Herzlichen dank für den positiven Einfluss und den Beitrag, die mich zu dem gemacht hat, was ich bin.

Marc Sader
Ghent 29/08/2019

## Table of Contents

Acknowledgements ..... v
Table of Contents ..... vii
Summary ..... xiii
List of symbols ..... xxi
List of Tables ..... xxiii
List of Figures ..... xxxi
1 Introduction ..... 1
1.1 Background ..... 1
1.2 Problem setting ..... 5
1.3 The structure of this dissertation ..... 6
1.4 A brief overview of this dissertation ..... 8
1.4.1 Part|I| Preliminaries ..... 8
1.4.2 Part|II| Gathering sensory and instrumental data ..... 9
1.4.3 Part III Analysis of sensory data ..... 9
1.4.4 $\quad$ Part|IV| Prediction of sensory evaluation ..... 11
I Preliminaries ..... 13
2 The nature of sensory and instrumental data ..... 15
2.1 Introduction and overview ..... 15
2.2 Sensory evaluation and data ..... 15
2.2.1 Introduction to sensory evaluation ..... 16
2.2.2 Sensory measurement and analysis ..... 17
2.2.3 Determination of overall food quality ..... 21
2.3 Instrumental evaluation and data ..... 23
2.3.1 Microbial spoilage of food ..... 23
2.3.2 Instrumental techniques for quantifying VOCs ..... 25
2.3.3 Chemical analysis ..... 26
3 Mathematical prerequisites ..... 29
3.1 Introduction ..... 29
3.2 Notations and mathematical conventions ..... 30
3.3 Mathematical optimization methods ..... 33
3.3.1 Motivation ..... 33
3.3.2 General definitions ..... 34
3.3.3 Solving convex optimization problems ..... 37
3.3.4 Subclasses of convex optimization problems ..... 38
3.4 Statistical learning theory ..... 39
3.4.1 The standard framework ..... 39
3.4.2 Estimation and hypothesis space ..... 40
3.4.3 Overfitting and regularization ..... 42
3.5 Statistical learning as an optimization problem ..... 43
3.5.1 Regression problems ..... 43
3.5.2 Classification problems ..... 44
II Gathering sensory and instrumental data ..... 47
4 Sensory evaluation and chemical analysis of food products ..... 49
4.1 Introduction ..... 49
4.2 Experimental sensory evaluation set-up ..... 50
4.2.1 Preliminaries ..... 50
4.2.2 Chicken breast ..... 51
4.2.3 Atlantic cod ..... 53
4.2.4 Atlantic brown shrimp ..... 56
4.2.5 Atlantic salmon ..... 58
4.3 Quantification of spoilage-related VOCs ..... 63
III Analysis of sensory data ..... 65
5 Aggregation of ordinal labels ..... 67
5.1 Introduction ..... 67
5.2 Assigning a consensus label ..... 69
5.3 Consensus state problem. ..... 74
5.4 Interesting consensus states ..... 76
5.4.1 Unanimity ..... 76
5.4.2 Majority ..... 77
5.4.3 Marginal majority ..... 77
5.4.4 Monotonicity ..... 78
5.4.5 Marginal monotonicity ..... 80
5.4.6 A general comparison ..... 83
5.5 The optimization problem ..... 84
5.6 Independence of the search for different consensus states ..... 85
5.6.1 Unanimity, (marginal) majority and (marginal) monotonicity ..... 85
5.6.2 Marginal majority and majority ..... 86
5.6.3 Marginal monotonicity and monotonicity ..... 87
5.7 Case studies ..... 88
5.7.1 Dating system ..... 88
5.7.2 Movie preferences ..... 89
5.8 Application to sensory data ..... 91
5.8.1 Atlantic cod ..... 91
5.8.2 Chicken breast ..... 96
5.8.3 Brown shrimp ..... 96
5.8.4 Atlantic salmon ..... 98
5.9 Conclusions ..... 105
6 Aggregation of rankings ..... 109
6.1 Introduction ..... 109
6.2 Determining a consensus ranking ..... 110
6.2.1 The Borda count ..... 110
6.2.2 The method of Condorcet ..... 112
6.2 .3 The method of Kemeny ..... 116
6.2.4 Monotonicity of the votrix ..... 119
6.3 Consensus state problem ..... 122
6.4 The optimization problem ..... 123
6.5 Application of ranking rules to sensory data ..... 124
6.5.1 Chicken breast ..... 124
6.5.2 Atlantic cod ..... 129
6.5.3 Brown shrimp ..... 131
6.5.4 Atlantic salmon ..... 132
6.6 Conclusions ..... 135
7 Combining scores and rankings ..... 137
7.1 Introduction ..... 137
7.2 Obtaining consensus scores ..... 139
7.3 Obtaining consensus rankings ..... 141
7.4 Integrating scores and rankings ..... 143
7.4.1 Improving the quality of the assessment of ..... $\square$
$\square$ a consensus vector of scores ..... 143
7.4.2 Improving the quality of the assessment of a consensus ranking ..... 148
7.5 Information about samples ..... 151
7.5.1 Knowledge of storage days ..... 153
7.5.2 Results of clustering analysis ..... 154
7.5.3 Other sensory evaluation tests ..... 156
7.5.4 The constrained mode, median and mean ..... 162
7.6 Application to sensory data ..... 165
7.7 Integrating rankings for assigning consensus scores ..... 166
7.7.1 Integrating scores for determining consensus rankings ..... 171
7.7.2 Incorporating knowledge of storage days for assigning joint consensus scores ..... 172
7.7.3 Incorporating results of a clustering analysis for ..... $\square$
assigning joint consensus scores ..... 175
7.7.4 Incorporating consensus rankings for ..... 77
assigning joint consensus scores ..... 177
7.7.5 Comparing the consensus vectors of scores ..... 178
7.8 Conclusions ..... 180
IV Prediction of sensory evaluations ..... 183
8 Learning to predict ordinal labels ..... 185
8.1 Introduction ..... 185
8.2 Predictive modelling of ordinal labels ..... 187
8.2.1 The ordinal regression problem ..... 187
8.2.2 Performance measures for sensory evaluations ..... 190
8.2.3 Regularization in ordinal regression ..... 191
8.3 Experimental analysis ..... 192
8.3.1 General experimental setup ..... 192
8.3.2 Experimental settings using synthetic data ..... 194
8.4 Application to sensory data ..... 195
8.4.1 Chicken breast ..... 197
8.4.2 Atlantic cod ..... 197
8.4.3 Brown shrimp ..... 198
8.4.4 Atlantic salmon ..... 199
8.5 Conclusions ..... 199
9 Learning to predict rankings ..... 201
9.1 Introduction ..... 201
9.2 Approaches to predictive modelling of rankings ..... 202
9.2.1 The pointwise approach ..... 202
9.2.2 The pairwise approach ..... 203
9.2.3 From preferences to rankings ..... 204
9.2.4 Performance measures for preferences and rankings ..... 205
9.3 Application to sensory data ..... 205
9.3.1 Chicken breast ..... 205
9.3.2 Atlantic cod ..... 206
9.3.3 Brown shrimp ..... 208
9.3.4 Atlantic salmon ..... 209
9.4 Conclusions ..... 211
10 Integrating rankings in ordinal regression models ..... 213
10.1 Introduction ..... 213
10.2 The ordinal regression problem ..... 214
10.3 Incorporating relative evaluations ..... 215
10.3.1 Rankings ..... 215
10.3.2 Constraints ..... 216
10.4 Experimental analysis ..... 219
10.4.1 General experimental setup ..... 220
10.4.2 Experimental settings using synthetic data ..... 222
10.5 Application to sensory data ..... 225
10.5.1 Chicken breast ..... 225
10.5.2 Atlantic cod ..... 226
10.5.3 Atlantic brown shrimp ..... 229
10.5.4 Atlantic salmon ..... 230
10.6 Conclusions ..... 232
V Epilogue ..... 233
11 General conclusions and perspectives ..... 235
11.1 Analysis of sensory evaluations ..... 235
11.2 Prediction of sensory evaluations ..... 238
Bibliography ..... 243
A Appendix ..... 263
A. 1 Sensory evaluation for chicken breasts ..... 263
A. 2 Sensory evaluation for Atlantic cod ..... 266
A. 3 Sensory evaluation for Atlantic brown shrimp ..... 269
A. 4 Sensory evaluation for Atlantic salmon ..... 271
B Appendix ..... 279
B. 1 Quantification of VOCs in chicken breasts ..... 279
B. 2 Quantification of VOCs in Atlantic cod ..... 283
B. 3 Quantification of VOCs in Atlantic brown shrimp ..... 290
B. 4 Quantification of VOCs in Atlantic salmon ..... 293
Curriculum Vitae ..... 309

## Summary

In the following, we provide a summary of this dissertation in English and Dutch.

## English Summary

In recent years, sensory evaluation has seen an increase in the development of its science and application, and an increasing interest in the use of multiple sensory evaluation tests has resulted in the collection of different types of sensory data. In addition, chemical experiments have been prominent in gathering valuable instrumental data about the quality of food. In this dissertation, we discuss several approaches for analysing and predicting sensory evaluations in settings where there is a limited number of available trained panellists who provide sensory evaluations.

We start off with Chapter 1 by motivating the problem described above and highlighting the most important research questions that this dissertation answers. We then provide the structure and a brief overview of this dissertation.

In Chapter 2, we introduce the nature of sensory and instrumental data. First, we provide a review of the discriminative, affective and descriptive methods of sensory measurement and discuss the different types of sensory data generated from using these methods for obtaining appropriate information about sensory quality of food. Second, we introduce microbial spoilage as the most well-known cause of spoilage in food and discuss how storage conditions affect the smell or flavour of food. We briefly introduce several instrumental evaluation techniques, namely chemical testing techniques for quantifying volatile organic compounds and discuss several ways in which chemical analysis can be used to help interpret the results of sensory analysis.

In Chapter 3, we provide mathematical (optimization) methods as they are the foundation for understanding the main contributions presented in this dissertation. We review general definitions and properties found in the field of mathematical optimization and discuss solving convex optimization problems. Subsequently, we review statistical learning theories that are used to build predictive models and provide a motivation for the selection of different ingredients (loss function, hypothesis space and a regularisation parameter). It is supposed to serve as a tool for the reader to understand the different recipes used in this dissertation.

In Chapter 4. we describe the different foods, namely, chicken breasts, Atlantic cod
(Gadus morhua), Atlantic brown shrimp (Crangon crangon) and Atlantic salmon (Salmo salar) that are studied in this dissertation and delve into the techniques used to gather the sensory and instrumental data. In this chapter, we collect a number of datasets to answer the different research questions in this dissertation.

In Chapter 5. we discuss methods to determine the overall quality of food samples. Normally, trained panellists are asked to provide absolute evaluations, in the form of ordinal labels. These labels are used to assign a consensus label that describes the overall quality of a food sample. We introduce the most prominent methods used for the aggregation of ordinal labels to reach a consensus label. In addition, we propose a novel approach, based on the search for monotonicity, to assign (joint) consensus labels of multiple objects. Previously used techniques, such as the median and the mode, are compared to our novel approach on data sets from three real-world problems and are then applied to analyse the sensory data gathered in Chapter 4. We illustrate the different methods to show possible consensus labels of samples in settings where the consensus label is not clear.

In Chapter 6. we discuss methods to determine the overall ranking of food samples. Normally, untrained (or less trained) panellists are asked to provide a relative evaluation, in the form of a ranking. These rankings are used to determine a consensus ranking of the samples in terms of freshness. We introduce the most prominent methods used for the aggregation of rankings: the Borda count, the method of Condorcet, and the method of Kemeny. In addition, we consider the latest method that has been developed in-house, where the monotonicity of a representation of rankings provided by the panellists is exploited to reach a consensus ranking. The aforementioned methods are compared by applying them to analyse the sensory data gathered in Chapter 4 . We illustrate the different methods to show possible consensus rankings of the samples in settings where the consensus preference is not clear.

In Chapter 7, we present the problem of having a limited number of available trained panellists, and, thus, having a limited amount of data available to reach a consensus evaluation. Interestingly, it is quite common to invoke a more costefficient source of information. Thus, untrained panellists are used to gather some additional information. However, untrained panellists are obviously not as skilled as trained panellists, and might be unable to accurately evaluate a given sample. Since it is a conceptually easier task, untrained panellists are then just asked to rank different samples according to their personal appreciation. We propose to combine ordinal labels and rankings provided by trained and untrained panellists, respectively, to assign an improved consensus ordinal label that describes the overall quality of a food sample. We make use of the most prominent aggregation methods of labels and rankings to make a first attempt at combining absolute data, in the form of scores, and relative data, in the form of rankings. We first propose an approach for integrating rankings with scores to improve the quality assessment
of the consensus score. We propose a second approach for integrating scores with rankings to improve the assessment of the consensus ranking. In addition, we propose a third approach for incorporating different types of (relative) information to improve the quality of the consensus score. The application of these methods is illustrated on the sensory data gathered in Chapter 4, specifically, on scoring and ranking data of salmon samples (Section 4.2.5). We illustrate the different methods to show possible consensus scores of the samples, and provide guidelines for choosing an optimal source of additional information.

In Chapter 8, the problem is similar to that discussed in Chapter 5, where trained panellists are typically asked to provide absolute evaluations, in the form of ordinal labels, however, we propose a method to predict an ordinal label of a new sample. We present an ordinal regression model that takes instrumental data (i.e., features of the studied samples) as input and sensory data (i.e, assigned ordinal labels by trained panellists) as output. This model can then predict the quality of a food sample based solely on its chemical information. However, as we have previously discussed, training and (subsequently) collecting information from panellists usually carries big expenses. For this reason, there usually is a limited amount of data available to learn a good statistical model without the issue of overfitting the data. We present an approach for including $\ell_{1}$ regularisation into ordinal regression models to reduce overfitting and improve its performance. The ordinal regression problem is validated and applied on synthetic data and on real-life experimentation on the sensory and instrumental data gathered in Chapter 4. We show that the results of the studies are consistent and that this strategy is useful, especially for problems where the number of samples is very small in comparison to the total number of features.

In Chapter 9 , the problem is similar to that discussed in Chapter 6. where untrained (or less trained) panellists are asked to provide relative evaluations, in the form of rankings, however, here we discuss methods to predict a ranking of a set of samples. We discuss that ordinal regression can be used to predict rankings, however, we illustrate the pairwise approach as an alternative approach, where a ranking model takes instrumental data (i.e., difference in features of the pairs of samples) as input and sensory data (i.e, rankings that constitute preferences of pairs of samples) as output. This model can predict the preference of couples of samples, and, subsequently, predict a ranking. Since the number of samples is very small in comparison to the total number of features, we include $\ell_{1}$ regularisation. The pairwise approach is illustrated on the sensory and instrumental data gathered in Chapter 4

In Chapter 10, the problem is similar to that discussed in Chapter 7, where a limited number of trained panellists is available, however, we discuss the issue of having a limited amount of data to learn a good statistical model. We propose an approach that allows to integrate relative information, in the form of rankings, to
augment ordinal regression models. This approaches allows to reduce overfitting and improve the performance of ordinal regression models. This approach is validated and applied on synthetic data and on real-life experimentation on the sensory and instrumental data gathered in Chapter 4. We show that this strategy is useful, especially for problems where the number of samples is very small in comparison to the total number of features.

In Chapter 11, we end this dissertation with some general conclusions and a discussion on future research directions.

## Dutch summary <br> - Nederlandstalige samenvatting

In de afgelopen jaren zijn zowel het onderzoek naar als de toepassing van sensorische evaluatie toegenomen. Bovendien heeft een toenemende belangstelling voor het gebruik van meerdere sensorische evaluatietests geresulteerd in de verzameling van verschillende soorten van sensorische gegevens. Daarnaast zijn chemische experimenten prominent aanwezig geweest bij het verzamelen van waardevolle instrumentele gegevens over de kwaliteit van voedsel. In dit proefschrift bespreken we verschillende benaderingen voor het analyseren en voorspellen van sensorische evaluaties in onderzoeken/evaluaties waar slechts een beperkt aantal van de beschikbare opgeleide panelleden sensorische evaluaties aanleveren.

In Hoofdstuk 1 wordt het hierboven beschreven probleem gemotiveerd en de belangrijkste onderzoeksvragen geïntroduceerd die in dit proefschrift beantwoord zullen worden. Vervolgens bespreken we de structuur en geven we een kort samenvatting van dit proefschrift.

In Hoofdstuk 2 introduceren we de aard van sensorische en instrumentele gegevens. Ten eerste geven we een overzicht van de discriminerende, affectieve en beschrijvende methoden van sensorische meting en bespreken we de verschillende soorten van sensorische gegevens die gegenereerd worden door het gebruik van deze methoden voor het verkrijgen van gepaste informatie over de sensorische kwaliteit van voedsel. Ten tweede introduceren we microbieel bederf als de meest bekende oorzaak van bederf in voedsel en bespreken we hoe opslagomstandigheden de geur of smaak van voedsel beïnvloeden. We introduceren kort een aantal instrumentele evaluatietechnieken, namelijk chemische testtechnieken voor het kwantificeren van vluchtige organische stoffen en we bespreken verschillende manieren waarop chemische analyse kan worden gebruikt om de resultaten van sensorische analyse te interpreteren.

In Hoofdstuk 3 leveren we wiskundige optimalisatiemethoden aan, aangezien ze de basis vormen voor het begrijpen van de belangrijkste bijdragen die in dit proefschrift gepresenteerd zullen worden. We bekijken algemene definities en eigenschappen
die gevonden zijn op het gebied van wiskundige optimalisatie en we bespreken het oplossen van convexe optimalisatieproblemen. Vervolgens bekijken we statistische leer theorieën die gebruikt zullen worden om voorspellende modellen op te bouwen en leveren we een motivatie aan voor de selectie van verschillende ingrediënten (verliesfunctie, hypotheseruimte en een regularisatieparameter). Het hoofdstuk zal als hulpmiddel moeten zijn voor de lezer om de verschillende recepten te begrijpen die in dit proefschrift gebruikt zullen worden.

In Hoofdstuk 4 beschrijven we de verschillende voedingsmiddelen, namelijk kippenborsten, Atlantische kabeljauw (Gadus morhua), Atlantische bruine garnaal (Crangon crangon) en Atlantische zalm (Salmo salar) die in dit proefschrift bestudeerd zullen worden en gaan we dieper in op de technieken die gebruikt zullen worden om de sensorische en instrumentele gegevens te verzamelen. In dit hoofdstuk verzamelen we een aantal datasets om de verschillende onderzoeksvragen in dit proefschrift te beantwoorden.

In Hoofdstuk 5 bespreken we methoden om de algemene kwaliteit van voedselmonsters te bepalen. Normal gesrpoken worden opgeleide panelleden gevraagd absolute evaluaties, in de vorm van ordinale labels, aan te leveren. Deze labels worden gebruikt om een consensuslabel toe te wijzen dat de algemene kwaliteit van een voedselmonster beschrijft. We introduceren de meest prominente methoden die gebruikt worden voor de aggregatie van ordinale labels om een consensuslabel te bereiken. Daarnaast stellen we een nieuwe aanpak voor, gebaseerd op het zoeken naar monotoniciteit, om (gezamenlijke) consensuslabels van meerdere objecten toe te wijzen. Eerder gebruikte technieken, zoals de mediaan en de modus, worden vergeleken met onze nieuwe benadering door datasets uit drie praktijkproblemen en worden vervolgens toegepast om de sensorische gegevens te analyseren die in Hoofdstuk 4 verzameld werden. We illustreren de verschillende methoden om mogelijke consensuslabels van monsters te laten zien in instellingen waar het consensuslabel niet duidelijk is.

In Hoofdstuk 6 bespreken we methoden om in de algemene rangschikking van voedselmonsters te bepalen. Normal gesrpoken worden niet opgeleide (of minder opgeleide) panelleden gevraagd een relatieve evaluatie, in de vorm van een rangschikking, aan te leveren. Deze rangschikkingen worden gebruikt om een consensusrangschikking van de monsters te bepalen qua versheid. We introduceren de meest prominente methoden die gebruikt worden voor aggregatie van rangschikkingen: het Borda-aantal, de methode van Condorcet en de methode van Kemeny. Daarnaast beschouwen we de nieuwste methode die intern ontwikkeld wordt, waarbij de monotonie van een representatie van rangschikkingen door de panelleden wordt benut om een consensus te bereiken. De bovengenoemde methoden worden vergeleken door ze toe te passen om de sensorische gegevens te analyseren die in Hoofdstuk 4 verzameld werden. We illustreren de verschillende methoden om mogelijke consensusrangschikkingen van de monsters te tonen in instellingen waar
de consensusvoorkeur niet duidelijk is.
In Hoofdstuk 7 presenteren we het probleem van een beperkt aantal aan beschikbare opgeleide panelleden, waarbij er dus slechts een kleine hoeveelheid gegevens beschikbaar is om tot een consensusevaluatie te komen. Het is opmerkelijk dat het vrij gebruikelijk is om een meer kosteneffcinte bron van informatie aan te roepen. Niet-opgeleide panelleden worden dus gebruikt om wat extra informatie te verzamelen. Niet-opgeleide panelleden zijn echter duidelijk niet zo bekwaam als opgeleide panelleden en ze zouden mogelijks niet in staat zijn om een bepaald monster accuraat te evalueren. Omdat het een conceptueel eenvoudigere taak is, worden niet-opgeleide panelleden vervolgens slechts gevraagd om verschillende monsters te rangschikkingen op basis van hun persoonlijke waardering. We stellen voor om ordinale labels en rangschikkingen die opgegeven worden door zowel opgeleide en niet-opgeleide panelleden te combineren, om een verbeterd algemeen ordinaal label toe te wijzen dat de algemene kwaliteit van een voedingsmonster beschrijft. We maken gebruik van de meest prominente aggregatiemethoden van labels en rangschikkingen om een eerste poging te doen om absolute gegevens, in de vorm van scores, en relatieve gegevens, in de vorm van rangschikkingen, te combineren. We stellen eerst een benadering voor om rangschikkingen met scores te integreren om de kwaliteitsbeoordeling van de consensusscore te verbeteren. We stellen een tweede benadering voor om scores met rangschikkingen te integreren om de beoordeling van de consensusrangschikking te verbeteren. Daarnaast stellen we een derde benadering voor om verschillende soorten (relatieve) informatie op te nemen om de kwaliteit van de consensusscore te verbeteren. De toepassing van deze methoden wordt geillustreerd aan de hand van de sensorische gegevens die in Hoofdstuk 4 verzameld werden, met name over het scoren en rangschikken van gegevens van zalmmonsters (sectie 4.2.5). We illustreren de verschillende methoden om mogelijke consensusscores van de monsters te laten zien, en geven richtlijnen voor het kiezen van een optimale bron van aanvullende informatie.

In Hoofdstuk 8 is het probleem vergelijkbaar met het probleem dat in Hoofdstuk 5 besproken wordt, met name waar opgeleide panelleden doorgaans worden gevraagd om absolute evaluaties, in de vorm van ordinale labels, aan te leveren. In Hoofdstuk 8 daarentegen, stellen we een methode voor om een ordinaal label van een nieuw monster te voorspellen. We presenteren een ordinaal regressiemodel dat instrumentele gegevens (d.w.z. kenmerken van de onderzochte monsters) als invoer en sensorische gegevens (d.w.z. toegewezen ordinale labels door opgeleide panelleden) als uitvoer neemt. Dit model kan vervolgens de kwaliteit van een voedselmonster voorspellen op basis van alleen de chemische informatie. Zoals we eerder hebben besproken, brengt de opleiding en (vervolgens) het verzamelen van informatie van panelleden meestal grote kosten mee. Om deze reden is er meestal een kleine hoeveelheid gegevens beschikbaar om een goed statistisch model te leren zonder de kwestie van overfitting van de gegevens. We presenteren een aanpak om $\ell_{1}$ regularisatie op te nemen in ordinale regressiemodellen om overfitting te
verminderen en de prestaties ervan te verbeteren. Het ordinale regressieprobleem wordt gevalideerd en toegepast op synthetische gegevens en op experimenten in de praktijk op de sensorische en instrumentele gegevens die in oofdstuk 4 verzameld werden. We laten zien dat de resultaten van de onderzoeken consistent zijn en dat deze strategie nuttig is, vooral voor problemen waarbij het aantal monsters erg klein is in vergelijking met het totale aantal functies.

In Hoofdstuk 9 is het probleem vergelijkbaar met het probleem dat in Hoofdstuk 6 besproken wordt, met name waar niet opgeleide (of minder opgeleide) panelleden worden gevraagd om relatieve evaluaties, in de vorm van rangschikkingen aan te leveren. In Hoofdstuk 9 daarentegen bespreken we methoden om een rangschikking van een reeks monsters voor te spellen. We bespreken dat ordinale regressie kan gebruikt worden om rangschikkingen voor te spellen. We illustreren echter de paarsgewijze benadering als een alternatieve benadering, waarbij een rangschikkingsmodel instrumentele gegevens (d.w.z. verschil in kenmerken van de paren samples) als invoer en sensorische gegevens (d.w.z. rangschikkingen die voorkeuren van paren monsters vormen) als uitvoer neemt. Dit model kan de voorkeur van paren van monsters voorspellen en vervolgens een rangschikking voorspellen. Omdat het aantal monsters erg klein is in vergelijking met het totale aantal features, bevatten we $\ell_{1}$ regularisatie. De paarsgewijze benadering wordt geillustreerd op de sensorische eninstrumentele gegevens die in Hoofdstuk 4 verzameld werden.

In Hoofdstuk 10 is het probleem vergelijkbaar met het probleem dat in Hoofdstuk 7 besproken wordt, namelijk dat van een beperkt aantal aan beschikbare opgeleide panelleden, maar in Hoofdstuk 10 bespreken we het probleem waarin slechts een kleine hoeveelheid gegevens beschikbaar is om een goed statistisch model te leren . We stellen een benadering voor die toelaat om relatieve informatie, in de vorm van rangschikkingen, te integreren om de prestaties van ordinale regressiemodellen te verbeteren. Deze benadering maakt het mogelijk om overfftting te verminderen en de prestaties van ordinale regressiemodellen te verbeteren. Deze benadering wordt gevalideerd en toegepast op synthetische gegevens en op experimenten in de praktijk op de sensorische en instrumentele gegevens die in Hoofdstuk 4 verzameld werden. We laten zien dat deze strategie nuttig is, vooral voor problemen waarbij het aantal monsters erg klein is in vergelijking met het totale aantalfuncties.

In Hoofdstuk 11 beëindigen we dit proefschrift met enkele algemene conclusies en een discussie over toekomstige onderzoeksrichtingen.

## List of symbols

| $\mathbb{H}$ | Hypothesis space |
| :--- | :--- |
| $\mathbb{N}$ | Set of natural numbers |
| $\mathbb{R}$ | Set of real numbers |
| $\mathbb{R}^{+}$ | Set of positive real numbers |
| $\mathbb{Z}$ | Set of integers |
| $\mathbb{Z}^{+}$ | Set of positive integers |
| $\mathscr{A}$ | Set of samples |
| $\mathscr{D}$ | Dataset |
| $\mathscr{T}$ | Training dataset |
| $\mathscr{L}$ | Set of ordinal labels |
| $\mathscr{P}$ | Set of couples of samples |
| $\mathscr{R}$ | Set of all strict rankings |
| $\mathscr{R}$ | Set of all rankings with ties |
| $\mathcal{S}$ | Set of constraints |
| $\mathcal{E}$ | Error term |
| $\mathcal{G}$ | Latent random variable |
| $\mathcal{L}$ | Loss function |
| $\mathcal{P}$ | Power set |
| $\mathcal{R}$ | Relation (either binary or ternary) |
| $\mathcal{S}$ | Consensus state |
| $\mathcal{X}$ | Input random vector |
| $\mathcal{Y}$ | Output random vector |
| $a^{k}$ |  |
| $a_{j}$ | Sample stored for $k$ days |
| $B_{\mathbf{r}}$ | j-th sample in set $\mathscr{A}$ |
| $B_{T}$ | Borda count |
| $B_{U}$ | Upper bound on the number of possible vector of scores |
| $\mathbf{C}$ | Upper bound on the number of possible rankings with ties |
| $C_{\alpha}(\mathbf{s})$ | Matrix of costs |
| $D_{\alpha}(\precsim)$ | Convex combination of distances associated with vector of scores s |
| $\partial$ | Convex combination of distances associated with ranking with ties $\precsim$ |
| $d_{\text {KL }}$ | Distance function |
| $d_{\text {EmD }}$ | Earth mivergence |
| $g$ | Latent variable function |
| $K$ | Sum of Kendall (Kemeny in the case of rankings with ties) distances |
|  |  |


| $M, \mathbf{M}, \mathbb{M}$ | Monometric |
| :--- | :--- |
| $n_{T}$ | Number of trained panellists |
| $n_{U}$ | Number of untrained panellists |
| $P$ | Probability distribution function |
| $r_{e m p}$ | Empirical risk |
| $r_{r e g}$ | Regularized empirical risk |
| $\mathbf{r}$ | Profile of strict rankings |
| $\widetilde{\mathbf{r}}$ | Profile of rankings with ties |
| $R$ | Regularisation term |
| $s_{i j}$ | Score assigned by the $i$-th panellist to the $j$-th sample |
| $\mathbf{s}_{i}(j)$ | Score assigned to the $j$-th sample in the $i$-th vector of scores |
| $L$ | Ordinal label |
| $\mathbf{V}$ | Votrix |
| $\widetilde{\mathbf{V}}$ | Weak votrix |
| $\mathbf{w}$ | Vector of model weight parameters |
| $\mathbf{x}$ | Feature vector |
| $y$ | Ordinal label |
| $\mathbf{z}$ | List of labels |
| $\mathbf{Z}$ | Matrix of labels |
| $\alpha$ | Parameter for convex combinations |
| $\boldsymbol{\beta}$ | Vector of non-negative model weight parameters |
| $\epsilon$ | Threshold on absolute distances |
| $\boldsymbol{\theta}$ | Vector of model threshold parameters |
| $\vartheta_{\mathbf{s}}$ | Set of all rankings with ties that do not contradict s |
| $\varphi \precsim$ | Set of all vectors of scores that do not contradict $\precsim$ |
| $\lambda$ | Regularization parameter |
| $\xi, \xi^{\prime}, \zeta$ | Slack variables |
| $\pi, \pi^{\prime}, \mu$ | Weight variables for the slack variables |
| $\ell$ | Likelihood function |
| $\prec$ | Strict ranking |
| $\precsim$ | Ranking with ties |
| $\wp_{l e x}$ | Lexicographic order on a set of couples |
| $\leq \mathscr{L}$ | Total order relation on the ordinal scale $\mathscr{L}$ |
| $\sqsubseteq$ | Order relation |
| $\sqsubseteq_{\mathbf{z}}$ |  |

## | List of Tables

2.1 Summary of the most common tests for sensory evaluation, asdescribed in Sensory Evaluation of Food [1].19
4.1 Sampling chicken samples packaged at day 0 for quantification of VOCs and sensory evaluation (labelling and ranking) tests. ..... 52
4.2 Chicken samples for labelling tests and the number of panellists in ..... 52
4.3 Chicken samples for ranking tests and the number of panellists in each storage experiment. ..... 53
4.4 Sampling cod samples packaged at day 0 for quantification of VOCs ..... 54
4.5 Cod samples for labelling tests and the number of panellists in each storage experiment. ..... 55
4.6 Cod samples for ranking tests and the number of panellists in each storage experiment. ..... 56
4.7 Sampling shrimp samples packaged at day 0 for quantification of VOCs and sensory evaluation (labelling and ranking) tests. ..... 57
4.8 Shrimp samples for labelling tests and the number of panellists in each storage experiment. ..... 58
4.9 Shrimp samples for ranking tests and the number of panellists in each storage experiment. ..... 58
4.10 Sampling salmon samples packaged at day 0 for quantification of VOCs and sensory evaluation (labelling, ranking and scoring) tests. Experiment AN4 denoted with * consists of scoring tests instead oflabelling tests.59
4.11 Salmon samples for labelling tests and the number of panellists in each storage experiment. ..... 60
4.12 Salmon samples for ranking tests and the number of panellists in each storage experiment. ..... 61
4.13 The procedure to select at different storage days the samples from ..... 61
4.14 The order of grouping the salmon samples from different storage days (represented by the corresponding superindex) and the day of the week each group was provided to the panellists. ..... 62
5.1 Expressed lists of labels and their frequency for the three candidates. ..... 89
5.2 Expressed lists of labels and their frequency for the four movies. ..... 90
5.3 Expressed list of labels for the four samples. ..... 91
5.4 Closest monotone matrix of labels w.r.t. ( $L_{5}, L_{3}, L_{3}, L_{1}$ ) (left) and( $L_{5}, L_{3}, L_{2}, L_{1}$ ) (right) given in the form of a list of evaluations.93
5.5 Consensus labelling of $\mathscr{A}$ for the different methods. ..... 93
5.6 Consensus labelling of the set of cod samples from company A described in Table 4.5 for each storage experiment by aggregatingthe labels gathered in Table $\mid$ A. 4 |using the different methods. . . . 95
5.7 Consensus labelling of the set of chicken samples described in Ta- ble 4.2 for experiment H4 by aggregating the labels gathered inTable|A.1 using the different methods.96
5.8 Consensus labelling of the set of shrimp samples described in Ta-

| ble 4.8 for every session (1 and 2) and for each storage experiment |
| :---: | :--- | :--- |
| by aggregating the labels gathered in Table\|A.7|using the different |
| methods. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 97 |

5.9 Consensus labelling of the set of salmon samples described in Ta- ble 4.11 for session 1 of each storage experiment by aggregating thelabels gathered in Table|A.9|using the different methods.99
5.9 (Continued) Consensus labelling of the set of salmon samples de- scribed in Table 4.11 for session 1 of each storage experiment by aggregating the labels gathered in Table $\mid$ A. $9 \mid$ using the different methods. ..... 100
5.10 Consensus labelling of the set of salmon samples described in Ta- ble 4.11 for session 2 of each storage experiment by aggregating thelabels gathered in Table|A.9|using the different methods.101
5.11 Consensus labelling of the set of salmon samples described in Ta- ble 4.11 for session 3 of each storage experiment by aggregating thelabels gathered in Table|A.9|using the different methods.103
5.12 Consensus labelling of the set of salmon samples described in Ta- ble 4.11 for session 4 of each storage experiment by aggregating the labels gathered in Table|A.9 using the different methods. ..... 106

5.13 Consensus vectors of scores of the sets of salmon samples described| in Table 4.14 for experiment AN4* and for every group (1-5) by |
| :---: |aggregating the scores gathered in Table|A.11|using the differentmethods.107

6.1 Profile of rankings on $\mathscr{X}$ given by fifteen panellists. ..... 111
6.2 Profile of rankings on $\mathscr{A}$ given by fifteen panellists. ..... 115
6.3 Profile of rankings on $X$ given by thirteen panellists. ..... 115
6.4 The number of reversals needed to make each ranking a unanimous ..... 116
6.5 The number of reversals needed to reach each possible ranking. The minimum number of reversals is shown in bold. ..... 117
6.6 Rankings on $X$ given by four panellists. ..... 119
6.7 Kemeny distance $d_{K}$ vs. weighted Kemeny distance $d_{K, \mathbf{w}}$. ..... 119
6.8 The number of changes needed to impose monotonicity w.r.t. eachranking on $X$.121
6.9 Consensus rankings of chicken samples described in Table 4.3 for experiments L4 and H4 in every session (1 and 2), in Table A.2 for the different methods. The symbol ${ }^{* *}$ means that the ranking is the closest Condorcet ranking.125
6.9 (Continued) Consensus rankings of chicken samples described in Table 4.3 for experiments L4 and H4 in every session (1 and 2), in Table A.2, for the different methods. The symbol ${ }^{(*)}$ means that the ranking is the closest Condorcet ranking. 126
6.9 (Continued) Consensus rankings of chicken samples described in Table 4.3 for experiments L4 and H4 in every session (1 and 2), in Table A. 2 , for the different methods. The symbol ${ }^{{ }^{* *}}$ means that the ranking is the closest Condorcet ranking.
6.9 (Continued) Consensus rankings of chicken samples described in Table 4.3 for experiments L4 and H4 in every session (1 and 2), in Table|A.2, for the different methods. The symbol ${ }^{(*)}$ means that the ranking is the closest Condorcet ranking.
6.10 Consensus rankings of chicken samples described in Table 4.3 for experiments L8 and H8, in Table|A.3 for the different methods. The symbol ${ }^{\text {(*) }}$ means that the ranking is the closest Condorcet ranking. 128
6.11 Consensus rankings of cod samples described in Table 4.6 for experiments L4, H4 and H8 in every session (1 and 2) and experiments L8 and A4, in Table|A.6, for the different methods. The symbol '*, means that the ranking is the closest Condorcet ranking.
6.12 Consensus rankings of shrimp samples described in Table 4.9 for experiments L4 and H4 in every session (1 and 2), in Table A.8 for the different methods. 131
6.13 Consensus rankings of salmon samples described in Table 4.12 for experiments H4, AN4, ANH4, A4, L4 and M4 in session 1 in Table A. 10 , for the different methods. The symbol ${ }^{\text {'*' }}$ means that the ranking is the closest Condorcet ranking.

6.14 Consensus rankings of salmon samples described in Table 4.12 for
experiments H4, AN4, ANH4 and A4 in sessions 2, 3 and 4 in
Table $\mid$ A.10, for the different methods. The symbol ${ }^{{ }^{*} \text {, } \text { means that }}$
the ranking is the closest Condorcet ranking.
133
6.14 Consensus rankings of salmon samples described in Table 4.12 for experiments H4, AN4, ANH4 and A4 in sessions 2, 3 and 4 in Table $\mid$ A.10, for the different methods. The symbol ${ }^{{ }^{*} \text {, } \text { means that }}$ the ranking is the closest Condorcet ranking. . . . . . . . . . . . . 134
6.14 (Continued) Consensus rankings of salmon samples described in Table 4.12 for experiments H4, AN4, ANH4 and A4 in sessions 2, 3 and 4 in Table A.10] for the different methods. The symbol ${ }^{(*)}$ means that the ranking is the closest Condorcet ranking. ..... 134
6.15 The consensus rankings of salmon samples ( $\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}$ ) in each group (1-4). ..... 136
7.1 The scores assigned to the samples $a_{1}, a_{2}, a_{3}$ and $a_{4}$ by the trained panellists in Example|7.1| ..... 140
7.2 The zero-one distance $d_{0}\left(\mathbf{s}, \mathbf{s}_{i}\right)$, absolute difference $d_{1}\left(\mathbf{s}, \mathbf{s}_{i}\right)$ and squared difference $d_{2}\left(\mathbf{s}, \mathbf{s}_{i}\right)$ between $\mathbf{s}=(2,2,3,4)$ and each vector of scores $\mathbf{s}_{1}, \mathbf{s}_{2}$ and $\mathbf{s}_{3}$ provided by the trained panellists. ..... 141
7.3 The rankings of samples $a_{1}, a_{2}, a_{3}$ and $a_{4}$ expressed by the untrained ..... 142
7.4 The Kemeny distance $d_{K}\left(\precsim, \precsim_{i}\right)$ between $\prec=a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ and each ranking $\precsim_{1}, \precsim_{2}, \precsim_{3}, \precsim_{4}$ and $\precsim_{5}$ provided by the untrained panellists. ..... 143
7.5 Sum of Kemeny distances between each ranking $\prec$ that does not contradict $\mathbf{s}=(2,2,3,4)$ and the rankings provided by the untrained panellists. ..... 146
7.6 Sum of absolute distances between each vector of scores $\mathbf{s}$ that does not contradict $\precsim=a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ and the vectors of scoresprovided by the trained panellists.151
7.7 The scores assigned by the panellists in Example 7.5. ..... 163
7.8 Sum of zero-one distance $\partial_{0}$, absolute distance $\partial_{1}$ and squared difference $\partial_{2}$ between the scores provided by the panellists and all possible vectors of scores. The minimizers are shown in bold and the vectors of scores in which the first sample is assigned a score greater than or equal to the score of the second sample are highlighted in blue. ..... 164
7.9 The scores assigned by the panellists in Example|7.6| ..... 165
7.10 The constraints based on the storage days of the samples. ..... 174
7.11 The median and the constrained median for each sample after incor- porating knowledge of storage days. The differences are highlighted.174
7.12 Clustered samples based on the similarity of their VOC profile. ..... 176
7.13 The median and the constrained median for each sample after in- corporating the results of clustering analysis. The differences are
176
highlighted.
7.14 The minimizers s* and $\swarrow^{*}$ of salmon samples of fillets (A, B, C, D) on each day of the week in the order shown in Table 4.14$]$ ..... 177

A. 7 Labels assigned by panellists to shrimp samples in each storage experiment described in Table 4.8. ..... 269
A. 8 Rankings of shrimp samples expressed by the panellists in each experiment described in Table 4.9. ..... 270
A. 9 Labels assigned by panellists to salmon samples in each storage experiment described in Table 4.11$]$ ..... 271
A. 9 (Continued) Labels assigned by panellists to salmon samples in each storage experiment described in Table 4.11] ..... 272
A. 10 Rankings of salmon samples expressed by the panellists in each experiment described in Table 4.12 | ..... 273
A. 10 (Continued) Rankings of salmon samples expressed by the panellists in each experiment described in Table $4.12 \square$. ..... 274
A. 11 Vectors of scores assigned to groups of samples of salmon fillets (A, B, C, D) by the trained panellists on each day of the week in the order shown in Table|4.14] ..... 275
A. 12 Rankings with ties of samples of salmon fillets (A, B, C, D) expressed by the untrained panellists on each day of the week in the order shown in Table 4.14 ] ..... 276
A. 12 (Continued) Rankings with ties of samples of salmon fillets (A, B, C, D) expressed by the untrained panellists on each day of the week in the order shown in Table 4.14] ..... 277
B. 1 Volatile organic compounds (VOCs) quantified in packaged chicken samples with SIFT-MS: product ions, mass to charge ratios (m/z), branching ratios (b) and reaction rate coefficients (k). Product ions denoted with * were selected for quantifying the respective VOC. 279
B. 2 Measured concentrations $\left(\mu \mathrm{g} / \mathrm{m}^{3}\right)$ of VOCs detected with SIFT-MS on each day in the headspace of chicken samples used for labelling tests. ..... 280
B. 3 Measured concentrations $\left(\mu \mathrm{g} / \mathrm{m}^{3}\right)$ of VOCs detected with SIFT-MS on each day in the headspace of chicken samples used for ranking tests. ..... 281
B. 3 (Continued) Measured concentrations ( $\mathrm{\mu g} / \mathrm{m}^{3}$ ) of VOCs detected with SIFT-MS on each day in the headspace of chicken samples used for ranking tests. ..... 282
B. 4 Volatile organic compounds (VOCs) quantified in packaged codsamples with SIFT-MS: product ions, mass to charge ratios (m/z),branching ratios (b) and reaction rate coefficients (k). Product ionsdenoted with * were selected for quantifying the respective VOC. 283
B. 5 Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of cod samples used for labelling tests. ..... 284

|  |  |
| :---: | :---: |
| SIFT-MS on each day in the headspace of cod samples used for |  |
| labelling tests. |  |
| B. 5 (Continued) Measured concentrations (ppb) of VOCs detected with |  |
| SIFT-MS on each day in the headspace of cod samples used for |  |
| labelling tests. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . |  |
|  | Measured concentrations (ppb) of VOCs detected with SIFT-MS o |
| each day in the headspace of cod samples used for ranking tests. |  |
|  | (Continued) Measured concentrations (ppb) of VOCs detected with |
| SIFT-MS on each day in the headspace of cod samples used for |  |
| ranking tests. . . . . . . . . . . . . . . . . . . . . . . . . . . . |  |
| (Continued) Measured concentrations (ppb) of VOCs detected with |  |
| SIFT-MS on each day in the headspace of cod samples used for |  |
| ranking tests. . . . . . . . . . . . . . . . . . . . . . . . . . . . . |  |
| Volatile organic compounds (VOCs) quantified in packaged shrimp |  |
| samples with SIFT-MS: mass to charge ratios (m/z), branching |  |
| ratios (b), reaction rate coefficients (k) and product ions. Product |  |
| ions denoted with * were selected for quantifying the respective VOC. 290 |  |
| Measure |  |
| each day in the headspace of shrimp samples used for labelling tests. 291 |  |
| Measured concentrations (ppb) of VOCs detected with SIFT-MS on |  |
| each day in the headspace of shrimp samples used for ranking tests. 292 |  |
| 10 Volatile organic compounds (VOCs) quantified in packaged salmon |  |
| samples with SIFT-MS: product ions, mass to charge ratios (m/z), |  |
| branching ratios (b) and reaction rate coefficients (k). Product ions |  |
| denoted with * were selected for quantifying the respective VOC. |  |
| B. 11 Measured concentrations (ppb) of VOCs detected with SIFT-MS on |  |
| each day in the headspace of salmon samples used for ranking tests. 294 |  |
| 1 (Continued) Measured concentrations (ppb) of VOCs detected with |  |
| SIFT-MS on each day in the headspace of salmon samples used for |  |
| ranking tests. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . |  |
| B. 12 Measured concentrations (ppb) of VOCs detected with SIFT-MS on |  |
| each day in the headspace of salmon samples used for labelling tests. |  |
| B. 14 Measured concentrations (ppb) of VOCs detected with SIFT-MS on |  |
| each day in the headspace of salmon samples used in experiment |  |
| AN4* (scoring and ranking with ties). |  |

## List of Figures

1.1 An illustrative concept of a silicon photonics-based chemical sensor with a matrix of selective coating materials and the three modelling stages. ..... 4
1.2 The structure of this dissertation. ..... 7
2.1 The nature of sensory data and the training requirement of panellists. ..... 17
2.2 Example of rating seven descriptors on 5-point scoring scales ..... 21
3.1 Example of a Hasse diagram of the order relation $<$ on set $A$. ..... 32
4.1 Example of sensory evaluation facility showing the various activities
50 for gathering sensory evaluation data.
$4.2 \quad 5$-point scale used by panellists, where the extreme scores of " 1 " and " 5 " represent spoiled and fresh, respectively, and the intermediatescore of " 3 " represents a neutral response of neither spoiled nor fresh.62
4.3 The setup for monitoring spoilage metabolites in packaged chicken using syringes for the inlet of the SIFT-MS and for creating an opensystem.63
5.1 Natural interpretation of the use of a monometric on the ordinal ..... 72
5.2 Hasse diagram of $\sqsubseteq_{\mathbf{z}}$ for $\mathbf{z}=(G, G)$ (left) and $\mathbf{z}=(A, G)$ (right). ..... 79
5.3 Hasse diagram of $\sqsubseteq_{\mathbf{z}}$ for $\mathbf{z}=(G, G)$. ..... 80
5.4 Example of a monotone matrix of labels w.r.t. $\left(L_{4}, L_{4}\right)$. ..... 81
5.5 Hasse diagram of $\sqsubseteq_{L}$ for $L=G$ (left) and $L=A$ (right). ..... 81
5.6 Example of a marginally monotone matrix of labels w.r.t. $\left(L_{4}, L_{4}\right)$. ..... 82
5.7 Relations between the different consensus states. ..... 83
5.8 Hasse diagram of $\sqsubseteq_{\mathbf{z}}$ for $\mathbf{z}=\left(L_{3}, L_{3}\right)$ (left) and $\mathbf{z}=\left(L_{2}, L_{2}\right)$ (right). ..... 88
6.1 Hasse diagram of $\sqsubset$ for the ranking $d \prec c \prec b \prec a$. ..... 120
6.2 Votrix represented on the Hasse diagram of $\sqsubset$ for $d \prec c \prec b \prec a$. ..... 121
7.1 Illustrating the vectors of scores $\mathbf{s}_{\alpha}^{*}$ that minimize $C_{\alpha}(\mathbf{s})$ for $\alpha \in[0,1]$ in Example 7.3. ..... 147
7.2 A visualization of the scores assigned to each sample $a_{1}, a_{2}, a_{3}$ and$a_{4}$ in Example 7.3 for (a) $\alpha>\frac{4}{9}$ and for (b) and (c) $0<\alpha<\frac{4}{9}$. . . 148
7.3 Illustrating the rankings $\precsim_{\alpha}^{*}$ that minimize $D_{\alpha}(\precsim)$ for $\alpha \in[0,1]$ inExample [7.4]152

8.1 (a) A model underlying ordinal data for the case of 3 ordinal labels, the horizontal axis indicates the value of a (one-dimensional) feature vector $\mathbf{x}$, the vertical axes contain the random variable $\mathcal{Y}$ that falls in category $L_{k}$ when the latent variable $g$ falls in the $k$-th interval of values. The intervals are determined by the thresholds $\theta_{1}$ and $\theta_{2}$ on the latent variable. The latent random variable $\mathcal{G}$ is assumed to follow a logistic distribution with mean $g(\mathcal{X}=\mathbf{x})$. The highlighted areas indicate the probabilities for the case of $k=2$, such that $\operatorname{Pr}\left(\mathcal{Y}=L_{2} \mid \mathcal{X}=\mathbf{x}_{1}\right)$ and $\operatorname{Pr}\left(\mathcal{Y}=L_{2} \mid \mathcal{X}=\mathbf{x}_{2}\right)$. (b) A visualization of the proportional odds model with the latent variable on the horizontal axis and the cumulative probabilities on the vertical axis. 189
8.2 Experiment illustrating the mean KL divergence values of standard ordinal regression models as reference (blue line) with possible values (shaded blue) in comparison to the mean KL divergence values of $\ell_{1}$-regularized ordinal regression models (orange line) with possible values (shaded orange) for an increasing number of features p. . . . 196
10.1 Experiments comparing $\ell_{1}$-regularized ordinal regression models learned on all labels as reference (dashed black line) with models learned on half the labels and integrating an increasing number of ranking tasks of two samples (blue line), three samples (green line), and four samples (orange line).

| 10.2 Re-evaluation of the results shown in Figures | 10.1 (a) and | 10.1 (c) |
| :---: | :---: | :---: | using the Earth mover's distance metric. . . . . . . . . . . . . . . . 224

10.3 Experiment illustrating the KL divergence value of an $\ell_{1}$-regularized ordinal regression model as reference (dashed black line) in comparison to the mean (blue line) and range (shaded) of the KL divergence values of $\ell_{1}$-regularized ordinal regression models with integrated ranking information of chicken breasts.
10.4 Experiment illustrating the KL divergence value of an $\ell_{1}$-regularized ordinal regression model as reference (dashed black line) in comparison to the mean (blue line) and range (shaded) of the KL divergence values of $\ell_{1}$-regularized ordinal regression models with integrated ranking information of Atlantic cod.
10.5 Experiment illustrating the KL divergence value of an $\ell_{1}$-regularized ordinal regression model as reference (dashed black line) in comparison to the mean (blue line) and range (shaded) of the KL divergence values of $\ell_{1}$-regularized ordinal regression models with integrated ranking information of Atlantic brown shrimp.

| 10.6 Experiment illustrating the KL divergence value of an $\ell_{1}$-regularized |
| :---: |
| ordinal regression model as reference (dashed black line) in compari- |
| son to the mean (blue line) and range (shaded) of the KL divergence |
| values of $\ell_{1}$-regularized ordinal regression models with integrated |
| ranking information of Atlantic salmon. . . . . . . . . . . . . . . . 232 |

## 1 Introduction

## Table of Contents

\author{

1.1 Background <br> 1.2 Problem setting <br> 1.3 The structure of this dissertation <br> 1.4 A brief overview of this dissertation <br> 1.4.1 Part|II Preliminaries <br> 1.4.2 $\quad$ Part $\mid$ II $\mid$ Gathering sensory and instrumental data <br> 1.4.3 Part III Analysis of sensory data <br> | 1.4 .4 | Part | IV | Prediction of sensory evaluation |
| :--- | :--- | :--- | :--- |

}

### 1.1. Background

Over the past decades, there has been an excessive increase in the demand for fresh perishable food. Perishable foods are a class of food (including meat, fish, milk, eggs, and many raw fruits and vegetables) with limited shelf life after harvest, processing, or slaughter [2]. The metabolism timeline, although inherent, is greatly influenced by the immediate surroundings, whether they are temperature, light, humidity, atmospheric gases or bacteria. High spoilage rates resulting from the shelf life limitation can lead to large economic and environmental losses. This has required the industry to develop new and improved methods for extending the shelf life of food while maintaining food quality.

Packaging materials for perishable food play a major role in our society through their increasing functionality. They offer appropriate protection to food and an increased convenience towards the consumers. The packaging process alters the spoilage evolution of food, eventually leading to a longer shelf life. The huge variety in both packaging materials and packaging methods, however, leads to a range of processes, which occur in the packaged food. Currently, there is a lack of understanding of these processes that eventually lead to the production of a variety of volatile organic compounds (VOCs) and, subsequently, the spoilage of the food. This lack of understanding prevents the food industry from fully exploiting new packaging materials and methods.

Generally, the implementation of optimal storage conditions through modified atmospheres has been effective in reducing microbial spoilage processes and maximizing the shelf life, while preserving the quality of food [3, 4. This procedure is known as modified atmosphere packaging (MAP). The containment of the modified
atmosphere within the package is crucial; however, quality assessment based on sensory, microbial, or chemical analyses is not possible except by opening the package. Unfortunately, measurement with destructive gas analysis equipment cannot be used to check every single package. Therefore, both the food industries as well as the distribution chains need to adapt their logistics and sampling plans to predict product shelf life expectancies. In addition, large product and economic losses occur while using the destructive sampling. In 2011, the Food and Agriculture Organisation of the United Nations estimated that approximately 1 billion tons of food is wasted every year: a third of the global food production [5]. In the EU up to $39 \%$ of the wasted food occur during processing and quality assurance measures [6]. These staggering numbers speak for themselves: our food system is failing.

There is an increasing need for a technology that would provide a tool for fast, accurate and non-destructive analysis of the quality status of packaged food. This technology should be able to detect and measure the concentrations of a range of different VOCs that are representative of spoilage processes that occur in packaged food.

Intelligent packaging can be used, in principle, to inform all actors in the chain, like wholesalers, retailers and consumers, about the quality status of food and offers the possibility to take logistic actions based on dynamically estimated shelf life, thereby reducing waste of food. In recent years, there have been several food packaging innovations known as intelligent and active packaging, which monitor the quality of packaged food [7]. When used exhaustively, they enable the monitoring of all packages, which on the short run diminishes and eradicates sampling plans, and on the long run predicts actual expiry dates, increases the margin of food safety, indicates freshness level, and detects early spoilage.

Normally, food samples are examined and rated by sensory evaluation. Assuming that intelligent food packaging will have the ability to describe the sensory properties of food, however, it should be emphasized that sensory evaluations are still essential when it comes to ensuring that the foods being produced are acceptable to the consumer. Provided that the instrument performs according to the required sensory properties, accuracy, and reproducibility, it could be applied to partly replace a sensory panel in the industry [8]. However, it has been shown that both sensory evaluation and intelligent food packaging are considered, in some cases, to be complementary to each other 9.

Interestingly, intelligent food packaging technology is heading towards replacing human decision-making, however, that would give rise to liability questions. The use of machine learning in intelligent food packaging for determining food quality has the potential for causing physical harm. Therefore, the liability question is complicated by uncertainties as to the standard of reliability, performance and accuracy which the hardware (i.e., sensor) and software (i.e., machine learning) in
the intelligent food package are expected to achieve. For instance, if the intelligent food package fails to detect food spoilage or food diseases, the most obvious response of consumers is to sue the producer of the food product or the producer of the intelligent food package. Therefore, as a start in the implementation of intelligent food packaging technology, users (i.e., businesses and consumers) should be made aware that this technology is merely for assisting in the decision-making process and not replacing it.

A Flemish strategic basic research project called CheckPack, which was launched in November 2013 (as described in [10]), focused on the research and development of a silicon photonics-based chemical sensor with a small footprint sensing region to measure concentrations of multiple VOCs and CO2 in the headspace of food packages. As a research partner in this project, our aim was to develop mathematical models for machine learning to be implemented alongside the sensor to identify the internal situation of the package and predict spoilage of food. To develop mathematical models of chemical sensors in intelligent food packaging, a detailed knowledge of the structure of chemical sensors is essential, and includes the understanding of the following three phases:

1. translating transducer signals (of the sensor into receptor signals),
2. translating receptor signals (into gas quantities in the package), and
3. translating gas quantities (into a label reflecting the quality of food).

A representation of the sensor concept and the three modelling phases are illustrated in Figure 1.1.

It has proven difficult at this stage of the technology to develop a silicon photonicsbased chemical sensor with (microporous and mesoporous) coatings that are highly selective of (most of) the spoilage-specific VOCs present in the headspace of most food packages. As a result, the plan of building mathematical models (in phase 1) to translate transducer signals of the sensor into receptor signals in the first phase was abandoned. Due to the unavailability of the (receptor signals of the) sensor, it seemed natural to divide phase 2 into two parts: (a) translating receptor signals into adsorbed quantities on the chemical coating, and (b) translating adsorbed quantities into gas quantities in the package. Clearly, phase 2(a) was abandoned since it requires a sensor. The problem defined in phase 2(b) has been known to be one of the most difficult challenges in the field of materials sciences and chemical engineering [11, especially when dealing with a very large number of VOCs (at least 50) in a food package. The difficulty lies in studying the interaction of the VOCs during adsorption on a chemical coating (i.e., the inverse of phase 2(b)). This phenomenon is called multi-component adsorption [12, 13]. Typically, multicomponent adsorption is studied for a very small number of compounds (up to five) on a chemical coating that is (at best) selective to a specific group of compounds. Currently, there exist very few chemical coatings that are highly selective, most

Figure 1.1: An illustrative concept of a silicon photonics-based chemical sensor with a matrix of selective coating materials and the three modelling stages.

of which are selective to specific groups of compounds, and a very few of which are selective to single compounds, which are not relevant to food packaging. In addition, the sensitivity of the current chemical coatings is very low: only high concentrations (parts per million) can be detected, whereas the concentrations of the compounds found in packages were mostly in low concentrations (parts per billion). Therefore, gathering data through experiments proved to be impossible in the entire project period. In an attempt to find an alternative approach, adosprtion data of compounds on chemical coatings were gathered from online databases, such as NIST Chemistry WebBook. Consequently, quantitative structure-activity relationship (QSAR) models [14, 15] were developed to predict adsorption characteristics of a specific compound on a specific coating. However, the developed models did not reflect accuracy nor precision, since the gathered data needed to be pre-processed, which was not feasible in the remaining project period. As a result, phase 2(b) was not continued.

Finally, the research was restricted to the modelling process of translating quantities of different types of gases in the headspace (phase 3) into a label/score that reflects the status of food. This modelling is independent of the type of receptor and transducer incorporated in the chemical sensor. One implication of this restriction was realizing existing problems in sensory evaluation of food. As a result, the concept of improving the quality of the assessment of a sensory evaluation was
born.

### 1.2. Problem setting

Sensory evaluation has seen an increase in the development of its science and application over the past years [16. Today, in the field of food science and in the food and beverage industry, sensory evaluation is a core resource for gathering valuable information about the quality of food. As a result, numerous research studies have focused on different types of sensory data, namely, absolute evaluations, in the form of ordinal labels, and relative evaluations, in the form of rankings.

In this dissertation, we develop methods to combine absolute and relative evaluations for determining the quality of a (food) sample. As a first step, we focus on the analysis of these different types of data to determine the overall freshness of a sample. To analyse such data, an analyst typically has a variety of traditional statistical methods at their ${ }^{1}$ disposal. Interestingly, these statistical methods depend on the nature of the data and their collection. The problem arises that most statistical methods do not fully exploit all the available information. In addition, it is often the case that analysts deal with a limited amount of data that entails inadequate statistical analysis. In this dissertation, we will develop data-analysis techniques that can be used to provide an answer to such problems. Firstly, we present an approach to reach consensus evaluations $\xi^{2}$ (ordinal labels) of multiple samples, simultaneously. Secondly, we propose to combine ordinal labels and rankings to improve the consensus label.

The next logical step is to utilize additional measurements such as instrumental data. The main instrumental evaluation technique dealt with in this dissertation applies to generating chemical data in analytical chemistry experiments and investigations. In different packaged foods, a range of different VOCs are representative for spoilage processes that could occur. Using instrumental measuring systems during analytical chemistry experiments, the values of these VOCs can be quantified. In this dissertation, we build a bridge between instrumental data and sensory data with the aim of solving the problem of predicting sensory evaluations of new samples. Firstly, we present an approach to build an ordinal regression model that can better predict ordinal labels in settings where data is limited. Secondly, we propose to integrate rankings in ordinal regression models to further improve the prediction of ordinal labels.

[^0]As a result, we present a question that this dissertation revolves around:
How can the consensus and predicted ordinal label of a (food) sample be improved by integrating rankings?

### 1.3. The structure of this dissertation

This dissertation is organized as follows: one introductory part (Part $\mathbb{I}$ ), three main parts (Parts II, III and IV), and one concluding part (PartV). This structure is illustrated in Figure 1.2. In this section we provide a roadmap to reading this dissertation.

Part $\rrbracket$ contains two chapters to assist the reader in understanding the topics that this dissertation will cover. Chapter 2 begins with considering the different types of data and defines descriptions of data and their collection processes. Chapter 3 covers traditional mathematical techniques.

Part $\Pi$ consists of one chapter and contains the procedures for gathering the sensory and instrumental data used throughout this dissertation for the demonstration of applications in food science. Chapter 4 describes the different foods studied in this dissertation and delves into the techniques used to perform microbial, chemical, and sensory evaluations.

Parts III and IV contain the main contributions of this dissertation. Each of these parts focuses on a number of research objectives for the characterization of food quality. Therefore, these parts build upon the material presented in Chapters 2 and 4 for the demonstration of applications in food science. It should be noted that these parts can be read independently from each other.

Part III consists of three chapters covering the analysis of sensory data. Chapter 5 focuses on the analysis of absolute evaluation data in the form of ordinal labels and presents the most prominent methods for determining consensus labels. Chapter 6 continues with the analysis of relative evaluation data in the form of rankings and presents the most prominent methods for aggregating rankings. Chapter 7 builds upon the methods used in the previous two chapters to combine the two types of sensory data.

Part IV consists of three chapters covering the prediction of sensory evaluations of a food sample. Chapter 8 focuses on learning a predictive model from absolute evaluation data for the prediction of ordinal labels. Chapter 9 continues with learning a predictive model from relative evaluation data for the prediction of rankings. Chapter 10 focuses on learning a predictive model from absolute and relative evaluation data combined for the prediction of ordinal labels.


Figure 1.2: The structure of this dissertation.

Part $\bigvee$ provides general conclusions of the most important results in this dissertation and a discussion on future research directions.

### 1.4. A brief overview of this dissertation

### 1.4.1. Part I: Preliminaries

This part consists of two chapters to assist the reader in understanding the topics this dissertation will cover and answers the question:

Question I: How do we collect sensory data and instrumental data and how can we analyse these data?

Today, in the field of food science, sensory evaluation and chemical experiments are prominent in gathering valuable sensory and instrumental data about the quality of food. The nature of sensory and instrumental data are introduced in Chapter 2. As the data depend largely on the focus of the experiment and the particular measurement system, this chapter offers a description of several methods of sensory measurement, namely, discriminative, affective and descriptive methods, and discusses the different types of sensory data generated from using these methods for different purposes. In addition, typical sensory analysis methods are introduced, and the training requirement of panellists is briefly described. For instance, to be able to evalaute specific characteristics of foods, such as spoilage, sensory quality etc., trained panellists require extensive training, which is cost intensive and time consuming. Unlike trained panellists, untrained (or less trained) panellists may be not be well-equiped in accurately providing an evaluation describing food spoilage, however, they are less costly to gather and are capable of detecting differences among samples [21].

The next logical step after performing sensory evaluation of food is to understand what causes the quality of food to change. Microbial spoilage has been seen as the most important cause of spoilage in food [22] producing VOCs that result in the change of smell or flavour of food. Thus, instrumental evaluation techniques for quantifying these VOCs are essential. Chapter 2 introduces chemical testing techniques for measuring VOCs and discusses several ways in which chemical analysis can be used to help to interpret the results of sensory analysis. In short, this chapter answers the question:

Question I.1: What do sensory evaluation and instrumental evaluation unequivocally tell us?

The analysis methods introduced in Chapter 2 are borrowed from the field of mathematical optimization, which is introduced in Chapter 3. Mathematical (optimization) methods are the foundation for understanding the main contributions presented in this dissertation. Section 3.3 focuses on providing general definitions and properties found in the field of mathematical optimization and discusses solving convex optimization problems.

What follows the steps of analysing the different types of data is the use of instrumental data to build models for predicting the evaluation of a food sample based solely on the features of the sample. Building a model is not a standard recipe and requires the selection of ingredients (loss function, hypothesis space and a regularization parameter). Thus, different models can be built using statistical learning, which is introduced in Chapter 3. Statistical learning theories are the foundation for understanding the main contributions presented in Part IV of this dissertation. Readers who are familiar with basic notions of mathematical optimization theory may skip Section 3.3 (or consult this section whenever necessary). In the same way, being well experienced with statistical learning theory, readers may skip Section 3.4 In short, Chapter 3 answers the question:

Question I.2: How does statistical learning relate to mathematical optimization?

### 1.4.2. Part II; Gathering sensory and instrumental data

This part consists of one chapter and contains the procedures for gathering the sensory and instrumental data used throughout this dissertation for the demonstration of applications in food science, and answers the question:

Question II: Which foods were studied and what was collected?
Chapter 4 describes the different foods, namely, chicken breasts, Atlantic cod (Gadus morhua), Atlantic brown shrimp (Crangon crangon) and Atlantic salmon (Salmo salar), studied in this dissertation and delves into the techniques used to perform sensory evaluation and analytical chemical experiments. This chapter is one of the pillars of this dissertation and, thus, deserves its own part. For answering the different questions in this dissertation, a number of datasets were collected, some specific for one question, and some others useful for several questions. To give the reader a structured introduction to the data, we gather them in Chapter 4.

### 1.4.3. Part III: Analysis of sensory data

This part consists of three chapters covering the analysis of sensory data, in case no chemical information about the samples is collected. In this setting, panellists build (implicit) features in their mind, and, thus, we call this setting a featureless setting. Part III builds upon the preliminaries of Part $\square$ and the materials presented in Part $\Pi$ and focuses on answering the question:

Question III: How can we analyse sensory data to reach a consensus evaluation?

To determine the overall quality of a food sample, trained panellists are normally asked to provide absolute evaluations in the form of ordinal labels. The aim is to assign a consensus label that describes the overall quality of a food sample. Thus, the most prominent methods used for the aggregation of ordinal labels to reach a consensus label are introduced in Chapter 5. This chapter is an extended version of a manuscript published in Information Fusion [23], focusing on the analysis of ordinal labels and proposes a novel approach, based on the search for monotonicity, to assign (joint) consensus labels of multiple objects. Previously used techniques, such as the median and the mode, are compared to our novel approach on data sets from three real-world problems, and are then applied to analyse the sensory data gathered in Chapter 4 In short, Chapter 5 answers the question:

Question III.1: How can we assign (joint) consensus ordinal labels?
Unfortunately, trained panellists are limited in number and (in some cases) expensive to train. For this reason, there usually is a limited amount of data available to reach a consensus evaluation. It is thus quite common to invoke untrained panellists and to gather some additional information [24]. However, untrained panellists are obviously not as skilled as trained panellists, and might be unable to accurately evaluate a given sample. Since it is a conceptually easier task, untrained panellists are then just asked to rank different samples according to their personal appreciation. Untrained panellists may find the process of ranking a large number of samples burdensome. It has also been shown that methods for analyzing rankings on a large number of samples is difficult [25]. Therefore, a simple solution would be to ask untrained panellists to rank only a subset of the large number of samples 26]. In sensory analysis of food, due to olfactory fatigue of panellists, it is recommended to rank no more than six samples [24]. Thus, the aggregation of rankings to reach a consensus ranking is introduced in Chapter 6, where the most prominent methods date back to the 18 -th century. These methods include: the Borda count, the methods of Condorcet, and the method of Kemeny. However, we also consider the latest method that has been developed in-house, where Raúl Pérez-Fernández et al. 27] proposed to exploit the monotonicity of a representation of rankings provided by the panellists to reach a consensus ranking. The aforementioned methods are compared by applying them to analyse the sensory data gathered in Chapter 4 . In short, Chapter 6 answers the question:

Question III.2: How can we reach a consensus ranking?
However, rankings provide only relative information on the quality of food samples. Thus, it is more interesting to reach a consensus ordinal label that provides absolute information on the quality of a food sample. Therefore, we propose to combine ordinal labels and rankings provided by trained and untrained panellists, respectively, to assign an improved consensus ordinal label that describes the overall quality of a food sample. This novel approach is introduced in Chapter 7 by making a first attempt at combining the most prominent methods used in
the aggregation of ordinal labels (Chapter 5) and rankings (Chapter 6). This chapter is an extended version of a manuscript published in International Journal of Approximate Reasoning [28, which has also been published as a conference paper [29], and an extended version of a manuscript under revision. We first propose an approach for integrating rankings with labels, in the form of scores, to improve the quality assessment of the consensus score. We propose a second approach for integrating scores with rankings to improve the assessment of the consensus ranking. In addition, we propose a third approach for incorporating different types of (relative) information to improve the quality of the consensus score. The application of these methods is illustrated on the sensory data gathered in Chapter 4 , specifically on scoring and ranking data of salmon samples (Section 4.2.5). In short, Chapter 7 answers the research question:

Question III.3: How can we combine scores and rankings to reach an improved consensus evaluation?

As the aforementioned techniques are based on mathematical optimization methods, we suggest to the reader to consult Section 3.3 .

### 1.4.4. Part IV: Prediction of sensory evaluation

This part consists of three chapters covering the prediction of food quality and are parallel to the chapters in Part III. In addition to sensory data, instrumental data, in the form of chemical information of samples, are collected. In this setting, each sample is described by the quantities of the VOCs, which we refer to as features of the samples, and, thus, we call this a setting with features. Part IV builds upon the preliminaries of Part $\Pi$ and the materials presented in Part $I$ and focuses on answering the question:

## Question IV: How can we predict the evaluation of a sample?

In parallel to Chapter 5, where trained panellists are typically asked to provide absolute evaluations, in the form of ordinal labels, the aim now is to use sensory and instrumental data to build a model that would take the features of the studied samples as input and the ordinal labels as output. This model can then predict the ordinal label of a new sample based solely on its chemical information. However, as we have previously discussed, training and (subsequently) collecting information from panellists usually carries big expenses. For this reason, there usually is a limited amount of data available to learn a good statistical model. Chapter 8 introduces a novel recipe for building an ordinal regression model, while including a regularization parameter. This chapter originates from a manuscript published in Information Fusion [30]. A motivation for the selection of the ingredients of this recipe, including the loss function, hypothesis space and regularization parameter, is provided. The ordinal regression problem is illustrated on the sensory and
instrumental data gathered in Chapter 4 In short, Chapter 8 answers the research question:

Question IV.1: How can we predict an ordinal label?
In parallel to Chapter 6, where untrained (or less trained) panellists, who provide relative evaluations in the form of rankings, are typically used, the aim now is to use sensory and instrumental data to build a model that would take the features of the studied samples as input and the rankings as output. This model can then predict a ranking of new samples based solely on their chemical information. Chapter 9 introduces the pairwise approach, the most prominent approach for building a ranking model to predict preferences. The predicted preferences are then used to derive an associated ranking of the samples. The pairwise approach is illustrated on the sensory and instrumental data gathered in Chapter 4. In short, Chapter 9 answers the research question:

Question IV.2: How can we predict rankings?
In parallel to Chapter 7, where it is more interesting to study ordinal labels that provide absolute information (rather than just rankings that provide only relative information) on the quality of food samples, we propose a strategy of integrating rankings provided by untrained panellists to improve the prediction of an ordinal label of new samples based solely on their chemical information. Chapter 10 presents this novel strategy building upon the works in Chapters 8 and 9 for augmenting ordinal regression models. This chapter is an extended version of the manuscript published in Information Fusion [30. The proposed method is illustrated on the sensory and instrumental data gathered in Chapter 4. In short, Chapter 10 answers the research question:

Question IV.3: How can we combine ordinal labels and rankings to improve the prediction of an ordinal label?

As the aforementioned mathematical models are based on statistical learning theories (Section 3.4) and mathematical optimization (Section 3.3), we suggest to the reader to consult Chapter 3 .

## PART I

## PRELIMINARIES

## 2 The nature of sensory and instrumental data

## Table of Contents

| 2.1 | Introduction and overview |
| :--- | :--- |
| 2.2 | Sensory evaluation and data |
|  | 2.2 .1 |
|  | Introduction to sensory evaluation |
| 2.2 .2 | Sensory measurement and analysis |
|  | 2.2 .3 |
|  | Determination of overall food quality |
| 2.3 | Instrumental evaluation and data |
|  | 2.3 .1 |
|  | Microbial spoilage of food |
| 2.3 .2 | Instrumental techniques for quantifying VOCs |
| 2.3 .3 | Chemical analysis |

### 2.1. Introduction and overview

In this chapter we discuss mainly two issues: The first is "defining the nature of data": Quantitative, qualitative, subjective, objective, etc.; and the second is "describing the methods of data gathering": Discriminative, affective, descriptive, gas chromatography, etc.

We define the nature of sensory and instrumental data and highlight the different measurement systems for gathering the data, by answering the following question:

Question I.1: What do sensory evaluation and instrumental evaluation unequivocally tell us?

The question is answered in two parts: The first concerning sensory evaluation is answered in Section 2.2, and the second concerning instrumental evaluation is answered in Section 2.3. The benefits of sensory data and instrumental data are discussed below while briefly walking through typical analysis techniques.

### 2.2. Sensory evaluation and data

Sensory evaluation has been a traditional method for studying sensory properties of food and offers scientific approaches to obtain appropriate information about
sensory quality of food. The field of sensory evaluation has been comprehensively reviewed by Amerine et al. [24, Lawless et al. [1, Meilgaard et al. [31, Moskowitz et al. [32, and Stone and Sidel [16]. The mentioned works are recommended for readers who would like to learn more in detail about sensory evaluation.

### 2.2.1. Introduction to sensory evaluation

The main objective of sensory evaluation is to provide an input for decision making. Simply put, it is a scientific method that involves the principles and practices of evoking, measuring, analysing, and interpreting the responses of (food) samples as perceived through the five senses [16.

The first process in sensory evaluation is evoking. Sensory evaluation provides guidelines for minimizing biasing factors in the preparation and serving of samples under controlled conditions. For example, separate booths are normally used to separate panellists and avoid the influence of their judgements on each other. Another example is providing each panellist with samples in a different order to counteract for possible effects caused by the order of the samples. Moreover, samples are given random numbers to avoid the interaction of panellists' evaluations with the numbers.

The second process in sensory evaluation is (sensory) measurement. The measurements can be either quantitative or qualitative. Quantitative measurements are numeric in nature and can be either discrete, in the form of integers, or continuous, in the form of any numeric value. Qualitative measurements are non-numeric in nature and can either be categorical (or nominal), in the form of categories without order, or ordinal, in the form of categories with order. We will further discuss the different types of measurement and the data that can be generated in the following subsection.

The third process in sensory evaluation is (sensory) analysis. In sensory evaluation, proper analysis of the generated data is critical. Mathematical methods are commonly used to analyse sensory data. However, the different types of data require appropriate statistical analysis methods. As we are dealing with human perception, the data generated is often highly variable. Hence, without proper statistical data analysis techniques, reasonable judgements cannot always be made on the relation between the sample characteristics and the sensory response of the panellists. In the following subsection, we will further discuss the analyses of different types of sensory data.

Finally, the fourth process in sensory evaluation is the interpretation of results. Based on the generated data, the proper analyses, and the results, reasonable judgements must be made to draw conclusions. Conclusions involve the consideration of the applied methods, the limitations of the experiments, and the background and
contextual framework of the study.

### 2.2.2. Sensory measurement and analysis

Three major measurement tests are commonly used for the sensory evaluation of food: discriminative, affective, and descriptive tests [1]. These sensory tests have different goals and measure different variables. Thus, there are several aspects to the nature of data generated by these various sensory tests. We illustrate the different types of data that can be generated from sensory measurement in Figure 2.1 .


Training of panellists

Figure 2.1: The nature of sensory data and the training requirement of panellists.

## Discriminative tests

Discriminative tests are the simplest tests in sensory measurement. They are used when comparing two (or more) samples to determine whether there exists any perceivable difference or similarity between them. A typical example is the triangle test, where there are three samples, two of which are from the same batch or process, and panellists are asked choose the sample that is most different. An extension to the triangle test is the two-out-of-five test, where there are five samples, and panellists are asked to sort these samples into two groups. Another example
is the duo-trio test, where, similar to the triangle test, there are three samples, however, one is a reference that matches a second sample and the third is from a different batch or process, and panellists are asked to choose the sample that best matches the reference. A fourth popular example is the pairwise comparison test, where two samples are provided, and panellists are asked to choose which sample has a higher intensity of a given attribute. These four tests are summarized in Table 2.1.

The data generated from the discriminative method are generally discrete, more specifically, binary values. However, they can be interpreted as nominal data that represent different groups or categories. Since most of the tests force panellists to make a distinction, panellists may make mistakes due to chance. Thus, statistical methods based on binomial, chi square, or normal distributions allow us to determine whether the panellists were able to perceive a difference between the samples or the results were due to chance alone. These methods are known as significance tests [33].

## Affective tests

Affective tests are the most common tests in sensory measurement. They aim at measuring the degree of liking or disliking of a sample. These tests include: acceptance and ranking tests. In acceptability tests, typically a 9-point hedonic scale [34, 35] is used, where the points on the scale represent ordered categorical labels ranging from "like extremely" to "dislike extremely". Here, panellists are asked to assign a label to each sample.

In recent years, affective tests have been modified requiring panellists to evaluate samples according to the degree of fulfilment of specific criteria, such as perceived freshness. In ranking tests, the fact that a sample is ranked above another sample means that the former is perceived as fresher than the latter, and the fact that samples are tied means that they are perceived as equally fresh. In acceptance tests, ordered categorical labels can be expressed on an ordinal scale that represents an increasing or decreasing magnitude of freshness. Typical labels for these classes are "Spoiled" (SP), "Marginal" (M), "Satisfactory" (S), "Fresh" (F), and "Very Fresh" (VF), with the linear order: SP $\prec \mathrm{M} \prec \mathrm{S} \prec \mathrm{F} \prec \mathrm{VF}$.

In a ranking test, panellists are asked to order a number of samples (greater than two) based on their personal preference. Pairwise comparison can be seen as a ranking of two samples, however, in affective tests, ties are allowed and there is no forced choice. In some cases, panellists are asked to use an unlabelled ordinal scale where the samples are arranged in order of increasing or decreasing preference. These two affective tests are summarized in Table 2.1.

The data generated from affective tests are generally qualitative, specifically absolute evaluations in the form of ordinal labels and relative evaluations in the form of rankings. In the case of ordinal labels, t-tests or analysis of variance (ANOVA)

| Method | Test | Test type | Samples | Task |
| :---: | :---: | :---: | :---: | :---: |
|  | Triangle <br> Duo-trio <br> Pairwise comparison <br> Two-out-of-five | Oddity <br> Similarity <br> Forced choice <br> Sorting | $\begin{gathered} \mathrm{A}, \mathrm{~A}^{\prime}, \mathrm{B}\left(\text { or } \mathrm{A}, \mathrm{~B}, \mathrm{~B}^{\prime}\right) \\ \text { Ref-A (or Ref-B), } \mathrm{A}, \mathrm{~B} \\ \mathrm{~A}, \mathrm{~B} \\ \mathrm{~A}, \mathrm{~A}^{\prime}, \mathrm{B}, \mathrm{~B}^{\prime}, \mathrm{B}^{\prime \prime}\left(\mathrm{A}, \mathrm{~A}^{\prime}, \mathrm{A}^{\prime \prime}, \mathrm{B}, \mathrm{~B}^{\prime}\right) \end{gathered}$ | Choose the most different sample <br> Match sample to reference <br> Choose sample with highest given attribute <br> Sort samples into two groups |
|  | Ranking <br> Labelling/Scoring | Relative <br> Absolute | $\begin{aligned} & \mathrm{A}, \mathrm{~B}, \mathrm{C}, \mathrm{D}, \ldots \\ & \mathrm{~A}, \mathrm{~B}, \mathrm{C}, \mathrm{D}, \ldots \end{aligned}$ | Rank samples in a specific order (with ties) Assign a label or a score to the samples |

Table 2.1: Summary of the most common tests for sensory evaluation, as described in Sensory Evaluation of Food [1.

F-tests are used to test the variances among the means of the samples by assessing the variation within the samples, relative to the variation between the samples. It can be seen that the t-test is a special case of the ANOVA F-test when there are only two samples to compare.

In the case of rankings, the Friedman test [36] is used, in a similar way to the ANOVA F-test, to determine differences between samples, in which the ranks of the samples are used to test the variances among the means of the samples by assessing the variation within the samples in proportion to the variation between the samples.

## Descriptive tests

Descriptive tests are the most comprehensive and informative sensory tests. They aim to quantify perceived intensities of the sensory attributes of a sample. Generally, there are two main phases in descriptive tests: the first is qualitative in nature, where panellists identify attributes that can be used to describe the characteristics of samples, and the second is quantitative in nature, where panellists evaluate the intensity of these attributes by assigning scores. The most widely used methods combine these two phases into an integrated sensory method. The main methods are consensus profiling, descriptive profiling, and free choice profiling, and the main difference among these methods is whether the two phases are performed collectively or individually. Descriptive profiling is the most conventional method, which is a form of Quantitative Descriptive Analysis (QDA) [16], where the first phase is performed collectively, while the second phase is performed individually. In consensus profiling, both phases are performed collectively, however, in free-choice profiling, both phases are performed individually.

The collective processes in consensus profiling and descriptive profiling require reaching a consensus description/evaluation in the measurement process of sensory evaluation, in which panellists discuss and modify their personal description/evaluation in order to reach a consensus. This consensus reaching process is thus done by the panellists before and/or during the evaluation of samples. However, in this dissertation we discuss a consensus reaching process that is done by an analyst after the evaluation of samples has been performed by the panellists.

Descriptive testing is used to obtain qualitative descriptors and quantitative evaluations of a sample. Apart from assigning a score to the individual descriptors, some kind of "general scoring" is often desired. This may include: total intensity of color, aroma or flavour, general difference, hedonic score, etc. A typical example of a quantitative descriptive test is illustrated in Figure 2.2.

Qualitative descriptors are usually in the form of categorical labels, and quantitative evaluations are numerical and usually in the form of either continuous or discrete scores. Continuous scores are decimal values usually on interval or ratio scales. In both interval and ratio scales, the order and the exact distance between the


Figure 2.2: Example of rating seven descriptors on 5-point scoring scales
values is meaningful and known, however, only ratio scales have an absolute zero. Discrete scores are equally spaced non-decimal values on an interval or a discrete ratio scale, summarized as a cardinal scale.

Similar to the analysis of data generated from affective tests, the quantitative evaluations generated from descriptive tests are commonly analysed using t-tests on means for two samples, or analysis of variance F-tests followed by comparisons of means for more than two samples.

### 2.2.3. Determination of overall food quality

Describing the overall quality of food is the most common problem considered by researchers in food science. Typically, the evolution of quality of food is studied by storing food samples for a different number of days [37, 38, 39. To solve the problem of describing the result, a fundamental decision needs to be made whether the evaluation should be: a) quantitative, such as flavour intensity or firmness, or b) qualitative, such as liking, preference, or acceptability. It must be noted that a further type of evaluation should also be considered: c) subjective, such as focusing on consumer behaviour and psychology or d) objective, such as measuring one particular attribute of a food rather than overall quality of the product.

Quantitative data of numerical nature that are ultimately generated from descriptive tests will allow for straightforward analysis to determine the overall quality of a sample. Typically, panellists are asked to evaluate a sample and provide either discrete or continuous values, and the arithmetic mean is assigned as the overall quality of a sample. Other measures of central tendency, such as the median, which is the value that separates the lower half from the upper half of the data, and
the mode, which is the value that is mostly assigned by the panellists, are also considered. Quantitative data generated from descriptive tests can be considered as subjective evaluations, where the panellists provide their subjective judgement when describing a food sample. Typically, panellists are trained to use predefined descriptors in order to generate more objective evaluations.

We have to note that descriptive tests require extensive training of panellists over long periods of time, thus, often deterring researchers from using this method.

Consequently, affective tests, more specifically acceptance tests, have become an attractive alternative, in which training of panellists is required to a lesser extent. The panellists then are considered as "experts" in the evaluation of only one sensory attribute. However, since the data is in the form of ordinal labels, the relative distance between the labels is not always equal, thus, making it difficult to determine the overall quality of a sample. In the field of food science, such a scale is known as the Labeled Hedonic Scale [40]. We would like to bring the reader to the attention of recent studies that deal with non-uniform ordinal scales 41, 42, 43, 44, 45]. In these studies, the notion of ordinal proximity measure is introduced to deal with psychological proximities among linguistic terms of ordered qualitative scales. In the field of food science, the median and the mode are typically used. In some cases, the labels are identified with discrete values [46] (usually assumed to be equidistant) to assign the (rounded) arithmetic mean. This assumes the existence of a certain notion of distance between the labels.

Note that qualitative evaluations are typically considered to be subjective, when asking consumers to provide their personal opinion or judgement on their preference of a food sample. In cases where trained panellists are asked to provide a qualitative evaluation on an attribute of a food sample that has a major effect on quality, then such evaluations are considered to be objective.

Finally, in the case of rankings and discriminative tests, little to no training of panellists is required, making it easier to gather data. Unfortunately, the major weakness of these tests is that the resulting data provide no information on the degree of difference between the samples. These data are relative and an overall ranking or preference is assigned using methods based on the median, the mode and the mean.

Therefore, it is important for a researcher in the field of food science to first determine the type of data to be generated by sensory evaluation tests, and, subsequently, determine the type of sensory evaluation test.

Note that it can be seen in Figure 2.1 that more training of the panellists is required the more we move to the left of the diagram (i.e., when moving from qualitative data that are not ordered to quantitative data that are continuous). In other words, panellists providing an evaluation on a nominal scale require no training, whereas panellists providing a continuous score on an interval or ratio scale require the
most training.
Thus, it can be seen that describing the overall quality of food based on the results of the aforementioned tests is often based on the distribution of the frequencies of the sensory responses of the panellists. Typically, the frequencies are summarized using a histogram that clearly displays the frequency distribution of the responses and highlights any central tendency. In this problem setting, mathematical methods for determining the central tendency of the data are borrowed from statistics. The use of the above-mentioned methods will be illustrated in Part III.

Typically, sensory evaluation is constrained to an available budget. As a result, it is critical to keep in mind the time, cost and resources involved with training panellists and performing sensory evaluations. For instance, the cost of gathering a certain amount of data from panellists can be derived by considering the costs for training the panellists, processing the food samples and organizing the sensory evaluation tests. Furthermore, the time for generating the data can be derived by considering the time required for gathering and training the panellists, processing the food samples and performing sensory evaluation. As a result, a value can be derived on the cost and time for obtaining the desired data, which can be used for determining a budget. Therefore, the scale of experimentation needs to be balanced against the available budget, sensory resources and panellists.

### 2.3. Instrumental evaluation and data

So far, we have described sensory evaluation tests for studying sensory quality of food. Before we introduce mathematical methods for analysing sensory data, we first describe instrumental techniques for measuring spoilage compounds in food.

Microbial spoilage has been seen to be the most important cause of food spoilage [22]. It is defined as an ecological phenomenon in which spoilage-specific organisms (SSOs) prevail and produce VOCs. Even though microbial enzymes will not be discussed in this dissertation, it is important to understand the factors that influence microbial spoilage, and, thus result in the production of different VOCs.

### 2.3.1. Microbial spoilage of food

It is well established that the ecological determinants influencing microbial spoilage can be categorized into five main factors: intrinsic, processing, extrinsic, implicit, and emergent 47. It has been shown that the most important determinants are the intrinsic and extrinsic factors 48.

Intrinsic factors are the inherent physical, chemical and structural properties of the food, such as the SSOs, available substrates, water activity (or moisture content), nutrients, and pH . These parameters greatly influence the composition of the microbial population that will colonize the food. For instance, the microbial growth rate is dependent on water and the types of SSOs that will prevail each require a specific pH . In addition, different kinds and amounts of nutrients and available substrates act as a rich source of energy for microbial growth allowing certain SSOs to colonize the food and produce VOCs, subsequently, off-odours or off-flavours.

Extrinsic factors are environmental parameters that affect food, notably, temperature, humidity, and atmosphere composition. It is well known that temperatures below $-32^{\circ} \mathrm{C}$ cause bacteria to enter a dormant stage, and, thus, inhibit spoilage and temperatures between $2-4^{\circ} \mathrm{C}$ cause bacteria to gradually grow, and, thus, slow spoilage [48, 49]. Modified atmosphere packaging (MAP) is another effective way to inhibit microbial spoilag $\AA^{1}$, while preserving nutrients and other properties of food, such as odour, colour and texture 50, 51. MAP is a technique in which the composition of the packaging atmosphere is altered, using three most commonly used gasses: O2, CO2 and N2. To optimize the use of a MAP, an optimal gas composition and type of food is required. Carbon dioxide (CO2, E 290) has antimicrobial properties against some SSOs that delay the initiation of microbial spoilage. Oxygen (O2, E 948) is mainly important to preserve or enrich the color of food. Nitrogen (N2, E 941) is normally used as a filler gas, since it is inert and has minimal effects on the metabolic reactions in meat. This gas is also used to replace O 2 in MAP in order to create an anaerobic atmosphere and select for a more anaerobic microbial population.

The other three determinants are briefly described as follows: Processing factors refer to the physical or chemical treatments during production that change the characteristics of a sample. Implicit factors are the result of the initial development of microorganisms and their interactions together, such as antagonism and synergism, which affect the microbial activity in the food. Finally, emergent factors are the interactive factors on top of the expected individual action of the aforementioned factors.

To quantify the impact of these factors on the quality of food samples, microbiological analysis, chemical analysis and sensory evaluation have to be performed. Consequently, a correlation can be made between these factors and the instrumental and sensory evaluation of food samples. In the field of food science, research on the influence of intrinsic and extrinsic factors on microbiological growth, VOCs and sensory evaluation are limited [38, 52, 53, 54]. In these studies, the influence of individual factors was studied. For instance, a correlation was made between

[^1]intrinsic factors and microbiological parameters and sensory evaluation at a constant extrinsic factor (i.e, at a constant temperature). Unfortunately, research on the influence of all the factors combined on the quality of the food samples, as far as we know, is non-existent. The reason for this could be that a large number of experiments are required to study the different combinations of factors and their combined impacts on microbiological, chemical and sensory evaluations. This is time and cost intensive.

### 2.3.2. Instrumental techniques for quantifying VOCs

As previously discussed, VOCs are produced by SSOs, and when the concentrations of these compounds exceed certain levels, odours (and flavours) are often produced by the VOCs, resulting in an olfactory impact that is associated with the spoilage of food. Thus, the composition of the VOC profiles could be used to evaluate the quality of food samples. It is clear that different food samples have different VOCs. For instance, in most meats, the main groups of VOCs produced are alcohols, aldehydes, ketones, esters, volatile fatty acids, and sulfur compounds [55]. These groups of VOCs can be measured and quantified using chemical testing techniques.

The most widely used chemical testing technique is gas chromatography, such as flame ionization detector (GC-FID) and mass spectrometry (GC-MS), which has seen wide applications in food analysis [56, 57, 58]. However, GC is not very convenient, requiring calibration by highly skilled operators and having a rather slow chromatographic process. This has led to the development of simpler and faster analytical techniques that allow direct mass spectrometric measurement without chromatographic separation. One of the most well-known technique for direct analysis of volatile compounds in air is selected ion flow tube mass spectrometry (SIFT-MS) [59], which is based on the chemical ionization of gases using specific precursor (positive) ions $\left(\mathrm{H}_{3} \mathrm{O}^{+}, \mathrm{NO}^{+}\right.$, and $\left.\mathrm{O}_{2}^{+}\right)$. Hence, the quantification of gases by their reaction with precursor ions results in ionized masses. These ionized masses are monitored by a mass spectrometer, measuring the amplitude of the product ion signal and providing a quantitative measure for the amount of selected gases in the headspace. Recently, SIFT-MS has been described as a tool used for the determination of quality parameters in food by quantifiying the VOCs [60, 61, 62, 63]. Perhaps the most important advantage of SIFT-MS over other analytical techniques is that VOCs can be quantified in real time without external calibration.

Essential for SIFT-MS is the kinetics library of reactions of the precursor ions with a wide variety of compounds. This database provides the reaction rates and the (primary and secondary) ion products for the analysis of trace gas molecules. There are two basic types of scans used to acquire data in this analytical method: full
mass (FM) scans and selected ion mode (SIM) scans. FM scans are most commonly used to identify unknown compounds. Data from the scans can be matched against the kinetics library, identifying the compounds and their concentrations. Relative to FM scans, SIM scans target specific compounds by devoting more time to specific target ions. Thus, SIM scans offer a more sensitive analysis.

### 2.3.3. Chemical analysis

Interpreting the results of analysed sensory data is often a difficult task. Usually, when studying the evolution of the overall quality of food, samples from different origins of the same food (i.e., different chicken breasts, fish fillets, etc.) are not necessarily identical. For instance, one cannot simply presume that samples from different fish fillets that have been stored for the same number of days will be the same. Similarly, one cannot presume that a sample from one fillet is always preferred over a sample from another fillet that has been stored for longer. Therefore, interpreting the results of analysing sensory data cannot be done based solely on the storage day of the samples.

While sensory evaluation gives a subjective understanding of the overall quality of a food sample, instrumental evaluation gives an objective response. However, instrumental evaluation is only able to measure one aspect of a food sample. Thus, this may not always be sufficient to determine whether the quality of a food sample is acceptable. Both sensory and instrumental evaluation of food are essential in the food industry to routinely monitor the quality and ensure the acceptability of the produced food. Thus, it can be seen that these two methods of evaluation are complementary to each other. It has been shown that the combination of data from sensory and instrumental evaluation may provide insights into the chemical and physical properties that drive sensory attributes (such as appearance, odour, flavour, taste, and texture) 64. As odours (and flavours) are often produced by VOCs, the composition of the VOC profiles has been successfully used to evaluate the quality of food, such as seafood 65] and meats [66]. SIFT-MS has attracted the attention of many researchers for rapid and accurate characterization of VOCs and has been validated for fish metabolite research [61, 65].

As we have previously discussed, evaluations of samples, provided by panellists, are commonly analysed to determine the overall quality of these samples. As it can be understood, the production of VOCs in a sample results in off-odours and/or off-flavours affecting the assessment of the overall quality or freshness of the sample by panellists. Therefore, it would be interesting to study the impact of the evolution of the VOCs on the sensory evaluation of food samples. An effective way to interpret these evaluations is by chemical analysis of these samples. Thus, chemical analysis allows to establish a relation between the VOC profiles and the determined overall quality of samples. An effective way of doing this is by performing
hierarchical agglomerative clustering [67], a data analysis tool that merges similar groups of samples based on the similarity of their VOC profile, resulting in a dendrogram. Recently, there has been interest in determining the similarities and differences between VOC profiles by using clustering methods [68, 69]. However, in many real-world datasets, the clusters are not so well-defined. Typically, high similarity within a cluster is a criterion for the quality of clustering. Thus, several methods to measure the quality of clustering by using density, distances of samples within each cluster, the distance between the clusters themselves, etc., have been proposed 70, 71, 72, 73, 74.

Another way in which chemical analysis of samples can also help in the evaluation is to build mathematical models that can predict the overall quality of food. To do this, a relationship between the VOC profiles and the results of sensory analysis has to be determined. This problem setting involves the prediction of the quality of a sample based solely on the instrumental measurement of that sample. This problem can be classified as a prediction problem. More precisely, we are interested in constructing a function that maps data from a chemical measurement to a sensory evaluation. Stated differently, the food sample can be represented in two manners. Firstly, it can be represented by means of the VOC profile in the headspace of the food sample. Secondly, it can be represented by the evaluation of the sample by panellists. In this problem setting, mathematical methods for building models are borrowed from the field of statistical learning and will be covered in detail in the following chapter (Section 3.4).

Over the past years, mathematical methods have been developed to predict microbial spoilage. Notable examples include: predicting microbial spoilage based on spectral data [75, 76, predicting microbial spoilage based on concentrations of VOCs 77, 78, 79, 80, 81, 82, and predicting sensory evaluations based on spectroscopic data [83, 84]. However, research on learning a relationship between VOC profiles of food samples and sensory analysis is, as far as we know, limited to a few studies [85, 86]. In these studies, the models were developed to simply predict a score. We will discuss later on in Part IV the difficulties encountered while merging instrumental and sensory data, and predicting ordinal labels and rankings.

## 3 Mathematical prerequisites

## Table of Contents

| 3.1 | Introduction |
| :---: | :---: |
| 3.2 | Notations and mathematical conventions |
| 3.3 | Mathematical optimization methods |
|  | 3.3.1 Motivation |
|  | 3.3.2 General definitions |
|  | 3.3.3 Solving convex optimization problems |
|  | 3.3.4 Subclasses of convex optimization problems |
| 3.4 | Statistical learning theory |
|  | 3.4.1 The standard framework |
|  | 3.4.2 Estimation and hypothesis space |
|  | 3.4.3 Overfitting and regularization |
| 3.5 | Statistical learning as an optimization problem |
|  | 3.5.1 Regression problems |
|  | 3.5.2 Classification problems |

### 3.1. Introduction

This dissertation revolves around using mathematical optimization methods, which are briefly discussed in this chapter. The aim of this chapter is to serve as a general introduction to the field of mathematical optimization. By providing a collection of definitions and properties, this chapter may assist the reader to appreciate the contributions of this dissertation.

Mathematical optimization is the foundation of many statistical learning methods. More specifically, most statistical learning problems may be reduced to optimization problems. Indeed, statistical learning could be labelled as "optimzation enabaled learning". Mathematical optimization is at the heart of many engineering, scientific, economic, social and personal decision that is taken by an individual or a collective representation of people or machines.

In this chapter, we will answer the following question:
Question I.2: How does statistical learning relate to mathematical optimization?

### 3.2. Notations and mathematical conventions

Throughout this dissertation, we use consistent mathematical notations to formalize problems. In this section, we gather our notations and some mathematical facts and conventions.

## Sets

Sets are denoted as script capitalized characters. For example $\mathscr{X}$ will be used to denote a generic set. If $\mathscr{X}$ is a set and $x$ is an element of $\mathscr{X}$, then we write $x \in \mathscr{X}$. A set can be specified by $\mathscr{X}=\{x \mid x$ satisfies P$\}$, as the set of all elements $x$ satisfying property P . The set of real numbers (also referred to as scalars) is denoted by $\mathbb{R}$. Thus, we denote by $\mathscr{X} \subset \mathbb{R}^{n}$ that $\mathscr{X}$ is a proper subset of the $n$-dimensional real vector space. We denote by $[a, b]$ the set of real numbers $x$ satisfying $a \leq x \leq b$. An inverted square bracket denotes strict inequality in the definition. Thus, $] a, b]$, $[a, b[$, and $] a, b[$ denote the set of all $x$ satisfying $a<x \leq b$, $a \leq x<b$, and $a<x<b$, respectively.

When dealing with order relations in this dissertation, parentheses () are used for representing an ordered set of elements and curly brackets $\}$ are used for representing a set of elements without considering a specific order among them.

## Order relations

Here, we introduce some basic properties of order relations. We refer to the works of Birkhoff [87, Davey and Priestley [88, Roberts and Tesman 89], Ehrgott 90, and Anandalingam and Friesz 91 for more detailed information.

Binary relations are used in many branches of mathematics to describe whether a relation exists between two objects. For example, in arithmetic, binary relations are used to model concepts like "is greater than", "is equal to", "divides", etc.

Definition 3.1 (Binary relation). A binary relation $\mathcal{R}: \mathscr{A}^{2} \rightarrow\{0,1\}$ (on a set A) may satisfy the following properties:
(i) Reflexivity: $\forall a \in \mathscr{A}$, it holds that

$$
a \mathcal{R} a
$$

(ii) Irreflexivity: $\forall a \in \mathscr{A}$, it holds that

$$
\neg(a \mathcal{R} a)
$$

(iii) Transitivity: $\forall a, b, c \in \mathscr{A}$, it holds that

$$
(a \mathcal{R} b) \wedge(b \mathcal{R} c) \Rightarrow(a \mathcal{R} c)
$$

(iv) Antitransitivity: $\forall a, b, c \in \mathscr{A}$, it holds that

$$
(a \mathcal{R} b) \wedge(b \mathcal{R} c) \Rightarrow \neg(a \mathcal{R} c)
$$

(v) Symmetry: $\forall a, b \in \mathscr{A}$, it holds that

$$
(a \mathcal{R} b) \Rightarrow(b \mathcal{R} a)
$$

(vi) Asymmetry: $\forall a, b \in \mathscr{A}$, it holds that

$$
(a \mathcal{R} b) \Rightarrow \neg(b \mathcal{R} a) .
$$

(vii) Antisymmetry: $\forall a, b \in \mathscr{A}$, it holds that

$$
(a \mathcal{R} b) \wedge(b \mathcal{R} a) \Rightarrow a=b
$$

(viii) Completeness: $\forall a, b \in \mathscr{A}$, it holds that

$$
(a \mathcal{R} b) \vee(b \mathcal{R} a)
$$

Definition 3.2 (Order relations). A binary relation (on a set $\mathscr{A}$ ) is an order relation if it satisfies reflexivity, antisymmetry, and transitivity; a strict order relation is the irreflexive part of an order relation; a total order relation is an order relation that satisfies completeness; a binary relation is called a weak order relation if it satisfies transitivity and completeness (and, consequently, reflexivity); a binary relation is called a strict weak order relation if it satisfies transitivity and irreflexivity (and, consequently, asymmetry).

An example of an order relation is the subset relation $\subseteq$. An example of a strict order relation is the relation $<$ on a set of integers. An example of a total order relation is the relation $\leq$ on a set of integers. Finally, an example of a strict weak order relation is a relation $\prec$ that compares multiple objects to determine which object precedes the other. The complement of the strict weak order relation $\prec$ is a weak order relation $\precsim$, and vice versa.

Note that in this dissertation we use the term "ranking with ties" to refer to a weak order and the term "ranking (without ties)" to refer to a strict weak order.

Normally, a Hasse diagram is used to comprehend and view an order relation on a finite set. Intuitively, a Hasse diagram is a graphical representation of an order relation on a set in which the vertices of the graph represent the elements in the set, and an edge between two vertices represents a covering relation between comparable elements in the set. Typically, the element at the top is greater than the element at the bottom.

For example, consider the set $A=\{a, b, c\}$ and the relations $a<c$ and $b<c$. The resulting Hasse diagram of the order relation $<$ on set $A$ is shown in Fig. 3.1.


Figure 3.1: Example of a Hasse diagram of the order relation $<$ on set $A$.

## Vectors and matrices

Vectors are denoted as boldface characters and are assumed to be row vectors. For example, $\mathbf{x} \in \mathbb{R}^{n}$ represents a $1 \times n$ row vector, where the $i$-th element of the vector is denoted $x_{i}$ or $\mathbf{x}(i)$ (these notations are used interchangeably). A (row) vector can be specified in the form $\mathbf{x}=\left(x_{1}, \ldots, x_{n}\right)$. The transpose of vector $\mathbf{x}$ is denoted $\mathbf{x}^{\top}$. We denote by $\|\mathbf{x}\|_{1}=\sum_{i=1}^{n}\left|x_{i}\right|$ and by $\|\mathbf{x}\|_{2}=\sqrt{\mathbf{x} \cdot \mathbf{x}}{ }^{1}$ the $\ell_{1}$-norm and $\ell_{2}$-norm of $\mathbf{x}$, respectively. Vectors can be indexed. For example, $\mathbf{x}_{1} \in \mathbb{R}^{n}$ and $\mathbf{x}_{2} \in \mathbb{R}^{n}$ indicate that both $\mathbf{x}_{1}$ and $\mathbf{x}_{2}$ are $1 \times n$ vectors. Moreover, the $j$-th element of a vector $\mathbf{x}_{i} \in \mathbb{R}^{n}$ is denoted $x_{i j}$ and can be described as the element in the $i$-th row and $j$-th column of a matrix.

Matrices are denoted as boldface capitalized characters. For example, $\mathbf{X} \in \mathbb{R}^{m \times n}$ is an $m \times n$ matrix. The transpose of $\mathbf{X}$ is denoted by $\mathbf{X}^{\top}$. For a square $n \times n$ matrix $\mathbf{X}, \mathbf{X}^{-1}$ denotes the inverse of $\mathbf{X}$ (if it exists). The element of a matrix $\mathbf{X}$ in the $i$-th row and $j$-th column is denoted $x_{i j}$ or $\mathbf{X}_{i j}$ (these notations are used interchangeably). Matrices are indexed by using a superindex. For example, $\mathbf{X}^{1} \in \mathbb{R}^{m \times n}$ and $\mathbf{X}^{2} \in \mathbb{R}^{m \times n}$ are both $m \times n$ matrices.

Vectors $\mathbf{0}_{n}$ and $\mathbf{1}_{n}$ denote an $1 \times n$ vector of zeros and ones, respectively, and the matrix $\mathbf{I}_{n}$ is the $n \times n$ identity matrix.

Given two vectors $\mathbf{x}_{1}, \mathbf{x}_{2} \in \mathbb{R}^{n}$, we write $\mathbf{x}_{1}=\mathbf{x}_{2}$ if $x_{1 j}=x_{2 j}$ for any $j \in\{1, \ldots, n\}$. Moreover, we write $\mathbf{x}_{1} \leq \mathbf{x}_{2}$ if $x_{1 j} \leq x_{2 j}$ for any $j \in\{1, \ldots, n\}$. Note that $\mathbf{x}_{1}<\mathbf{x}_{2}$ means that $\mathbf{x}_{1} \leq \mathbf{x}_{2}$ and that there is at least one index $j$ for which $x_{1 j}<x_{2 j}$.

Random vectors (or multivariate random variables) are denoted as boldface italic capitalised characters. For example, $\mathcal{X}$ is a random vector that consists of realvalued random variables. In this dissertation, we will be mostly using notations like $\operatorname{Pr}(\mathcal{X} \leq \mathbf{x})$ to denote the probability that the random vector $\mathcal{X}$ takes a value that is smaller than or equal to $\mathbf{x} \in \mathbb{R}^{n}$.

## Functions

[^2]A function $f$ is specified in the form $f: \mathscr{X} \rightarrow \mathscr{Y}$ to indicate that $f$ is defined on a nonempty set $\mathscr{X}$ and takes values in a set $\mathscr{Y}$. We denote by $\operatorname{dom}(f)$ and $\operatorname{codom}(f)$ the domain and co-domain of $f$, respectively. We consider that $\mathscr{X} \subseteq \mathbb{R}^{n}$, and we use $\bar{X}$ for the closure of a set $\mathscr{X}$ (i.e., all the points that are either in or "near" $\mathscr{X}$ ) and $\mathscr{X}^{c}$ for the complement of a set $\mathscr{X}$ (i.e., points not in $\mathscr{X}$ ). Thus, we use $\mathbf{b d}(\mathscr{X})=\overline{\mathscr{X}} \cap \overline{X^{c}}$ for the boundary of a set $\mathscr{X}$. We denote the interior of the domain of $f$ as $\operatorname{int}(\mathscr{X})=\mathscr{X} \backslash \mathbf{b d}(\mathscr{X})$. Note that the boundary can also be written as $\operatorname{bd}(\mathscr{X})=\bar{X} \backslash \operatorname{int}(\mathscr{X})$.

The gradient vector of $f$ is denoted $\nabla f$ (i.e., the collection of all the partial derivatives of $f$ in a vector) and the Hessian matrix $\nabla^{2} f$ (i.e., the square matrix of second-order partial derivatives of $f$ ). Functions that are $\mathbb{R}^{m}$-valued (i.e., the co-domain of the functions are subsets of $\mathbb{R}^{m}$ ) for $m>1$ are denoted as boldface characters. For example, $\mathbf{f}(\mathbf{x}): \mathscr{X} \rightarrow \mathbb{R}^{m}$ represents the function $\mathbf{f}(\mathbf{x})=$ $\left(f_{1}(\mathbf{x}), \ldots, f_{m}(\mathbf{x})\right)^{\top}$.

### 3.3. Mathematical optimization methods

### 3.3.1. Motivation

We will start with a motivation for mathematical optimization. Optimization problems are ubiquitous in statistics and machine learning. They appear throughout the two disciplines in terms of the estimators we are trying to form, the problems we are trying to solve, and so forth. We are often taught in statistics and machine learning about how a certain conceptual idea translates into a particular optimization problem.

Numerous problems in engineering, statistics, machine learning, etc., involve solving an optimization problem. For example, given that we have been gathering sensory evaluation data of several food samples, we are now provided with a new food sample and asked to find the overall quality of this new food sample. This can be seen as fitting a function (e.g. a linear function) to our gathered sensory evaluation data, and using this function to predict an output (in this case the overall quality) given an input (the new food sample). Now, let us suppose we want to impose some kind of structure on the coefficients of that function. For instance, we want the coefficients to be sparse or shrunken. We normally learn how to take that conceptual idea and translate it into an optimization problem. Another example would be the problem where we are given chemical data and are asked to determine similarities between the different VOCs. This can be seen as a problem of fitting the best hyperplanes with the biggest margins separating the different classes that the VOCs belong to. In this setting, we normally learn how that idea translates into a particular problem that we want to solve to achieve that goal.

The main motivation for solving an optimization problem can be reduced down to two major points:

1. Different algorithms for solving particular problems can perform better or worse depending on the "flavour" of the problem. It can even be the same problem we are trying to solve, but with a different instantiation of the data (i.e., the data comes from a different distribution or of a different nature, however, the problem is the same). It could also be that the problem we are solving is very similar, in appearance, to another problem, but each problem requires a different effective algorithm.
2. Understanding how to solve a problem and things about its nature, in terms of optimality and different views that we can take from the perspective of convex analysis, can actually lend a deep understanding to the statistical side of the problem (i.e., the conceptual side).

### 3.3.2. General definitions

This section is written as a compilation of several definitions and properties that are found in the field of mathematical optimization. Most of these definitions are key to understanding mathematical optimization, and will be used frequently in this dissertation. These definitions can be found more in detail in the works of Rockafellar [92, Boyd and Vandenberghe 93] and Nocedal and Wright 94.

As a starting point, we define a mathematical optimization problem. We use the following notation:

- $X$ is a set called the domain of the optimization problem;
- $\mathbf{x}=\left(x_{1}, \ldots, x_{n}\right)^{\top}$ is a vector called the optimization variable;
- $f$ is the objective function that we want to minimize or maximize;
- $\mathbf{g}(\mathbf{x}) \leq \mathbf{b}$ is an inequality constraint;
- $\mathbf{h}(\mathbf{x})=\mathbf{A x}-\mathbf{b}$ is an equality constraint

The inequality constraint function $\mathbf{g}(\mathbf{x})$ is a function of $\mathbf{x}$ that defines certain inequalities that the vector $\mathbf{x}$ must satisfy. The equality constraint function $\mathbf{h}(\mathbf{x})$ is defined as $\mathbf{h}(\mathbf{x})=\mathbf{A x}-\mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{q \times n}$ is a matrix and $\mathbf{b} \in \mathbb{R}^{n}$ is a vector. Using this notation, we define a mathematical optimization problem.

Definition 3.3 (Mathematical optimization problem). Consider a set $\mathcal{X} \subseteq \mathbb{R}^{n}$, the functions $f: \mathscr{X} \rightarrow \mathbb{R}, \mathbf{g}: \mathcal{X} \rightarrow \mathbb{R}^{p}$ and $\mathbf{h}: \mathscr{X} \rightarrow \mathbb{R}^{q}$. A (mathematical) optimization problem is the problem of finding an element $\mathbf{x} \in \mathscr{X}$ that minimizes the function $f$ and satisfies the inequality constraint $\mathbf{g}(\mathbf{x}) \leq \mathbf{b}$ and the equality constraint $\mathbf{h}(\mathbf{x})=\mathbf{A x}-\mathbf{b}$. We denote such a problem as:

$$
\begin{array}{ll}
\underset{\mathbf{x} \in \mathscr{X}}{\operatorname{minimize}} & f(\mathbf{x}) \\
\text { subject to } & \mathbf{g}(\mathbf{x}) \leq \mathbf{b} \\
& \mathbf{h}(\mathbf{x})=\mathbf{A} \mathbf{x}-\mathbf{b},
\end{array}
$$

The solution of an optimization problem is called an optimum, which we refer to as a minimizer throughout this dissertation. A vector $\mathbf{x}^{*}$ is called a minimizer if for any $\mathbf{x}$ satisfying $\mathbf{g}(\mathbf{x}) \leq \mathbf{b}$ and $\mathbf{h}(\mathbf{x})=\mathbf{A} \mathbf{x}-\mathbf{b}$, we have $f\left(\mathbf{x}^{*}\right) \leq f(\mathbf{x})$. In other words, $\mathbf{x}^{*}$ is a solution if it respects the inequality constraints and has the smallest objective function value.

In this section, we will always consider $\mathscr{X} \subseteq \mathbb{R}^{n}$. We start off by defining convex sets.

Definition 3.4 (Convex set). $A$ set $\mathscr{X} \subseteq \mathbb{R}^{n}$ is convex if the line segment between any two points $\mathbf{x}_{1}, \mathbf{x}_{2} \in \mathcal{X}$ is made by any scalar $\alpha \in[0,1]$ that lies in $\mathcal{X}$. This is denoted as:

$$
\alpha \mathbf{x}_{1}+(1-\alpha) \mathbf{x}_{2} \in \mathscr{X}
$$

A convex combination is considered as a special type of linear combination. More precisely, it is a combination of points of the form $\sum_{i=1}^{k} \alpha_{i} \mathbf{x}_{i}$, where $\mathbf{x}_{1}, \ldots, \mathbf{x}_{k} \in$ $\mathbb{R}^{n}$, such that $\alpha_{i} \geq 0$, for $i \in\{1, \ldots, k\}$ and $\sum_{i=1}^{k} \alpha_{i}=1$. As can be expected, a set $\mathscr{X}$ is then convex if and only if it contains every convex combination of its points.

Definition 3.5 (Convex hull). For a set $\mathscr{X} \subseteq \mathbb{R}^{n}$, let $\operatorname{conv}(\mathscr{X})$ be the convex hull of the set $\mathscr{X}$. This is denoted as:

$$
\operatorname{conv}(\mathscr{X})=\left\{\sum_{i=1}^{k} \alpha_{i} \mathbf{x}_{i} \mid \mathbf{x}_{i} \in \mathscr{X}, \alpha_{i} \geq 0, i \in\{1, \ldots, k\} ; \sum_{i=1}^{k} \alpha_{i}=1\right\}
$$

The convex hull $\operatorname{conv}(\mathscr{X})$ is the smallest convex set that contains $\mathscr{X}$.

## Convex functions

Convex sets are considered the building blocks of convex functions. Convex functions are essential in the field of mathematical optimization. For convex functions, we will assume that $\operatorname{dom}(f)=\mathscr{X}$ is a convex subset of $\mathbb{R}^{n}$.

Definition 3.6 (Convex function). Let $\mathscr{X}$ be a convex subset of $\mathbb{R}^{n}$. A function $f: \mathscr{X} \rightarrow \mathbb{R}$ is called convex if for all $\mathbf{x}_{1}, \mathbf{x}_{2} \in \mathscr{X}$ and $\alpha \in[0,1]$ we have that

$$
f\left(\alpha \mathbf{x}_{1}+(1-\alpha) \mathbf{x}_{2}\right) \leq \alpha f\left(\mathbf{x}_{1}\right)+(1-\alpha) f\left(\mathbf{x}_{2}\right) .
$$

Definition 3.7 (Concave function). Let $\mathscr{X}$ be a convex subset of $\mathbb{R}^{n}$. A function $f: \mathscr{X} \rightarrow \mathbb{R}$ is called concave if $-f$ is convex.

There are two important modifiers in terms of convex functions: strictly convex functions and strongly convex functions.

Definition 3.8 (Strictly convex function). Let $\mathscr{X}$ be a convex subset of $\mathbb{R}^{n}$. $A$ function $f: \mathscr{X} \rightarrow \mathbb{R}$ is called strictly convex if for all $\mathbf{x}_{1}, \mathbf{x}_{2} \in \mathscr{X}$ and $\left.\alpha \in\right] 0,1[$ we have that

$$
f\left(\alpha \mathbf{x}_{1}+(1-\alpha) \mathbf{x}_{2}\right)<\alpha f\left(\mathbf{x}_{1}\right)+(1-\alpha) f\left(\mathbf{x}_{2}\right)
$$

Loosely speaking, a strictly convex function is more convex (i.e., has a greater curvature) than a linear or affine function.

Definition 3.9 (Strongly convex function). Let $\mathscr{X}$ be a convex subset of $\mathbb{R}^{n}$. $A$ function $f: \mathscr{X} \rightarrow \mathbb{R}$ is called strongly convex if there exists a parameter $m>0$, which means that $f-\frac{m}{2}\|\mathbf{x}\|_{2}^{2}$ is convex. Equivalently, $f$ is strongly convex if there exists $m>0$ such that for all $\mathbf{x}_{1}, \mathbf{x}_{2} \in \mathscr{X}$ and $\alpha \in[0,1]$, we have that

$$
f\left(\alpha \mathbf{x}_{1}+(1-\alpha) \mathbf{x}_{2}\right) \leq \alpha f\left(\mathbf{x}_{1}\right)+(1-\alpha) f\left(\mathbf{x}_{2}\right)-\frac{m}{2} \alpha(1-\alpha)\left\|\mathbf{x}_{1}-\mathbf{x}_{2}\right\|_{2}^{2}
$$

Meaning that a strongly convex function is at least as convex as a quadratic function. The motivation for studying strong convexity is that it allows us to prove fast convergence rates of certain algorithms. Clearly, there is a hierarchy that strong convexity implies strict convexity, which in turn implies convexity. However, the converse is not generally true.

The definitions of the modifiers are analogous for concave functions. One would simply substitute the convex function $f$ with its negative as shown in Definition 3.7

## Properties of convex functions

A key property of convex functions states that a function is convex if and only if its restriction to any line is convex. The converse is always true.
Definition 3.10 (Epigraph). Let $\mathscr{X}$ be a convex subset of $\mathbb{R}^{n}$. The epigraph of a given function $f: \mathscr{X} \rightarrow \mathbb{R}^{n}$ is defined as

$$
\operatorname{epi}(f)=\{(\mathbf{x}, \alpha) \in \mathscr{X} \times \mathbb{R} \mid f(\mathbf{x}) \leq \alpha\}
$$

The function $f$ is convex if and only if epi $(f)$ is a convex set.
Definition 3.11 (Sublevel set). Let $f: \mathscr{X} \rightarrow \mathbb{R}^{n}$ be a convex function and $\alpha \in \mathbb{R}$ be a scalar. The sublevel set of $f$ is defined as

$$
\{\mathbf{x} \in \mathscr{X} \mid f(\mathbf{x}) \leq \alpha\}
$$

If $f$ is convex, then its sublevel set is convex for all $\alpha \in \mathbb{R}$. The converse is not
necessarily true.
When a function $f: \mathscr{X} \rightarrow \mathbb{R}$ is differentiable over its domain, $\mathscr{X}$ is convex, and $f\left(\mathbf{x}_{2}\right) \geq f\left(\mathbf{x}_{1}\right)+\nabla f(\mathbf{x}) \cdot\left(\mathbf{x}_{2}-\mathbf{x}_{1}\right), \forall \mathbf{x}_{1}, \mathbf{x}_{2} \in \mathscr{X}$, then we have that $f$ is convex. Therefore, for a differentiable convex function, it holds that $\nabla f(\mathbf{x})=0$ if and only if $\mathbf{x}$ minimizes $f$.

Similarly, if $f$ is twice differentiable, convex and $\nabla^{2} f(\mathbf{x}) \geq 0$ (the Hessian is positive semi-definite) for every $\mathbf{x} \in \mathscr{X}$, then $f$ is convex.

### 3.3.3. Solving convex optimization problems

As a starting point, we define a convex optimization problem. Both constraint functions $\mathbf{g}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$ are functions of $\mathbf{x}$ and define certain inequalities and equations that the vector $\mathbf{x}$ must satisfy. If there are no constraints (i.e., $p=q=0$ ) the optimization problem is called unconstrained. Formally, a point $\mathrm{x} \in \mathcal{X}$ is said to be feasible if it satisfies the equality constraint $\mathbf{h}(\mathbf{x})=\mathbf{0}_{q}$ and inequality constraint $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}_{p}$. Thus, a convex optimization problem is said to be feasible if there exists at least one feasible point, and infeasible otherwise. Moreover, the set of all feasible points is called the feasible set. Using this notation, the convex optimization problem can be defined as follows:

Definition 3.12 (Convex optimization problem). A convex optimization problem is the problem of finding an element $\mathbf{x} \in \mathscr{X}$ that minimizes the function $f$ and satisfies the inequality constraint $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}_{p}$ and the equality constraint $\mathbf{h}(\mathbf{x})=\mathbf{0}_{q}$. We denote such a problem as:

$$
\begin{array}{ll}
\underset{\mathbf{x} \in \mathscr{X}}{\operatorname{minimize}} & f(\mathbf{x}) \\
\text { subject to } & \mathbf{g}(\mathbf{x}) \leq \mathbf{0}_{p} \\
& \mathbf{h}(\mathbf{x})=\mathbf{0}_{q}
\end{array}
$$

and this problem is called convex provided that $f(\mathbf{x}), \mathbf{g}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$ are convex.
Note that the set $\mathscr{X}=\operatorname{dom}(f) \cap \operatorname{dom}(\mathbf{g}) \cap \operatorname{dom}(\mathbf{h})$ is the common domain of all functions.

The minimum $p^{*}$ of the optimization problem in Definition 3.12 is defined as

$$
p^{*}=\inf \left\{f(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}_{p}, \mathbf{h}(\mathbf{x})=\mathbf{0}_{q}\right\}
$$

We allow that $p^{*}= \pm \infty$. More precisely, if there exist feasible points $\mathbf{x}_{k}$ with $f\left(\mathbf{x}_{k}\right) \rightarrow-\infty$ as $k \rightarrow+\infty$, then we say that the problem is unbounded below. Moreover, we say that $\mathbf{x}^{*}$ is a minimizer, or solves the problem, if it is feasible and $f\left(\mathbf{x}^{*}\right)=p^{*}$. We say that a feasible point $\mathbf{x}^{*}$ is a local minimizer if there exists a
$c>0$ such that:

$$
\begin{equation*}
f\left(\mathbf{x}^{*}\right)=\inf \left\{f(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}_{p}, \mathbf{h}(\mathbf{x})=\mathbf{0}_{q},\left\|\mathbf{x}-\mathbf{x}^{*}\right\|_{2}^{2} \leq c\right\} \tag{3.1}
\end{equation*}
$$

In other words, the feasible point $\mathbf{x}^{*}$ solves the optimization problem

$$
\begin{array}{lc}
\underset{\mathbf{x} \in \mathscr{X}}{\operatorname{minimize}} & f(\mathbf{x}) \\
\text { subject to } & \mathbf{g}(\mathbf{x}) \leq \mathbf{0}_{p},  \tag{3.2}\\
\mathbf{h}(\mathbf{x})=\mathbf{0}_{q}, \\
& \left\|\mathbf{x}-\mathbf{x}^{*}\right\|_{2}^{2} \leq c
\end{array}
$$

This means that $\mathbf{x}^{*}$ minimizes $f$ over nearby points in the feasible set.
Note that if $\mathbf{x}^{*}$ is feasible and $g_{i}\left(\mathbf{x}^{*}\right)=0$, where $i \in\{1, \ldots, p\}$, then the $i$-th inequality constraint $g_{i}\left(\mathbf{x}^{*}\right) \leq 0$ is called active at $\mathbf{x}^{*}$. Otherwise, if $g_{i}\left(\mathbf{x}^{*}\right)<0$, then we say that the constraint is inactive.

### 3.3.4. Subclasses of convex optimization problems

For all convex optimization problems in this dissertation, we shall assume that $\mathscr{X}=\mathbb{R}^{n}$. Situations that require $\mathscr{X} \subset \mathbb{R}^{n}$ will be solved by adding inequality constraints to reduce the feasible set. We now describe two typical classes of continuous optimization problems: least-squares and linear programming problems.

## Least-squares problems

A least-squares problem is an unconstrained optimization problem that is widely known and used. Given an objective function that is a sum of squares of terms $\mathbf{a}_{i} \cdot \mathbf{x}-b_{i}$, a least-squares problem is defined as

$$
\begin{equation*}
\underset{\mathbf{x} \in \mathcal{X}}{\operatorname{minimize}} f(\mathbf{x})=\|\mathbf{A} \mathbf{x}-\mathbf{b}\|_{2}^{2}=\sum_{i=1}^{k}\left(\mathbf{a}_{i} \cdot \mathbf{x}-b_{i}\right)^{2} \tag{3.3}
\end{equation*}
$$

where $\mathbf{A} \in \mathbb{R}^{k \times n}(k \geq n)$ and $\mathbf{b} \in \mathbb{R}^{n}$. The least-squares problem is the standard approach for regression analysis and the basis for many parameter estimation and data fitting methods. The solution of a least-squares problem (3.3) is considered to be simple. It can be reduced to solving a set of linear equations

$$
\left(\mathbf{A}^{\top} \mathbf{A}\right) \mathbf{x}=\mathbf{A}^{\top} \mathbf{b}
$$

Precisely, the problem is non-complex and has a straightforward solution. The criterion used to determine easy to solve problems from hard ones can be found in the theoretical reference (95].

## Linear programs

A linear program is an optimization problem where the objective function and all constraint functions are affine (linear):

$$
\begin{array}{ll}
\underset{\mathbf{x} \in \mathscr{X}}{\operatorname{minimize}} & f(\mathbf{x})=\mathbf{c} \cdot \mathbf{x}  \tag{3.4}\\
\text { subject to } & \mathbf{g}(\mathbf{x})=\mathbf{A x} \leq \mathbf{b}
\end{array}
$$

Here $\mathbf{c} \in \mathbb{R}^{n}, \mathbf{A} \in \mathbb{R}^{p \times n}$, and $\mathbf{b} \in \mathbb{R}^{p}$ are the parameters that specify the objective and constraint functions.

Both least-squares problems and linear programs can be solved efficiently and reliably. However, many processes in science and engineering are inherently nonlinear, therefore, linear models may not effectively fit the data well.

### 3.4. Statistical learning theory

This section gives an overview of the key ideas and insights of statistical learning theory. This introductory section is mainly intended for the reader who is unfamiliar with the mathematical concepts and notations of statistical learning theory, regularization and hypothesis spaces. No assumption is made that the reader has a deep background in mathematics, statistics, or computer science. It should be noted that this introduction is the result of grouping several technical studies of the mathematics of statistical learning theory, and the material can be found in the works of Vapnik [96, Schölkopf and Smola 97], Bishop [98] and Bousquet et al. 99.

### 3.4.1. The standard framework

Machine learning is a field in computer science that deals with the development and application of algorithms that can be used to learn from data, for instance, to gain new insights, and/or make predictions on (future) data. Three main categories can be distinguished here: unsupervised machine learning, supervised machine learning and semi-supervised machine learning. Unsupervised machine learning mainly focuses on the discovery of patterns in high-dimensional datasets. Dimensionality reduction techniques are probably the most well-known category of unsupervised learning techniques. Supervised machine learning deals with the problem of deriving a prediction function from a dataset. Therefore, the data space is partitioned into an input space and an output space, and it is assumed that there exists some (unknown) mapping between these spaces. The goal of supervised machine learning is to infer this mapping from the data. Semi-supervised machine learning falls
between unsupervised machine learning and supervised machine learning, and mainly deals with the problem of discovering patterns in high-dimensional datasets and deriving a prediction function from the dataset.

The three categories appear to exhaustively classify machine learning paradigms, but they do not. Reinforcement machine learning, which deals with learning by interaction and does not rely on examples of correct behaviour, is also considered to be another category of machine learning, alongside unsupervised machine learning, supervised machine learning and semi-supervised machine learning. In reinforcement machine learning, the learner instead of being told which actions to take, must discover which actions yield the most reward by learning from their own experience. Although one might be tempted to think of reinforcement machine learning as a kind of unsupervised machine learning, reinforcement machine learning is trying to maximize a reward signal instead of trying to find hidden structure.

Statistical learning theory is a framework that provides a statistical basis for dealing with (mainly) supervised machine learning problems. Let $\mathscr{X}$ be the input space and $\mathscr{y}$ be the output space. Statistical learning theory takes the assumption that there exists an (unknown) probability density function $P$ over $\mathscr{X} \times \mathscr{Y}$, such that $P_{\mathcal{X}, \mathcal{Y}}(\mathbf{x}, \mathbf{y})$ is the probability of observing the couple $(\mathbf{x}, \mathbf{y}) \in \mathscr{X} \times \mathscr{Y}$. The training set $\mathscr{T}$ is a random sample from $P$. The outputs $\mathbf{y}$ may not always be a deterministic function of the inputs $\mathbf{x}$, but can be random themselves. Notably, given an input $\mathbf{x}$, there exists multiple outputs with conditional probability distribution function $P_{\mathcal{Y} \mid \mathcal{X}}(\cdot \mid \mathbf{x})$.

### 3.4.2. Estimation and hypothesis space

The quality of a prediction function w.r.t. a given supervised learning problem is often determined using a loss function. We define a loss function as a function $\mathcal{L}: \mathscr{Y} \times \mathscr{Y} \rightarrow \mathbb{R}^{+}$that measures the dissimilarity between two elements in $\mathscr{Y}$. For a given input vector $\mathbf{x}$, we define the risk of a function as the average loss over a predicted vector $\hat{\mathbf{y}}$ according to the underlying probability density function $P_{\mathcal{X}, \mathcal{Y}}$ as follows:

$$
\begin{equation*}
\int_{\mathscr{Y}} \mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) P_{\mathcal{Y} \mid \mathcal{X}}(\mathbf{y} \mid \mathbf{x}) \mathrm{d} \mathbf{y} \tag{3.5}
\end{equation*}
$$

In other words, the risk of a function $\mathbf{f}$ is the expected loss when $\hat{\mathbf{y}}$ is used to predict the output given an input $\mathbf{x} \in \mathscr{X}$. Intuitively, a function is optimal when the risk is minimal. We define an optimal prediction function $\mathbf{f}^{*}: \mathscr{X} \rightarrow \mathscr{Y}$ as follows:

$$
\begin{equation*}
\mathbf{f}^{*}(\mathbf{x})=\underset{\hat{\mathbf{y}}}{\arg \min } \int_{\mathscr{Y}} \mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) P_{\mathcal{Y} \mid \mathcal{X}}(\mathbf{y} \mid \mathbf{x}) \mathrm{d} \mathbf{y} \tag{3.6}
\end{equation*}
$$

The use of Eq. (3.6) requires the conditional probability distribution $P_{\mathcal{Y} \mid \mathcal{X}}$ to be known for all $\mathbf{x} \in \mathcal{X}$. The risk of a function $\mathbf{f}$ is defined as follows:

$$
\begin{equation*}
r(\mathbf{f})=\int_{\mathscr{X} \times \mathcal{Y}} \mathcal{L}(\mathbf{f}(\mathbf{x}), \mathbf{y}) P_{\mathcal{X}, \mathcal{Y}}(\mathbf{x}, \mathbf{y}) \mathrm{d} \mathbf{x} \mathrm{~d} \mathbf{y} \tag{3.7}
\end{equation*}
$$

Up to this point, $\mathbf{f}$ could be any function, however, in practice $\mathbf{f}$ is typically limited to a set of admissible functions that is called the hypothesis space $\mathbb{H}$. For example, $\mathbb{H}$ can be the space of all convex functions $\mathbf{f}: \mathcal{X} \rightarrow \mathscr{Y}$. Subsequently, the optimal prediction function $\mathbf{f}_{r}^{*} \in \mathbb{H}$ that minimizes this risk is computed as follows:

$$
\begin{align*}
\mathbf{f}_{r}^{*}(\mathbf{x}) & =\underset{\mathbf{f} \in \mathbb{H}}{\arg \min } \int_{X \times \mathcal{Y}} \mathcal{L}(\mathbf{f}(\mathbf{x}), \mathbf{y}) P_{\mathcal{X}, \mathcal{Y}}(\mathbf{x}, \mathbf{y}) \mathrm{d} \mathbf{x} \mathrm{~d} \mathbf{y}  \tag{3.8}\\
& =\underset{\mathbf{f} \in \mathbb{H}}{\arg \min } \int_{\mathscr{X} \times \mathcal{Y}} \mathcal{L}(\mathbf{f}(\mathbf{x}), \mathbf{y}) P_{\mathcal{Y} \mid \mathcal{X}}(\mathbf{y} \mid \mathbf{x}) P_{\mathcal{X}}(\mathbf{x}) \mathrm{d} \mathbf{x} \mathrm{~d} \mathbf{y} \tag{3.9}
\end{align*}
$$

Interestingly, Eq. (3.9) the joint probability distribution $P_{\mathcal{X}, \mathcal{Y}}$ in Eq. 3.8) is transformed as a function of the conditional probability distribution $P_{\mathcal{Y} \mid \mathcal{X}}$ and $P_{\mathcal{X}}$, the probability distribution of the random vector $\mathcal{X}$. This shows a direct link between Eq. (3.8) and (3.6). As Eq. 3.7) (and Eq. 3.8) still requires a known joint distribution $P_{\mathcal{X}, \mathcal{Y}}$, the risk $r(\mathbf{f})$ cannot be computed. We can, however, try to infer from the training set $\mathscr{T}$ a function $f$ has the lowest risk. A more straightforward approach is to compute an approximation, called empirical risk (or, in some cases, the training error), by using the training set $\mathscr{T}$ to estimate the risk of a function $\mathbf{f} \in \mathbb{H}$. Formally, the loss function is averaged over $\mathscr{T}$, and the resulting quantity is defined as follows

$$
\begin{equation*}
r_{e m p}(\mathbf{f})=\frac{1}{n} \sum_{i=1}^{n} \mathcal{L}\left(\mathbf{f}\left(\mathbf{x}_{i}\right), \mathbf{y}_{i}\right) \tag{3.10}
\end{equation*}
$$

The learning algorithm should choose a hypothesis that minimizes the empirical risk, and that can be used to obtain a function to make predictions. This principle is called empirical risk minimization. We define an optimal prediction function $\mathbf{f}_{\text {emp }}^{*} \in \mathbb{H}$ as follows:

$$
\begin{equation*}
\mathbf{f}_{e m p}^{*}(\mathbf{x})=\underset{\mathbf{f} \in \mathbb{H}}{\arg \min } \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}\left(\mathbf{f}\left(\mathbf{x}_{i}\right), \mathbf{y}_{i}\right) \tag{3.11}
\end{equation*}
$$

### 3.4.3. Overfitting and regularization

In general, the aim is to find a function $\mathbf{f}_{r}^{*} \in \mathbb{H}$ that minimizes the risk $r(\mathbf{f})$. However, in practice, this is infeasible since $r(\mathbf{f})$ requires knowledge of the joint distribution. Therefore, an approximating function $\mathbf{f}_{e m p}^{*} \in \mathbb{H}$ is learned instead. Unfortunately, $\mathbf{f}_{e m p}^{*}$ is not always a good approximation of $\mathbf{f}_{r}^{*}$, especially as the hypothesis space $\mathbb{H}$ becomes larger, it is likely that $r_{e m p}(\mathbf{f})<r(\mathbf{f})$, and as a result $\mathbf{f}_{e m p}^{*}$ and $\mathbf{f}_{r}^{*}$ become very dissimilar. In such a case, we say that the learned model is complex and that $\mathbf{f}_{e m p}^{*}$ would be an overly optimistic estimate of $\mathbf{f}_{r}^{*}$.

In the terminology of statistics, this is known as the bias-variance trade-off. If the learned model shows low bias but high variance (complex model), then this would result in overfitting. Otherwise, if the model shows low variance but high bias (overly simple model), then this would lead to underfitting.

Minimizing the empirical risk $r_{e m p}(\mathbf{f})$ and then comparing it to the overall risk $r(\mathbf{f})$ using the performance measure has been argued to result, in some cases, in overfitting or underfitting leading to poorly built predictive models. An implicit and direct approach of solving this is the principle of regularization. We define the regularized empirical risk as

$$
\begin{equation*}
r_{r e g}(\mathbf{f})=r_{e m p}(\mathbf{f})+\lambda R(\mathbf{f}), \tag{3.12}
\end{equation*}
$$

where $R(\mathbf{f})$ is the so-called regularizer, and $\lambda>0$ is the regularization parameter ${ }^{2}$. In addition, the term $\lambda R(\mathbf{f})$ of the regularized empirical risk is called the regularization term.

By substituting Eq. (3.10) into Eq. 3.12), we get the regularized risk as follows

$$
\begin{equation*}
r_{r e g}(\mathbf{f})=\frac{1}{n} \sum_{i=1}^{n}\left(\mathcal{L}\left(\mathbf{f}\left(\mathbf{x}_{i}\right), \mathbf{y}_{i}\right)\right)+\lambda R(\mathbf{f}) . \tag{3.13}
\end{equation*}
$$

One can minimize the regularized risk that results in the optimal prediction function $\mathbf{f}_{\text {reg }}$ as follows

$$
\begin{equation*}
\mathbf{f}_{\text {reg }}=\underset{\mathbf{f} \in \mathbb{H}}{\arg \min } r_{\text {reg }}(\mathbf{f}) . \tag{3.14}
\end{equation*}
$$

The regularizer $R(\mathbf{f})$ is supposed to exclude all the 'unlikely' functions (i.e., the extremely irregular functions are eliminated). Moreover, the regularization parameter $\lambda$ is considered a bias-variance trade-off constant. More specifically, it handles the trade-off between the fit of a function that can be used for prediction, and the

[^3]complexity of that function. To fine-tune $\lambda$, cross-validation is normally performed. In a nutshell, cross-validation splits the available data, once or several times, for minimizing the empirical risk. Part of data (the training sample) is used to train the model, and the remaining part (the validation sample) is used to test the model by estimating the empirical risk. The optimal value of $\lambda$ is obtained by repeating this procedure for different values of $\lambda$ and then selecting the value that results in the smallest estimated risk 100 .

### 3.5. Statistical learning as an optimization problem

In statistical learning, generalization is by far the most important property used to validate a learning approach. In a typical learning problem, firstly, a training loss function has to be picked. Then, an appropriate optimal function that generalizes well, based on the given training data, has to be searched for. This function will be used to build a model. Typically, this process introduces errors that diminish the quality of the resulting optimal functions. We now consider three causes of such errors. The first cause of error stems from having an unknown distribution. Therefore, any choice of models and loss functions could introduce an inappropriate bias and may not be suitable for the problem. The second cause of error is the fact that there exists only a finite number of (possibly noisy) training examples. Therefore, for a given hypothesis space even after choosing appropriate models and loss functions, the model could still be unsuitable and the results also may be inappropriate. The third cause of error arises from the difficulty of searching for an appropriate optimal function that results in a small generalization error and an appropriate model. This problem can be improved (or even solved) by transforming it into a convex optimization problem. This is achieved by mainly defining the appropriate loss functions, hypotheses spaces, regularization functions, and constraints.

### 3.5.1. Regression problems

We start off with the regularized empirical risk function to transform the statistical learning model into a mathematical optimization problem:

$$
\begin{equation*}
\underset{f \in \mathbb{H}}{\operatorname{minimize}} \sum_{i=1}^{n} \mathcal{L}\left(f\left(\mathbf{x}_{i}\right), y_{i}\right)+\lambda\|\mathbf{a}\| \tag{3.15}
\end{equation*}
$$

We define the input and output space as $\mathscr{X}=\mathbb{R}^{p}$ and $\mathscr{Y}=\mathbb{R}$, respectively. A common choice of loss function in regression problems is the squared loss
function $\mathcal{L}(y, \hat{y})=\|y-\hat{y}\|_{2}^{2}$, and a hypothesis space of all affine functions such that $\mathbb{H}: \mathscr{X} \rightarrow \mathscr{Y}$. For instance, we can define the function $f(\mathbf{x})=\mathbf{a} \cdot \mathbf{x}+b$ for all $f \in \mathbb{H}$ and $\mathbf{x} \in \mathscr{X}$. To measure the irregularities of such a function, we take the squared Euclidean norm (also known as $\ell_{2}$-norm) of a. Therefore, the regularization function is defined as the $\ell_{2}$-norm $\|\mathbf{a}\|_{2}^{2}$. Given all these components, we describe one of the most popular regression methods used in practice, known as ridge regression [101. It is defined as the following optimization problem:

$$
\begin{equation*}
\underset{\mathbf{a}, b}{\operatorname{minimize}} \sum_{i=1}^{n}\left(\mathbf{a} \cdot \mathbf{x}_{i}+b-y_{i}\right)^{2}+\lambda\|\mathbf{a}\|_{2}^{2} \tag{3.16}
\end{equation*}
$$

Alternatively, we could also use other loss functions instead of the squared loss. One popular alternative is the $\epsilon$-insensitive loss function $\mathcal{L}(y, \hat{y})=\max (0,|y-\hat{y}|-\epsilon)$. The strategy described is known as support vector regression [102, 103, and is as follows:

$$
\begin{equation*}
\underset{\mathbf{a}, b}{\operatorname{minimize}} \sum_{i=1}^{n} \max \left(0,\left|\mathbf{a} \cdot \mathbf{x}_{i}+b-y_{i}\right|-\epsilon\right)+\lambda\|\mathbf{a}\|_{2}^{2} . \tag{3.17}
\end{equation*}
$$

Moreover, an alternative to the $\ell_{2}$-norm of the parameter vector a is the $\ell_{1}$-norm. There, the regularization function is defined as $R(\mathbf{a})=\|\mathbf{a}\|_{1}$. The applied strategy on (3.16) is known as lasso regression (104 and results in:

$$
\begin{equation*}
\underset{\mathbf{a}, b}{\operatorname{minimize}} \sum_{i=1}^{n}\left(\mathbf{a} \cdot \mathbf{x}_{i}+b-y_{i}\right)^{2}+\lambda\|\mathbf{a}\|_{1} . \tag{3.18}
\end{equation*}
$$

In an optimization setting, problem (3.16) can be solved as an unconstrained convex optimization problem. Moreover, problem (3.17) can be transformed into an equivalent linearly constrained convex optimization problem with a quadratic objective function 97.

### 3.5.2. Classification problems

Up to this point, we have been discussing typical regression problems, where the output space is $\mathbb{R}$. The case where the output space is the discrete set of labels is called a classification problem. One of the widely used methods for solving classification problems is logistic regression where the output space is the discrete set of binary labels. Classification problems solved suing logistic regression are dealt with in a similar as regression problems, however, the goal is to find a function $\mathbf{f}: \mathscr{X} \rightarrow\{0,1\}$. In statistical learning theory, this function is defined as $f(\mathbf{x})=\operatorname{Pr}(\mathcal{Y}=1 \mid \mathcal{X}=\mathbf{x})$. Interestingly, the opposite case is simply defined
as $1-f(\mathbf{x})$. Logistic regression transforms its output using the logistic sigmoid function to return a probability value which can then be mapped to two discrete classes. In machine learning, the sigmoid function is used for mapping predictions to probabilities. The sigmoid function maps any real value into another value between 0 and 1 and is defined as $f(\mathbf{x})=\frac{\exp (\mathbf{a} \cdot \mathbf{x}+b)}{1+\exp (\mathbf{a} \cdot \mathbf{x}+b)}$ for all $f \in \mathbb{H}$ and $\mathbf{x} \in \mathcal{X}$. Computing the component-wise squared loss function results in a highly non-convex problem that can only be solved locally, and the exponential terms can result in numerical instability. As an alternative, we use the notation $y_{i} \in\{0,1\}$ and generalize $f(\mathbf{x})$ by computing the likelihood that $\operatorname{Pr}\left(\mathcal{Y}=y_{i} \mid \mathcal{X}=\mathbf{x}_{i}\right)$ as follows:

$$
\begin{equation*}
\prod_{i=1}^{n}\left(f\left(\mathbf{x}_{i}\right)\right)^{y_{i}}\left(1-f\left(\mathbf{x}_{i}\right)\right)^{1-y_{i}} \tag{3.19}
\end{equation*}
$$

For computational convenience, the negative log-likelihood is considered in Eq. (3.19) as loss function. To follow the tradition in the field of mathematical optimization, the loss function is minimized as follows:

$$
\begin{equation*}
\underset{\mathbf{a}, b}{\operatorname{minimize}}-\sum_{i=1}^{n}\left(y_{i} \log f\left(\mathbf{x}_{i}\right)+\left(1+y_{i}\right) \log \left(1-f\left(\mathbf{x}_{i}\right)\right)\right)+\lambda\|\mathbf{a}\|_{1} . \tag{3.20}
\end{equation*}
$$

In machine learning, Eq. 3.20 is known as $\ell_{1}$-regularized logistic regression and has been used for many classification problems, particularly ones with many features. It can be seen that this problem is convex, however, the sparsityinducing regularization term is non-smooth. There exist several strategies to solve optimization problems with sparsity-inducing penalties [105, 106, 107, 108 . In logistic regression, one way of solving this problem requires solving a constrained convex optimization problem [109], where the regularization term is transformed into an equivalent constraint, as follows:

$$
\begin{align*}
& \underset{\mathbf{a}, b}{\operatorname{minimize}} \quad-\sum_{i=1}^{n}\left(y_{i} \log f\left(\mathbf{x}_{i}\right)+\left(1+y_{i}\right) \log \left(1-f\left(\mathbf{x}_{i}\right)\right)\right),  \tag{3.21}\\
& \text { subject to } \quad\|\mathbf{a}\|_{1} \leq B .
\end{align*}
$$

The optimization problems Eq. (3.20) and Eq. (3.21) are equivalent, in the sense that for any choice of $\lambda$ in Eq. 3.20, there is a choice of $B$ in Eq. 3.21) that will give the same $\lambda$. In practice, $B$ (similar to $\lambda$ ) can be chosen via cross-validation.

An extension of the binomial logistic regression is known as the multi-class logistic regression [110, which is very popular in multi-class classification. Multi-class logistic regression is used when the output variable has more than two (unordered) categories. Using the notation $y_{i} \in\{0,1\}^{q}$, a function $\mathbf{f}: \mathbb{R}^{p} \rightarrow\{0,1\}^{q}$, such that $\mathbf{f} \in \mathbb{H}$, can be written as follows:

$$
f_{j}(\mathbf{x})= \begin{cases}\frac{\exp \left(\mathbf{a}_{j} \cdot \mathbf{x}+b_{j}\right)}{1+\sum_{j=1}^{q-1} \exp \left(\mathbf{a}_{j} \cdot \mathbf{x}+b_{j}\right)}, & \text { for } j=1, \ldots, q-1  \tag{3.22}\\ \frac{1}{1+\sum_{j=1}^{q-1} \exp \left(\mathbf{a}_{j} \cdot \mathbf{x}+b_{j}\right)}, & \text { for } j=q\end{cases}
$$

where $\mathbf{a}_{j} \in \mathbb{R}^{p}$ and $b_{j} \in \mathbb{R}$.
Similar to the case of logistic regression, the use of a component-wise squared loss function results in a highly non-convex problem that can be numerically unstable. Therefore, the multinomial deviance [110] can be used as a loss function, such that $\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}})=-\sum_{j=1}^{q} y_{j} \log \left(\hat{y}_{j}\right)$ and leads to the following optimization problem:

$$
\begin{equation*}
\underset{\mathbf{a}, b}{\operatorname{minimize}} \sum_{j=1}^{q} \sum_{i=1}^{n}\left(y_{i j} \log f_{j}\left(\mathbf{x}_{i}\right)\right)+\lambda\|\mathbf{a}\|_{1} . \tag{3.23}
\end{equation*}
$$

Problem (3.23) (similar to Problem (3.20) is convex and can be solved as a constrained convex optimization problem.

The setting where the output variable has multiple ordered (rather than unordered) categories is considered in Part $I V$ of this dissertation. In this setting, the disadvantage of a multi-class regression model is that information about the ordering is neglected. Therefore, we will delve into another extension of the binomial logistic regression, namely, ordinal logistic regression (or ordinal regression for short) that incorporates the information about the ordering of categories. We will also discuss a method for transforming a sparsity-inducing penalty term into an equivalent constraint similar to Eq. 3.22).

## PART II

## GATHERING SENSORY AND INSTRUMENTAL DATA

# 4 Sensory evaluation and chemical analysis of food products 

## Table of Contents

4.1 Introduction<br>4.2 Experimental sensory evaluation set-up<br>4.2.1 Preliminaries<br>4.2.2 Chicken breast<br>4.2.3 Atlantic cod<br>4.2.4 Atlantic brown shrimp<br>4.2.5 Atlantic salmon<br>4.3 Quantification of spoilage-related VOCs

### 4.1. Introduction

The determination of food quality is not a perfect science. Panellists have different preferences, and samples of the same food provided to these panellists can be different. Therefore, it becomes difficult to compare the quality of different food samples. As a result, guidelines have been set for consistent evaluation of food.

In this chapter, we present the guidelines followed for studying different foods and compile the different experimental set-ups for data acquisition. As a result, we answer the question:

Question II: Which foods were studied and what was collected?
The experiments described in this chapter revolve around two main points:

1. Sensory evaluation of the following packaged foods:

- Chicken breasts: labelling and ranking
- Atlantic Cod (Gadus morhua): labelling and ranking
- Atlantic Brown Shrimp (Crangon crangon): labelling and ranking
- Atlantic Salmon (Salmo salar): labelling, scoring and ranking

2. Chemical experimentation using SIFT-MS techniques for the above foods.

All the food samples were packaged under different conditions and stored for a different number of days before beginning the evaluation. The origin, packaging,
storage temperature, storage day, sample size and panel size are all scrutinized during the process.

### 4.2. Experimental sensory evaluation set-up

### 4.2.1. Preliminaries

In this section, we describe the different experimental sensory evaluation set-ups for the different foods: chicken breasts, cod, brown shrimp and salmon. Samples of these foods are studied under different packaging conditions and for different numbers of storage days.

As a starting point, we provide an example of a sensory evaluation facility in Figure 4.1 showing various activities, particularly, a product preparation area, where researchers can package the food product at specific storage conditions, a sensory evaluation preparation area, where researchers can divide the food product into samples for use in sensory evaluation, and booths, where panellists are provided specific samples and asked to evaluate the samples at their own booth.

Figure 4.1: Example of sensory evaluation facility showing the various activities for gathering sensory evaluation data.


Now, we provide an overview of the notations used to describe the sensory evaluation tests and the variables in this chapter:

Tests:

- Labelling tests: Samples were assigned a label on an ordinal scale with labels "Spoiled" (SP), "Marginal" (M), "Satisfactory" (S), "Fresh" (F), "Very Fresh"
(VF), such that $\mathrm{SP} \prec \mathrm{M} \prec \mathrm{S} \prec \mathrm{F} \prec \mathrm{VF}$. Chicken samples were assigned a label on an ordinal scale with labels SP, S and F only.
- Ranking tests: Samples were ranked from least fresh to most fresh. The fact that a sample was ranked above another sample means that the former was preferred over the latter.
- Scoring tests: Samples were assigned a score on a 5 -point scale, where the extreme scores of " 1 " and " 5 " represent spoiled and fresh, respectively, and the intermediate score of " 3 " represents a neutral response of neither spoiled nor fresh.

Variables:

- $a^{k}$ : the sample $a$ that has been stored for $k$ days
- $\mathbf{z}_{i}$ : the list $\mathbf{z}$ of $n$ labels assigned by the $i$-th trained panellist to the $n$ samples
- $\prec_{i}$ : the ranking provided by the $i$-th untrained panellist for the $n$ samples
- $\mathbf{s}_{i}$ : the vector of scores $\left(s_{i 1}, \ldots, s_{i n}\right)$ assigned by the $i$-th trained panellist to the $n$ samples that are ordered according to increasing storage days


### 4.2.2. Chicken breast

Fresh double skinless chicken breast fillets weighing around 300-400g were first prepared at a local company, without the use of any additives. The fillets were delivered the same day to the Research Unit Food Microbiology and Food Preservation (FMFP, UGent) in a cooler box with cooling elements. Then, the fillets were hand mixed (with disinfected latex gloves) by moving them around in the box for five minutes to ensure homogeneous contamination on the surface and were then each packaged ${ }^{1}$.

These packages were stored under different MAP conditions and temperatures, where the notation of the MAP conditions indicates high (H) or low (L) O2 content and temperature in Celsius degrees ( 4 or 8 ), as follows:

- L4: 40/30/30 (CO2/O2/N2) at $4 \pm 1^{\circ} \mathrm{C}$;
- L8: $40 / 30 / 30(\mathrm{CO} 2 / \mathrm{O} 2 / \mathrm{N} 2)$ at $8 \pm 1^{\circ} \mathrm{C}$;
- H4: 30/70 (CO2/O2) at $4 \pm 1^{\circ} \mathrm{C}$;
- H8: 30/70 (CO2/O2) at $8 \pm 1^{\circ} \mathrm{C}$.

For each individual storage experiment, trays were randomly sampled on specific storage days for quantification of VOCs (Section 4.3) and were then frozen at

[^4]$-32^{\circ} \mathrm{C}$ under vacuum ${ }^{2}$ The sampling of the chicken samples was performed on the storage days shown in Table 4.1 .

|  | Storage days |  |
| :---: | :---: | :---: |
| Experiment | Labelling tests | Ranking tests |
| L4: | $0,5,7,9,11,13,15$ |  |
| L8: | $0,2,4,5,6$ |  |
| H4: | $0,5,7,8,9,11,13,15$ | $0,5,7,9,11,13,15$ |
| H8: | $0,2,4,5,6$ |  |

Table 4.1: Sampling chicken samples packaged at day 0 for quantification of VOCs and sensory evaluation (labelling and ranking) tests.

Sensory evaluation, namely, ranking and labelling tests, was then performed on the chicken samples in several experiments. It must be noted that the samples used for the labelling tests are different from those used for the ranking tests. In these experiments, the samples on the day of evaluation were thawed at $2^{\circ} \mathrm{C}$ overnight, cut to $15.0 \pm 0.1 \mathrm{~g}$ portions and presented to the panellists at $4^{\circ} \mathrm{C}$ in odour-free and transparent plastic cups, closed with lid $\}^{3}$ and labelled with three-digit random codes according to an $n$-sample Latin Square Design, where $n$ is the number of samples provided to panellists at one time in each experiment. Sensory evaluation was based on olfactory evaluation and performed in individual booths under red light (SensoLab UGent).

## Labelling

The labelling test was carried out only for storage experiment condition H4. The eight samples were served to 33 panellists, as shown in Table 4.2. The panellists were recruited from the Faculty of Bioscience Engineering at Ghent University with previous experience in performing sensory evaluation of chicken and were considered as trained panellists. The panellists were asked to assign to each sample provided to them a label ("Spoiled" (SP), "Satisfactory" (S), or "Fresh" (F)) on an ordinal scale, such that: $\mathrm{SP} \prec \mathrm{S} \prec \mathrm{F}$.

| Samples | Panellists |
| :---: | :---: |
| $a^{0}, a^{5}, a^{7}, a^{8}, a^{9}, a^{11}, a^{13}, a^{15}$ | 33 |

Table 4.2: Chicken samples for labelling tests and the number of panellists in storage experiment H4.

[^5]The labels provided by the panellists are gathered in Table A.1.

## Ranking

The ranking tests were carried out for each storage experiment condition separately. The tests in $4^{\circ} \mathrm{C}$ conditions were planned in a randomized design of two sessions, where four samples were selected and served in session 1 and the three remaining samples were served in session 2. It must be noted that the samples provided to the panellists are the same, however, in the order described in Table 4.3, while experiments in $8^{\circ} \mathrm{C}$ conditions were planned in one session, where five samples were served simultaneously to the panellists. Fourteen panellists were recruited from the Faculty of Bioscience Engineering at Ghent University with previous experience in performing sensory evaluation of chicken. However, as the ranking tests in $4^{\circ} \mathrm{C}$ conditions were planned in a randomized design of two sessions, the panellists were divided among the different groups of samples that were randomly selected for each panellist. The panellists were asked to rank the samples according to their perceived freshness from least fresh to most fresh.

|  | Samples |  |  | Panellists |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Group | Session 1 | Session 2 |  | L4 | H4 |
| 1 | $a^{0}, a^{5}, a^{7}, a^{9}$ | $a^{11}, a^{13}, a^{15}$ |  | 4 | 3 |
| 2 | $a^{0}, a^{5}, a^{7}, a^{11}$ | $a^{9}, a^{13}, a^{15}$ |  | 4 | 4 |
| 3 | $a^{0}, a^{7}, a^{9}, a^{15}$ | $a^{5}, a^{11}, a^{13}$ |  | 3 | 2 |
| 4 | $a^{5}, a^{7}, a^{11}, a^{13}$ | $a^{0}, a^{9}, a^{15}$ |  | 4 | 4 |
|  |  |  |  | Panellists |  |
| Group | Samples |  | L8 | H8 |  |
| 1 | $a^{0}, a^{2}, a^{4}, a^{5}, a^{6}$ |  | 14 | 14 |  |

Table 4.3: Chicken samples for ranking tests and the number of panellists in each storage experiment.

The rankings provided by the panellists in experiments L 4 and H 4 in Table 4.3 are gathered in Table A.2, and the rankings provided by the panellists in experiments L8 and H8 in Table 4.3 are gathered in Table A.3.

### 4.2.3. Atlantic cod

Atlantic cod was caught in the North Atlantic Ocear ${ }^{4}$. weighing around 4.5 kg , was gutted, filleted and skinned at a commercial seafood processing company, which we

[^6]will refer to as company A, without the use of any additives. Similarly, Atlantic cod, weighing around 4 kg , was gutted, filleted and skinned at a second commercial seafood processing company, which we will refer to as company B. The fresh fillets, weighing around $217 \pm 5 \mathrm{~g}$, were transported to Belgium by air and delivered to FMFP at Ghent University in polystyrene boxes under ice. The cod fillets were first manually mixed for one minute to ensure homogeneous contamination on the surfaces of each fillet and were each then packaged ${ }^{5}$.

These packages were stored under different MAP conditions and temperatures, where the notation of the MAP conditions indicates high (H) or low (L) O2 content or O 2 content in air (A) and temperature in Celsius degrees (4 or 8), as follows:

- L4: $60 / 5 / 35(\mathrm{CO} 2 / \mathrm{O} 2 / \mathrm{N} 2)$ at $4 \pm 0.7^{\circ} \mathrm{C}$;
- L8: 60/5/35 (CO2/O2/N2) at $8 \pm 1.4^{\circ} \mathrm{C}$;
- H4: 60/40 (CO2/O2) at $4 \pm 0.7^{\circ} \mathrm{C}$;
- H8: 60/40 (CO2/O2) at $8 \pm 1.4^{\circ} \mathrm{C}$;
- A4: air at $4 \pm 0.7^{\circ} \mathrm{C}$.

For each individual storage experiment, trays were randomly sampled on specific storage days for quantification of VOCs (Section 4.3) and were then frozen at $-32^{\circ} \mathrm{C}$ under vacuum ${ }^{6}$. The sampling of the cod samples was performed on the storage days shown in Table 4.4 .

|  | Storage days |  |
| :---: | :---: | :---: |
| Experiment | Labelling tests | Ranking tests |
| L4: | $0,4,8,13$ | $0,4,5,6,7,8,13$ |
| L8: | $0,3,5,7$ | $0,3,5,7$ |
| H4: | $0,4,8,13$ | $0,4,6,7,8,11,13$ |
| H8: | $0,3,5,7$ | $0,3,4,5,6,7$ |
| A4: | $0,1,2,3$ | $0,1,2,3$ |

Table 4.4: Sampling cod samples packaged at day 0 for quantification of VOCs and sensory evaluation (labelling and ranking) tests.

Sensory evaluation, namely, ranking and labelling tests, was then performed on

[^7]the cod samples in several experiments. It must be noted that the samples used for the labelling tests are different from those used for the ranking tests. In these experiments, the samples on the day of evaluation were thawed at $2^{\circ} \mathrm{C}$ overnight, cut to $5.0 \pm 0.1 \mathrm{~g}$ portions and presented to the panellists at $4^{\circ} \mathrm{C}$ in odour-free and transparent plastic cups, closed with lid $\left\{{ }^{7}\right.$ and labelled with three-digit random codes according to a 4 -sample Latin Square Design. Sensory evaluation was based on olfactory evaluation and performed in individual booths under red light (SensoLab UGent).

## Labelling

The labelling test was carried out for cod samples of fillets from company A for all storage experiment conditions. The labelling test was also carried out for cod samples of fillets from company B for storage experiments $\mathrm{H} 4, \mathrm{H} 8$, and A4. A number of panellists (between 8 and 12) were recruited from the Faculty of Bioscience Engineering at Ghent University with previous experience in performing sensory evaluation of cod. The samples provided to the panellists and the number of panellists in each labelling test are described in Table 4.5 .

|  | Company A |  |  | Company B |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Experiment | Samples | Panellists |  | Samples | Panellists |
| L4 | $a^{0}, a^{4}, a^{8}, a^{13}$ | 8 |  |  |  |
| L8 | $a^{0}, a^{3}, a^{5}, a^{7}$ | 10 |  |  |  |
| H4 | $a^{0}, a^{4}, a^{8}, a^{13}$ | 10 |  | $a^{0}, a^{4}, a^{8}, a^{13}$ | 10 |
| H8 | $a^{0}, a^{3}, a^{5}, a^{7}$ | 12 |  | $a^{0}, a^{3}, a^{5}, a^{7}$ | 10 |
| A4 | $a^{0}, a^{1}, a^{2}, a^{3}$ | 10 |  | $a^{0}, a^{1}, a^{2}, a^{3}$ | 10 |

Table 4.5: Cod samples for labelling tests and the number of panellists in each storage experiment.

The panellists were asked to assign to each sample provided to them a label ("Spoiled" (SP), "Marginal" (M), "Satisfactory" (S), "Fresh" (F), "Very Fresh" (VF)) on an ordinal scale, such that: SP $\prec \mathrm{M} \prec \mathrm{S} \prec \mathrm{F} \prec \mathrm{VF}$. The labels assigned by the panellists are gathered in Table A.4. Unfortunately, data gathered from some labelling tests were provided only in the form of frequency distributions. These data are gathered in Table A.5.

## Ranking

The ranking tests were carried out for cod samples from company A for each storage experiment condition separately. For each of the storage experiments, four samples were studied. For storage experiments L4, H4 and H8 a second session

[^8]was performed to study the critical storage days that were identified in the first session. Note that samples in the second session with the same storage day as the samples in the first session are actually different samples.

The samples provided to the panellists and the number of panellists in each ranking test are described in Table 4.6. The panellists were asked to rank the samples according to their perceived freshness from least fresh to most fresh.

|  | Session 1 |  |  | Session 2 |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Experiment | Samples | Panellists |  | Samples | Panellists |
| L4 | $a^{0}, a^{4}, a^{8}, a^{13}$ | 10 |  | $a^{5}, a^{6}, a^{7}, a^{8}$ | 8 |
| L8 | $a^{0}, a^{3}, a^{5}, a^{7}$ | 9 |  |  |  |
| H4 | $a^{4}, a^{6}, a^{8}, a^{13}$ | 8 |  | $a^{6}, a^{7}, a^{8}, a^{11}$ | 9 |
| H8 | $a^{0}, a^{3}, a^{5}, a^{7}$ | 8 |  | $a^{3}, a^{4}, a^{5}, a^{6}$ | 8 |
| A4 | $a^{0}, a^{1}, a^{2}, a^{3}$ | 8 |  |  |  |

Table 4.6: Cod samples for ranking tests and the number of panellists in each storage experiment.

The rankings provided by the panellists are gathered in Table A.6.

### 4.2.4. Atlantic brown shrimp

Atlantic brown shrimps were caught in the North Atlantic Ocear ${ }^{8}$ in October and November 2015 were sorted according to size and washed before cooking without the use of any additives according to normal Belgian fishing practices. The fresh shrimp were then cooked and, subsequently, cooled and stored overnight in plastic bags under ice. The following morning, the shrimp were directly transported to FMFP at Ghent University, where they were hand-peeled. During peeling, the shrimp were kept on ice in plastic bags while avoiding direct contact between the shrimp and the ice. Shrimp portions, weighing around $150 \pm 2 \mathrm{~g}$, were then packaged at $2: 1$ headspace-product ratid ${ }^{9}$

These packages were stored under different MAP conditions at $4^{\circ} \mathrm{C}$, where the notation of the MAP conditions indicates high (H) or low (L) CO2 content, as follows:

- L4: 30/0/70 (CO2/O2/N2) at $4 \pm 0.7^{\circ} \mathrm{C}$;

[^9]- H4: 50/0/50 (CO2/O2/N2) at $4 \pm 0.7^{\circ} \mathrm{C}$.

For each individual storage experiment, trays were randomly sampled on specific storage days for quantification of VOCs (Section 4.3) and were then frozen at $-32^{\circ} \mathrm{C}$ under vacuum ${ }^{10}$. The sampling of the shrimp samples was performed on the storage days shown in Table 4.7 .

|  | Storage days |  |
| :---: | :---: | :---: |
| Experiment | Labelling tests | Ranking tests |
| L4 | $0,3,5,10$ | $0,3,5,10$ |
| H4 | $0,3,5,7,12$ | $0,3,5,7,12$ |

Table 4.7: Sampling shrimp samples packaged at day 0 for quantification of VOCs and sensory evaluation (labelling and ranking) tests.

Sensory evaluation, namely, ranking and labelling tests, was then performed on the shrimp samples in several experiments. It must be noted that the samples used for the labelling tests are different from those used for the ranking tests. In these experiments, the samples on the day of evaluation were thawed at $2^{\circ} \mathrm{C}$ overnight, cut to $5.0 \pm 0.1 \mathrm{~g}$ portions and presented to the panellists at $4^{\circ} \mathrm{C}$ in odourfree and transparent plastic cups, closed with lids ${ }^{11}$ and labelled with three-digit random codes according to a 4 -sample Latin Square Design. Sensory evaluation was based on olfactory evaluation and performed in individual booths under red light (SensoLab UGent).

## Labelling

The labelling test was carried out for shrimp samples for each MAP condition separately. The storage experiments were divided into two sessions in which four samples were studied in each session. Note that samples in the second session with the same storage day as the samples in the first session are actually different samples. A number of panellists (nine or ten) were recruited from the Faculty of Bioscience Engineering at Ghent University with previous experience in performing sensory evaluation of shrimp. The samples provided to the panellists and the number of panellists in each ranking test are described in Table 4.8.

The panellists were asked to assign to each sample provided to them a label ("Spoiled" (SP), "Marginal" (M), "Satisfactory" (S), "Fresh" (F), "Very Fresh" (VF)) on an ordinal scale, such that: $\mathrm{SP} \prec \mathrm{M} \prec \mathrm{S} \prec \mathrm{F} \prec \mathrm{VF}$. The labels assigned by the panellists are gathered in Table A.7.

[^10]|  | Session 1 |  |  | Session 2 |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Experiment | Samples | Panellists |  | Samples | Panellists |
| L4 | $a^{0}, a^{3}, a^{5}, a^{10}$ | 10 |  | $a^{0}, a^{3}, a^{5}, a^{10}$ | 10 |
| H4 | $a^{0}, a^{3}, a^{7}, a^{12}$ | 9 |  | $a^{0}, a^{3}, a^{5}, a^{7}$ | 10 |

Table 4.8: Shrimp samples for labelling tests and the number of panellists in each storage experiment.

The ranking tests were carried out for shrimp samples for each MAP condition separately. The storage experiments were divided into two sessions in which four samples were studied in each session. Note that samples in the second session with the same storage day as the samples in the first session are actually different samples.

The samples provided to the panellists and the number of panellists in each ranking test are described in Table 4.9. The panellists were asked to rank the samples according to their perceived freshness from least fresh to most fresh.

|  | Session 1 |  |  | Session 2 |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Experiment | Samples | Panellists |  | Samples | Panellists |
| L4 | $a^{0}, a^{3}, a^{5}, a^{10}$ | 9 |  | $a^{0}, a^{3}, a^{5}, a^{10}$ | 10 |
| H4 | $a^{0}, a^{3}, a^{7}, a^{12}$ | 9 |  | $a^{0}, a^{3}, a^{5}, a^{7}$ | 10 |

Table 4.9: Shrimp samples for ranking tests and the number of panellists in each storage experiment.

The rankings provided by the panellists are gathered in Table A.8.

### 4.2.5. Atlantic salmon

Atlantic salmon, caught in the North Atlantic Ocear ${ }^{12}$, was gutted, filleted and skinned at a commercial seafood processing company, without the use of any additives. Fresh fillets, weighing around $1-1.5 \mathrm{~kg}$, were delivered after four days from harvest to FMFP at Ghent University in Styrofoam boxes, wrapped in plastic and covered with crushed ice. The salmon fillets were first manually mixed by moving them around in the boxes for one minute to ensure homogeneous contamination on the surfaces of each fillet and were each then divided into five samples (ca. $203 \pm 2 \mathrm{~g}$ each) and packaged in 610 ml tray: $4^{13}$,

[^11]These packages were stored under different MAP conditions at $4^{\circ} \mathrm{C}$, where the notation of the MAP conditions indicates high (H), medium (M) or low (L) O2 content, O 2 content in air (A), and anaerobic condition (AN) with no O2 content and with high (ANH) CO2 content, as follows:

- L4: $60 / 5 / 35(\mathrm{CO} 2 / \mathrm{O} 2 / \mathrm{N} 2)$ at $4 \pm 0.7^{\circ} \mathrm{C}$;
- M4: $60 / 21 / 19(\mathrm{CO} 2 / \mathrm{O} 2 / \mathrm{N} 2)$ at $4 \pm 0.7^{\circ} \mathrm{C}$;
- H4: 60/40/0 (CO2/O2/N2) at $4 \pm 0.7^{\circ} \mathrm{C}$;
- AN4: 0/0/100 (CO2/O2/N2) at $4 \pm 0.7^{\circ} \mathrm{C}$;
- ANH4: 60/0/40 (CO2/O2/N2) at $4 \pm 0.7^{\circ} \mathrm{C}$;
- A4: air at $4 \pm 0.7^{\circ} \mathrm{C}$.

For each individual storage experiment, trays were randomly sampled on specific storage days for quantification of VOCs (Section 4.3) and were then frozen at $-32^{\circ} \mathrm{C}$ under vacuum ${ }^{14}$. The sampling of the salmon samples was performed on the storage days shown in Table 4.10

|  | Storage days |  |  |
| :---: | :---: | :---: | :---: |
| Experiment | Labelling tests | Ranking tests | Scoring tests |
| L4 | $1,5,7,11$ | $1,5,7,11$ |  |
| M4 | $1,5,9,11$ | $1,5,9,11$ |  |
| H4 | $1,3,5,7,9,11$ | $1,3,5,7,9,11$ |  |
| AN4 | $1,3,5,7,9,11,13$ | $1,3,5,7,9,11,13$ |  |
| AN4 | - | $1,2,3,4,5,6,7,8$ | $1,2,3,4,5,6,7,8$ |
| ANH4 | $1,3,5,7,9,11$ | $1,3,5,7,9,11$ |  |
| A4 | $1,3,5,7,9,11$ | $1,3,5,7,9,11$ |  |

Table 4.10: Sampling salmon samples packaged at day 0 for quantification of VOCs and sensory evaluation (labelling, ranking and scoring) tests. Experiment AN4 denoted with * consists of scoring tests instead of labelling tests.

Sensory evaluation, namely, ranking and labelling tests, was then performed on the salmon samples in several experiments. It must be noted that the samples used for the labelling tests are different from those used for the ranking tests. In addition, scoring and ranking tests were then performed on the salmon samples in storage experiment AN4*. It must be noted that the samples used for the scoring tests are different from those used for the ranking tests. In these experiments, the samples on the day of evaluation were thawed at $2^{\circ} \mathrm{C}$ overnight, cut to $5.0 \pm 0.1 \mathrm{~g}$ portions and

[^12]presented to the panellists at $4^{\circ} \mathrm{C}$ in odour-free and transparent plastic cups, closed with lids ${ }^{15}$ and labelled with three-digit random codes according to a 4 -sample Latin Square Design. Sensory evaluation was based on olfactory evaluation and performed in individual booths under red light (SensoLab UGent).

## Labelling

The labelling test was carried out for salmon samples for each MAP condition separately. Similar to the ranking test, storage experiments H4, AN4, ANH4 and A4 were divided into four sessions in which four samples were studied in each session. Note that samples in all the sessions are different. The samples provided to the panellists and the number of panellists in each ranking test are described in Table 4.11

| Experiment | Session 1 |  | Session 2 |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Samples | Panellists | Samples | Panellists |
| H4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 8 | $a^{1}, a^{3}, a^{5}, a^{7}$ | 8 |
| AN4 | $a^{1}, a^{5}, a^{9}, a^{13}$ | 12 | $a^{3}, a^{7}, a^{9}, a^{11}$ | 12 |
| ANH4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 9 | $a^{1}, a^{3}, a^{5}, a^{7}$ | 5 |
| A4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 11 | $a^{1}, a^{3}, a^{5}, a^{7}$ | 10 |
| L4 | $a^{1}, a^{5}, a^{7}, a^{11}$ | 9 |  |  |
| M4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 8 |  |  |
|  | Session 3 |  | Session 4 |  |
| Experiment | Samples | Panellists | Storage days | Panellists |
| H4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 9 | $a^{3}, a^{5}, a^{7}, a^{9}$ | 10 |
| AN4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 8 | $a^{1}, a^{3}, a^{5}, a^{7}$ | 9 |
| ANH4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 8 | $a^{1}, a^{3}, a^{5}, a^{7}$ | 9 |
| A4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 9 | $a^{1}, a^{3}, a^{5}, a^{7}$ | 9 |

Table 4.11: Salmon samples for labelling tests and the number of panellists in each storage experiment.

The panellists were asked to assign to each sample provided to them a label ("Spoiled" (SP), "Marginal" (M), "Satisfactory" (S), "Fresh" (F), "Very Fresh" (VF)) on an ordinal scale, such that: SP $\prec \mathrm{M} \prec \mathrm{S} \prec \mathrm{F} \prec \mathrm{VF}$. The labels assigned by the panellists are gathered in Table A.9.

## Ranking

The ranking tests were carried out for salmon samples for each MAP condition separately. Storage experiments H4, AN4, ANH4 and A4 were divided into four sessions in which four samples were studied in each session. Note that samples in all the sessions are different. The samples provided to the panellists and the number of panellists in each ranking test are described in Table 4.12 .

[^13]| Experiment | Session 1 |  | Session 2 |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Samples | Panellists | Samples | Panellists |
| H4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 9 | $a^{1}, a^{3}, a^{5}, a^{7}$ | 8 |
| AN4 | $a^{1}, a^{5}, a^{9}, a^{13}$ | 12 | $a^{3}, a^{7}, a^{9}, a^{11}$ | 12 |
| ANH4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 9 | $a^{1}, a^{3}, a^{5}, a^{7}$ | 5 |
| A4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 10 | $a^{1}, a^{3}, a^{5}, a^{7}$ | 9 |
| L4 | $a^{1}, a^{5}, a^{7}, a^{11}$ | 9 |  |  |
| M4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 8 |  |  |
|  | Session 3 |  | Session 4 |  |
| Experiment | Samples | Panellists | Samples | Panellists |
| H4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 9 | $a^{3}, a^{5}, a^{7}, a^{9}$ | 10 |
| AN4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 7 | $a^{1}, a^{5}, a^{5}, a^{7}$ | 8 |
| ANH4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 8 | $a^{1}, a^{3}, a^{5}, a^{7}$ | 9 |
| A4 | $a^{1}, a^{5}, a^{9}, a^{11}$ | 9 | $a^{1}, a^{3}, a^{5}, a^{7}$ | 9 |

Table 4.12: Salmon samples for ranking tests and the number of panellists in each storage experiment.

The panellists were asked to rank the samples according to their perceived freshness from least fresh to most fresh. The rankings provided by the panellists are gathered in Table A. 10 .

## Scoring and ranking with ties

The scoring and ranking tests for experiment AN4* were carried out for salmon samples from four fresh salmon fillets (A, B, C and D). The samples were selected in the order described in Table 4.13 .

| Storage day | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sample | A | A,B | A,B,C | A,B,C,D | A,B,C,D | B,C,D | C,D | D |

Table 4.13: The procedure to select at different storage days the samples from salmon fillets A, B, C and D packaged at day 0 .

To gather sensory evaluation data, a number of panellists (nine or ten depending on the day) were recruited from FMFP at Ghent University with previous training in evaluating the overall quality of salmon and were considered as trained panellists. A number of panellists (between 23 and 28 depending on the day) were recruited from multiple departments at the Faculty of Bioscience Engineering with no prior experience in sensory evaluation of salmon and were considered as untrained panellists.

The different samples of salmon fillets are differentiated by adding a superindex representing the corresponding storage day. Four samples were grouped, one sample
from each fillet, into five groups and were provided to the panellists at random in the order shown in Table 4.14. The groups of samples were not provided in chronological order to prevent the panellists from recognizing a pattern in the experiments that could affect their evaluations.

| Group | Samples | Day |
| :---: | :---: | :---: |
| 1 | $\left(\mathrm{~A}^{1}, \mathrm{~B}^{2}, \mathrm{C}^{3}, \mathrm{D}^{4}\right)$ | Tuesday |
| 2 | $\left(\mathrm{~A}^{2}, \mathrm{~B}^{3}, \mathrm{C}^{4}, \mathrm{D}^{5}\right)$ | Thursday |
| 3 | $\left(\mathrm{~A}^{3}, \mathrm{~B}^{4}, \mathrm{C}^{5}, \mathrm{D}^{6}\right)$ | Monday |
| 4 | $\left(\mathrm{~A}^{4}, \mathrm{~B}^{5}, \mathrm{C}^{6}, \mathrm{D}^{7}\right)$ | Wednesday |
| 5 | $\left(\mathrm{~A}^{5}, \mathrm{~B}^{6}, \mathrm{C}^{7}, \mathrm{D}^{8}\right)$ | Friday |

Table 4.14: The order of grouping the salmon samples from different storage days (represented by the corresponding superindex) and the day of the week each group was provided to the panellists.

The trained panellists were each provided with a group of four samples of salmon, one sample at a time. Since the trained panellists were experienced enough to assign to a score on a 5 -point scale, the panellists were asked to assign to each sample a score on the 5 -point scale shown in Figure 4.2. In Figure 4.2, the scores are equidistantly spaced from each other, and the labels "Spoiled", "Fresh" and "Neither spoiled nor fresh" are provided as anchoring labels on the extreme scores " 1 " and " 5 " and the intermediate score " 3 ", respectively, to point to the trained panellists the representation of these scores. It is important to note that in the scoring test only scores are assigned to each sample, and subsequently collected. The untrained panellists were each asked to express a ranking with ties of the four samples of salmon by ordering them from least fresh to most fresh.


Figure 4.2: 5 -point scale used by panellists, where the extreme scores of " 1 " and " 5 " represent spoiled and fresh, respectively, and the intermediate score of " 3 " represents a neutral response of neither spoiled nor fresh.

The scores provided by the trained panellists are gathered in Table A. 11 and the rankings with ties provided by the untrained panellists in Table A.12.

### 4.3. Quantification of spoilage-related VOCs

In this section, we describe the use of Selective Ion Flow Tube-Mass Spectrometry (SIFT-MS) as a tool to monitor spoilage metabolites in packaged food. We describe the use of SIFT-MS on the aforementioned foods, namely, chicken breasts, cod, brown shrimp and salmon.

The VOCs were quantified from the package headspace using SIFT-MS (Voice 200, Syft Technologies ${ }^{\text {TM }}$, Christchurch, New Zealand) operating in multiple ion mode (MIM) for each sample of the aforementioned foods. A non-destructive technique was performed that allowed the measurement of the VOCs without opening the package, where PTFE/Silicone septa were applied on the two opposite corners of each tray, one for the inlet of the SIFT-MS and the other for a syringe, creating an open system that prevented hypo-pressure within the tray while measuring, as shown in Fig. 4.3. As the VOCs were introduced through the inlet of the SIFT-MS, the syringe allowed environmental air inside the package to compensate for the amount of gas lost by the measurement. The gas, containing the VOCs, that was introduced through a heated inlet into the flow tube reacted with precursor ions $\left(\mathrm{H}_{3} \mathrm{O}^{+}, \mathrm{NO}^{+}, \mathrm{O}_{2}^{+}\right)$resulting in ionized masses. These masses were then detected by a mass-spectrometer at the end of the flow tube. The headspace of each package was sampled at a flow rate of $30 \mathrm{ml} / \mathrm{min}$ for 90 s . After every measurement, the syringe used for the SIFT-MS was changed and flushed with compressed air to remove any VOCs that may contaminate future measurements. Laboratory air was routinely measured before every experiment to check for the out-of-the-ordinary existence of any VOCs.

Figure 4.3: The setup for monitoring spoilage metabolites in packaged chicken using syringes for the inlet of the SIFT-MS and for creating an open system.


The VOCs were selected on the basis of literature survey and previous research
regarding spoilage of the aforementioned foods [38, 52, 111. The average concentration of each product ion was determined by sampling the package headspace for approximately 300 s through a septum inserted on the package lid. To avoid package collapse as well as contamination from the external atmosphere, the headspace was connected with a needle to a bag filled with $100 \%$ N2. Prior to analysing the samples, the flow rate was measured on each day with Gilibrator-2 (Sensidyne, St. Petersburg, FL, USA) and all measured concentrations were corrected according to it. Relative standard deviations were calculated for each product ion, and one product ion was selected for quantifying each VOC according to:

1. average relative standard deviation $<25 \%$ during a scan,
2. high branching ratio, and
3. minimum amount of mass overlaps

If multiple product ions fulfilled the aforementioned criteria equally, the one with the lowest concentration was selected.

Based on the developed procedure, 23 compounds were analysed in packaged chicken breasts, 20 compounds were analysed in packaged cod, 20 compounds were analysed in packaged brown shrimp, and 25 compounds were analysed in packaged salmon. The reported results are summarized in Table B.1, Table B.4 Table B. 7 and Table B. 10 for chicken breasts, cod, brown shrimp and salmon, respectively.

The VOCs were targeted through multiple ion monitoring mode (MIM), while their quantification was carried out using the reaction rate coefficients and the branching ratios of the reaction between the selected VOCs and the precursor ions that generated specific ionized masses. The results are reported in Table B. 2 (for chicken samples used for labelling tests), Table B.3 (for chicken samples used for ranking tests), Table B. 5 (for cod samples used for labelling tests), Table B. 6 (for cod samples used for ranking tests), Table B.8 (for shrimp samples used for labelling tests), Table B. 9 (for shrimp samples used for ranking tests), Table B. 13 (for salmon samples used for labelling tests), Table B.11(for salmon samples used for ranking tests) and Table B. 14 (for salmon samples used for scoring and ranking tests).

Note that all the measured concentrations are averages of measurements performed on three replicates of each sample for every food product, except for the salmon samples used in experiment AN4* (scoring and ranking with ties).

## PART III

## ANALYSIS OF SENSORY <br> DATA

## 5 Aggregation of ordinal labels

## Table of Contents



### 5.1. Introduction

Consensus labels are usually assigned based on one of the following three simple methods. The first is the method of the mode, often used by researchers to assign to each sample the label expressed by the largest number of panellists. The second is the method of the median, and its value is considered as that which separates the lower half from the upper half of the assigned labels. And, finally, the arithmetic mean, which is used when labels are identified with numbers (usually assumed to be equidistant). When consensus labels are generated from arithmetic values there
is one important criteria or assumption that we should be aware of: the existence of a certain notion of distance between labels.

Actually, the three methods can be understood as the search for the label that minimizes the distance to all labels expressed by the trained panellists (considering the zero-one distance function for the mode, the absolute distance function for the median and the squared difference function for the mean). Unfortunately, when these labels represent abstract concepts, as mentioned above, this distance assumption may be too strong. Dealing with perceptions is not an easy task. The search for new methods reducing the reliance on the chosen distance function thus becomes of utmost importance.

In this chapter, we propose a new method for obtaining the consensus label of each sample (from now on referred to as consensus labelling of multiple samples) based on the search for a consensus state. In the field of social choice theory, where voters almost never unanimously agree, different types of consensus states have been analysed [27, 112]. In that context, a consensus state is a set of voting results, where (even when unanimity does not hold) the result of the election is still clear. In our context, the labels expressed by the trained panellists are said to be in (or to belong to) a consensus state if determining the consensus labelling is obvious.

Monometrics [27, which are closely related to distance functions, will be a key tool for measuring how close the labels expressed by the trained panellists are said to be in (or belong to) the chosen consensus state. Here, five consensus states are discussed: unanimity, majority, marginal majority, marginal monotonicity and monotonicity. We advocate for the use of the latter, which is similar to that of monotonicity of a profile of rankings discussed in [113] for the aggregation of rankings, resulting in a method for simultaneously exploiting the labels expressed for all different samples, something that was not previously considered. A consensus labelling of the samples obtained by exploiting the information expressed for all samples simultaneously will be referred to as joint consensus labelling.

In this chapter, we will answer the following question:
Question III.1: How can we assign (joint) consensus ordinal labels?
We will discuss the methods for aggregating ordinal labels assigned by trained panellists to food samples to obtain a consensus labelling. The application of these methods will be illustrated on the sensory data in Chapter 4 . Since ordinal labels are also assigned to samples other than food, we explore applications typically found in movie recommendation systems to determine the consensus rating of a movie and on online dating sites to provide the consensus ordinal label of a candidate profile.

### 5.2. Assigning a consensus label

Let $\mathscr{A}=\left\{a_{1}, \ldots, a_{n}\right\}$ be a set of $n$ samples and $\mathscr{L}=\left\{L_{1}, \ldots, L_{l}\right\}$ be an ordinal scale equipped with a total order relation $\leq \mathscr{L}$. The labels are indexed in such a way that $L_{l_{1}} \leq \mathscr{L} L_{l_{2}}$, when $l_{1} \leq l_{2}$, for any $l_{1}, l_{2} \in\{1, \ldots, l\}$. The converse of the relation $\leq_{\mathscr{L}}$ is denoted by $\geq \mathscr{L}$.

We consider the setting where $n_{T}$ trained panellists have expressed a label in $\mathscr{L}$ for each sample in $\mathscr{A}$. The goal is to agree on the label that should be assigned to each sample. For any $i \in\left\{1, \ldots, n_{T}\right\}$, we denote by $f_{i}: \mathscr{A} \rightarrow \mathscr{L}$ the label function associating to each sample in $\mathscr{A}$ the label expressed by the $i$-th trained panellist. Any list of $n$ labels where the $j$-th label corresponds to sample $a_{j}$ is referred to as a labelling of the samples. The matrix of $n_{T} \times n$ labels where the label at the $i$-th row and $j$-th column corresponds to the label expressed by the $i$-th trained panellist for sample $a_{j}$ is referred to as the (matrix of) labellings expressed by the trained panellists:

$$
\mathbf{Z}=\left(\begin{array}{ccc}
f_{1}\left(a_{1}\right) & \ldots & f_{1}\left(a_{n}\right) \\
\vdots & \ddots & \vdots \\
f_{n_{T}}\left(a_{1}\right) & \ldots & f_{n_{T}}\left(a_{n}\right)
\end{array}\right)
$$

We denote by $f: \mathscr{A} \rightarrow \mathscr{L}$ the consensus label function assigning to each sample in $\mathscr{A}$ a consensus label in $\mathscr{L}$ according to the labellings expressed by the trained panellists. When the number of samples is small, a (consensus) label function is usually given in the form of a list $\left(f\left(a_{1}\right), \ldots, f\left(a_{n}\right)\right)$.

In mathematics, the notion of distance function or metric is a well-known concept, and is defined as follows:

Definition 5.1 (Distance function). A function $\partial: \mathscr{A} \times \mathscr{A} \rightarrow \mathbb{R}$ is called a distance function (on the set $\mathscr{A}$ ) if the following conditions are met:
(i) Non-negativity: for any $a, b \in \mathscr{A}$, it holds that

$$
\partial(a, b) \geq 0 .
$$

(ii) Coincidence: for any $a, b \in \mathscr{A}$, it holds that

$$
\partial(a, b)=0 \quad \Leftrightarrow \quad a=b .
$$

(iii) Symmetry: for any $a, b, c \in \mathscr{A}$, it holds that

$$
\partial(a, b)=\partial(b, a) .
$$

(iv) Triangle inequality: for any $a, b, c \in \mathscr{A}$, it holds that

$$
\partial(a, c) \leq \partial(a, b)+\partial(b, c)
$$

The simplest and most widely used methods for assigning a consensus label are based on the analysis of the frequency distribution of the labels expressed by the trained panellists [1]. Usually, the most frequent label expressed by the trained panellists (the mode) is assigned as the consensus label [114]. This procedure is equivalent to assigning the label that minimizes the sum of the zero-one distances to the labels expressed by the trained panellists, where the zero-one distance function is defined as $\partial_{0}\left(L_{l_{1}}, L_{l_{2}}\right)=0$, if $l_{1}=l_{2}$, and $\partial_{0}\left(L_{l_{1}}, L_{l_{2}}\right)=1$, otherwise. The choice of the median, which is the label that separates the higher half from the lower half of the given labels, is also common in practice [1]. This procedure is equivalent to assigning the label that minimizes the sum of absolute distances to the labels expressed by the trained panellists, where the absolute distance function is defined by $\partial_{1}\left(L_{l_{1}}, L_{l_{2}}\right)=\left|l_{1}-l_{2}\right|$. Often, the labels are identified by indices, and the available information is treated quantitatively by calculating the (rounded) arithmetic mean of the indices of the labels expressed by the trained panellists 1, 40. This procedure is equivalent to assigning the label that minimizes the sum of the squared difference to the labels expressed by the trained panellists, where the squared difference function is defined by $\partial_{2}\left(L_{l_{1}}, L_{l_{2}}\right)=\left(l_{1}-l_{2}\right)^{2}$. A generalization of the median and the mean is equivalent to assigning the label that minimizes the sum of the $p$-th order difference to the labels expressed by the trained panellists, where the $p$-th order difference function is defined by $\partial_{p}\left(L_{l_{1}}, L_{l_{2}}\right)=\left|l_{1}-l_{2}\right|^{p}$, for $p>0$.

Note that the use of the absolute distance function or the squared difference function on $\mathscr{L}$ assumes that all labels are equidistant. However, when these labels represent abstract notions, such as "Bad", "Acceptable", or "Good", this may be too strong an assumption. Wichchukit et al. [115] argue that it is preferable to represent the labellings expressed by the trained panellists using a histogram. A histogram clearly displays the frequency distribution of the labels and highlights any central tendency. Along this line of research in which the entire frequency distribution is analysed, we discuss a new method for assigning a consensus label to each sample that does not rely on the 'distance' between the different labels, but, instead, better exploits the information expressed by the trained panellists, resulting in a joint consensus labelling of all samples at the same time.

Although the distances between labels could not be determined accurately, perhaps due to a qualitative nature of the scale, there always exists the natural notion of a label being in between two other labels. This notion is captured by a ternary relation called betweenness relation [116, 117].

Definition 5.2 (Betweenness relation). A ternary relation $\mathcal{R}$ on a set $\mathscr{A}$ is called
a betweenness relation on $\mathscr{A}$ if it satisfies the following two properties:
(i) Symmetry in the end points: for any $a, b, c \in \mathscr{A}$, it holds that

$$
(a, b, c) \in \mathcal{R} \quad \Leftrightarrow \quad(c, b, a) \in \mathcal{R} .
$$

(ii) Closure: for any $a, b, c \in \mathscr{A}$, it holds that

$$
((a, b, c) \in \mathcal{R} \wedge(a, c, b) \in \mathcal{R}) \quad \Leftrightarrow \quad b=c .
$$

Interestingly, if, for any $a, b \in \mathscr{A}$, it always holds that $(b, a, a) \in \mathcal{R}$ due to the closure axiom, then $(a, a, b) \in \mathcal{R}$ due to the symmetry in the end points. Thus, since it always holds that $(a, a, b) \in \mathcal{R}$, due to the closure axiom, we conclude that if it holds that $(a, b, a) \in \mathcal{R}$, then $a=b$.

For any betweenness relation $\mathcal{R}$, the fact that a triplet $(a, b, c)$ belongs to $\mathcal{R}$ is referred to as ' $b$ is in between $a$ and $c$ '.

In this chapter, the betweenness relations considered are always induced by an order relation [118].

Proposition 5.1. For any total order relation $\leq$ on a set $\mathscr{A}$, the following two statements hold:
(i) The relation $\mathcal{R}_{\leq}$defined as

$$
\mathcal{R}_{\leq}=\left\{(a, b, c) \in \mathscr{A}^{3} \mid \min (a, c) \leq b \leq \max (a, c)\right\}
$$

is a betweenness relation on $\mathscr{A}$.
(ii) The relation $\mathcal{R}_{\leq^{n}}$ defined as

$$
\mathcal{R}_{\leq^{n}}=\left\{(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in\left(\mathscr{A}^{n}\right)^{3} \mid(\forall i \in\{1, \ldots, n\})\left(\left(a_{i}, b_{i}, c_{i}\right) \in \mathcal{R}_{\leq}\right)\right\}
$$

is a betweenness relation on $\mathscr{A}^{n}$.
Here, we are interested in the betweenness relation induced by the total order relation $\leq_{\mathscr{L}}$ on the ordinal scale $\mathscr{L}$. In that way, any label of the ordinal scale is considered to be in between two other labels if it is greater than or equal to the minimum of the two labels and smaller than or equal to the maximum of the two labels.

A function satisfying the axioms of non-negativity and coincidence of a distance function and, at the same time, an axiom of compatibility with a betweenness relation is called a monometric [27]. Thus, a monometric is closely related to a distance function or metric, but does not impose symmetry nor the triangle inequality. Therefore, not every monometric is a distance function, and vice versa.

Definition 5.3 (Monometric). Let $\mathcal{R}$ be a betweenness relation on a set $\mathscr{A}$. A function $M: \mathscr{A} \times \mathscr{A} \rightarrow \mathbb{R}$ is called a monometric on $\mathscr{A}$ (w.r.t. $\mathcal{R}$ ) if it satisfies the following three properties:
(i) Non-negativity: for any $a, b \in \mathscr{A}$, it holds that

$$
M(a, b) \geq 0
$$

(ii) Coincidence: for any $a, b \in \mathscr{A}$, it holds that

$$
M(a, b)=0 \Leftrightarrow a=b .
$$

(iii) Compatibility: for any $a, b, c \in \mathscr{A}$ such that $(a, b, c) \in \mathcal{R}$, it holds that

$$
M(a, b) \leq M(a, c)
$$

For any two elements $a, b \in \mathscr{A}, M(a, b)$ is referred to as the cost of changing the element $a$ into the element $b$.

In order to measure the cost of changing a label of the ordinal scale into another one, we consider a monometric $M$ on $\mathscr{L}$ w.r.t. the betweenness relation $\mathcal{R}_{\leq \mathscr{L}}$. The axiom of compatibility w.r.t. this betweenness relation is illustrated in Fig. 5.1.


Figure 5.1: Natural interpretation of the use of a monometric on the ordinal scale $\mathscr{L}=\left\{L_{1}, L_{2}, L_{3}, L_{4}, L_{5}\right\}$.

Remark The zero-one distance function, the absolute distance function and the squared difference function are three natural examples of a monometric w.r.t. the betweenness relation induced by $\leq \mathscr{L}$.

The choice of a proper monometric is not an easy task. Actually, there is no truly best monometric when dealing with a qualitative scale. In several fields of application, the choice of a monometric ${ }^{1}$ is subject to a preceding thorough

[^14]validation by a group of trained panellists. For instance, we refer to a survey that shows different ways of dealing with the (9-point) hedonic scale [115].

A natural method for assigning a consensus label to a sample $a_{j} \in \mathscr{A}$ based on the labellings expressed by the trained panellists is then defined by means of a monometric $M$ on $\mathscr{L}$ w.r.t. $\mathcal{R}_{\leq \mathscr{L}}$ :

$$
\begin{equation*}
f\left(a_{j}\right)=\arg \min _{L \in \mathscr{L}} \sum_{i=1}^{n_{T}} M\left(f_{i}\left(a_{j}\right), L\right) \tag{5.1}
\end{equation*}
$$

In case $M$ is the zero-one distance function, the absolute distance function or the squared difference function, the consensus labelling of the samples coincides with that obtained by the usual methods for assigning a consensus labelling mentioned in the beginning of this section.

Remark In real-life problems, it is common to have a list of weights $\left(w_{i}\right)_{i=1}^{n_{T}}$ (adding up to one) weighing the performance of each trained panellist. In this case, the following alternative to Eq. (5.1) is considered:

$$
f\left(a_{j}\right)=\arg \min _{L \in \mathscr{\mathscr { L }}} \sum_{i=1}^{n_{T}} w_{i} M\left(f_{i}\left(a_{j}\right), L\right)
$$

Note that each trained panellist is actually expressing a label for each sample in $\mathscr{A}$ and that, due to several factors, the labellings of the different samples might interact with each other. This is a common problem in, for instance, smell experiments where trained panellists might not be able to express an appropriate label after smelling more than four to six food products during a short period of time due to olfactory fatigue [24]. This is also a common problem in the field of recommender systems, such as movie recommendation and online dating systems, where personality is an important factor that influences the labelling of different movies or profiles of candidates [119, 120]. In addition to personality, social and environmental contexts as well as visual experience and exposure are important sources of individual differences in online dating systems [119, 121]. Therefore, it seems natural to simultaneously exploit the information provided for all samples when assigning a consensus labelling of these samples.

With the intention of assigning a consensus label to all samples simultaneously, a monometric $\mathbf{M}: \mathscr{L}^{n} \times \mathscr{L}^{n} \rightarrow \mathbb{R}$ w.r.t. $\mathcal{R}_{\leq_{\mathscr{L}^{n}}}$ needs to be fixed. In that way, the consensus label assigned to a sample $a_{j} \in \mathscr{A}$ based on the labellings expressed by the trained panellists is defined as

$$
f\left(a_{j}\right)=\left(\arg \min _{\mathbf{z} \in \mathscr{L}^{n}} \sum_{i=1}^{n_{T}} \mathbf{M}\left(\left(f_{i}\left(a_{1}\right), \ldots, f_{i}\left(a_{n}\right)\right), \mathbf{z}\right)\right)(j)
$$

Obviously, the particular case where $\mathbf{M}$ is defined as

$$
\mathbf{M}\left(\left(f_{i}\left(a_{1}\right), \ldots, f_{i}\left(a_{n}\right)\right), \mathbf{z}\right)=\sum_{j=1}^{n} M\left(f_{i}\left(a_{j}\right), \mathbf{z}(j)\right)
$$

for a given monometric $M$ on $\mathscr{L}$, reduces to the previous setting where the labels of all samples are considered independent w.r.t. each other when calculating the cost.

### 5.3. Consensus state problem

As discussed in [27, the aggregation of rankings can generally be understood as a two-step procedure that measures the closeness to a desired state of the world in which determining the result of the aggregation of the given rankings is obvious. Usually, we refer to these desired states of the world leading to an obvious result of the aggregation process as consensus states.

Here, we adapt this notion of consensus state to our problem setting. A consensus state will now be the set of all matrices of labels for which identifying the consensus labelling of the samples is obvious. Naturally, all matrices of labels in which all rows coincide (i.e., all trained panellists agree on the label that should be assigned to each sample), need to belong to any consensus state. Various consensus states may be considered. Ideally, the chosen consensus state should be linked to a natural way of deciding the consensus labelling of the samples. According to some consensus states, more than one consensus labelling of the samples could be considered 'natural'. For instance, in case we consider the most frequent label for each sample, this natural way of deciding the consensus labelling of the samples might not be unique, resulting in a function where the codomain is the power set $\mathcal{P}\left(\mathscr{L}^{n}\right)$ of $\mathscr{L}^{n}$, rather than $\mathscr{L}^{n}$ itself.

Definition 5.4 (Consensus state). A subset $\mathcal{S}$ of $\mathscr{L}^{n \times n_{T}}$ is called a consensus state if there exists an underlying consensus label function $f: \mathcal{S} \rightarrow \mathcal{P}\left(\mathscr{L}^{n}\right)$ satisfying that, for any $\mathbf{Z} \in \mathscr{L}^{n_{T} \times n}$ such that $\mathbf{Z}_{i_{1} j}=\mathbf{Z}_{i_{2} j}$ for any $i_{1}, i_{2} \in\left\{1, \ldots, n_{T}\right\}$ and any $j \in\{1, \ldots, n\}$, it holds that $\mathbf{Z} \in \mathcal{S}$ and $f(\mathbf{Z})=\left\{\left(\mathbf{Z}_{11}, \ldots, \mathbf{Z}_{1 n}\right)\right\}$.

The goal of a consensus state problem is to find the matrix of labels belonging to the consensus state that is closest to the labellings expressed by the trained panellists. Such a matrix is referred to as a closest matrix of labels (in the chosen consensus state), and will lead to a natural consensus labelling of the samples. Considering a broader consensus state leads to methods that are less dependent on the chosen monometric and more dependent on the consensus state itself. Analogously, considering a narrower consensus state leads to methods that are more dependent on the chosen monometric and less dependent on the consensus
state itself. When dealing with, for instance, qualitative scales, where the choice of the proper monometric is unclear, we aim to reduce the importance of the monometric, while still considering a meaningful consensus state.

In order to search for a closest matrix of labels in the chosen consensus state, the cost of changing the matrix of labels expressed by the trained panellists into a new matrix of labels needs to be defined. This cost is measured by means of a monometric on $\mathscr{L}^{n_{T} \times n}$. In this case, the considered betweenness relation is defined as:

$$
\mathcal{R}_{\leq_{\mathscr{C}^{n} T^{\prime}}}=\left\{\left(\mathbf{Z}^{1}, \mathbf{Z}^{2}, \mathbf{Z}^{3}\right) \in\left(\mathscr{L}^{n_{T} \times n}\right)^{3} \left\lvert\, \begin{array}{l}
\left(\forall i \in\left\{1, \ldots, n_{T}\right\}\right)(\forall j \in\{1, \ldots, n\}) \\
\left(\left(\mathbf{Z}_{i j}^{1}, \mathbf{Z}_{i j}^{2}, \mathbf{Z}_{i j}^{3}\right) \in \mathcal{R}_{\leq_{\mathscr{L}}}\right)
\end{array}\right.\right\}
$$

A closest matrix of labels belonging to the chosen consensus state $\mathcal{S}$ is then defined by means of a monometric $\mathbb{M}$ on $\mathscr{L}^{n_{T} \times n}$ w.r.t. $\mathcal{R}_{\leq_{\mathscr{L}^{n} T^{\times n}}}$ :

$$
\arg \min _{\mathbf{Z} \in \mathcal{S}} \mathbb{M}\left(\left(\begin{array}{ccc}
f_{1}\left(a_{1}\right) & \ldots & f_{1}\left(a_{n}\right)  \tag{5.2}\\
\vdots & \ddots & \vdots \\
f_{n_{T}}\left(a_{1}\right) & \ldots & f_{n_{T}}\left(a_{n}\right)
\end{array}\right), \mathbf{Z}\right)
$$

Remark Note that the minimizer of Eq. 5.2 does not need to be unique.
The consensus labelling is then given by the underlying consensus label function $f: \mathcal{S} \rightarrow \mathcal{P}\left(\mathscr{L}^{n}\right)$ evaluated in the minimizer(s) of Eq. 5.2).

Obviously, the particular case where $\mathbb{M}$ is defined as

$$
\begin{equation*}
\mathbb{M}\left(\mathbf{Z}, \mathbf{Z}^{\prime}\right)=\sum_{i=1}^{n_{T}} \sum_{j=1}^{n} M\left(\mathbf{Z}_{i j}, \mathbf{Z}_{i j}^{\prime}\right) \tag{5.3}
\end{equation*}
$$

for a given monometric ${ }^{2} M$ on $\mathscr{L}$, reduces to the previous setting where all trained panellists and all labels of all samples are considered independent w.r.t. each other when calculating the cost.

As previously mentioned, the choice of the right monometric on $\mathscr{L}$ might be unclear. Here, we propose a monometric that minimizes the number of labels being changed into another label that is far away in the scale, while allowing for a larger number of labels being changed into a neighbouring label. Intuitively, we consider that, for the qualitative scale $\mathscr{L}=\{$ "Bad", "Acceptable", "Good" $\}$, in case the 'true" label of a sample is "Bad", it is more likely that several trained panellists express the label "Acceptable" for this sample rather than a unique trained panellist expressing the

[^15]label "Good". Formally, we aim to minimize, first, the number of labels changed into a label at maximum absolute distance $(l-1)$, second, the number of labels changed into a label at absolute distance $l-2$, third, the number of labels changed into a label at absolute distance $l-3$, etc.

This hierarchical optimization problem is addressed straightforwardly by considering a suitable monometric on $\mathscr{L}$ that assigns a cost to changing each label as follows: the cost assigned to changing a unique label into another label at a certain absolute distance $\partial_{1}$ should be larger than the sum of costs assigned to changing all the $n_{T} \times n$ labels into other labels at an absolute distance strictly smaller than $\partial_{1}$. We do this by considering exponentially increasing costs for each absolute distance. For more details on the hierarchical combination of monometrics, we refer to [27]. The proposed monometric on $\mathscr{L}^{n_{T} \times n}$ is then defined as:

$$
\begin{equation*}
\mathbb{M}\left(\mathbf{Z}, \mathbf{Z}^{\prime}\right)=\sum_{i=1}^{n_{T}} \sum_{j=1}^{n}\left(\left(n_{T} \cdot n+1\right)^{\partial_{1}\left(\mathbf{Z}_{i j}, \mathbf{Z}_{i j}^{\prime}\right)}-1\right) \tag{5.4}
\end{equation*}
$$

### 5.4. Interesting consensus states

In this section, we discuss five natural consensus states: unanimity, majority, marginal majority, monotonicity and marginal monotonicity.

### 5.4.1. Unanimity

The smallest consensus state, called the unanimity consensus state, is the set of all matrices of labels where all the rows coincide, i.e., all trained panellists agree on the label that should be assigned to each sample. Formally, this consensus state is defined as:

$$
\begin{equation*}
\mathcal{S}_{0}=\left\{\mathbf{Z} \in \mathscr{L}^{n_{T} \times n} \mid\left(\forall i_{1}, i_{2} \in\left\{1, \ldots, n_{T}\right\}\right)(\forall j \in\{1, \ldots, n\})\left(\mathbf{Z}_{i_{1} j}=\mathbf{Z}_{i_{2} j}\right)\right\} \tag{5.5}
\end{equation*}
$$

The underlying consensus label function $f: \mathcal{S}_{0} \rightarrow \mathcal{P}\left(\mathscr{L}^{n}\right)$ assigns to each matrix $\mathbf{Z}$ of labels in $\mathcal{S}_{0}$ the (unique) list of labels that appears $n_{T}$ times as row of $\mathbf{Z}$.

The unanimity consensus state is the only consensus state that unquestionably identifies a consensus labelling of the samples. Obviously, in case all trained panellists express the same labelling of the samples, the consensus labelling of these samples is clear. Unfortunately, this unanimity situation almost never happens in real-life problems. Therefore, any method focusing on the search for the unanimity consensus state strongly depends on the chosen monometric. Thus, when dealing
with qualitative scales, where there is no clear cost of changing a label into another one, the need for a broader consensus state than the unanimity consensus state is patent.

### 5.4.2. Majority

Another natural consensus state is the set of all matrices of labels where more than half of the rows coincide, i.e., more than half of the trained panellists agree on the labels that should be assigned to all samples:

$$
\mathcal{S}_{1}=\left\{\begin{array}{l|l}
\mathbf{Z} \in \mathscr{L}^{n_{T} \times n} & \begin{array}{l}
\left(\exists I \subseteq\left\{1, \ldots, n_{T}\right\}\right) \\
\left(\left(|I|>\frac{n_{T}}{2}\right) \wedge\left(\forall i_{1}, i_{2} \in I\right)(\forall j \in\{1, \ldots, n\})\left(\mathbf{Z}_{i_{1} j}=\mathbf{Z}_{i_{2} j}\right)\right)
\end{array} \tag{5.6}
\end{array}\right\} .
$$

The underlying consensus label function $f: \mathcal{S}_{1} \rightarrow \mathcal{P}\left(\mathscr{L}^{n}\right)$ assigns to each matrix $\mathbf{Z}$ of labels in $\mathcal{S}_{1}$ the (unique) list of labels that appears more than $\frac{n_{T}}{2}$ times as row of $\mathbf{Z}$.

Although the majority consensus state is a broader consensus state than the unanimity consensus state, it still focuses too strongly on the need of the trained panellists to agree on their labellings of the samples.

### 5.4.3. Marginal majority

Another natural consensus state is the set of all matrices of labels where for each of the columns more than half of the elements coincide, i.e., more than half of the trained panellists agree on the label that should be assigned to each sample:

$$
\mathcal{S}_{2}=\left\{\begin{array}{l|l}
\mathbf{Z} \in \mathscr{L}^{n_{T} \times n} & \begin{array}{l}
(\forall j \in\{1, \ldots, n\})\left(\exists I \subseteq\left\{1, \ldots, n_{T}\right\}\right) \\
\left(\left(|I|>\frac{n_{T}}{2}\right) \wedge\left(\left(\forall i_{1}, i_{2} \in I\right)\left(\mathbf{Z}_{i_{1} j}=\mathbf{Z}_{i_{2} j}\right)\right)\right)
\end{array} \tag{5.7}
\end{array}\right\} .
$$

The underlying consensus label function $f: \mathcal{S}_{2} \rightarrow \mathcal{P}\left(\mathscr{L}^{n}\right)$ assigns to each matrix $\mathbf{Z}$ of labels in $\mathcal{S}_{2}$ the (unique) list of labels in which the $j$-th label is the one appearing more than $\frac{n_{T}}{2}$ times in the $j$-th column of $\mathbf{Z}$.

Note that this consensus state treats all samples independently and does not exploit the fact that each trained panellist expresses a label for all samples in $\mathscr{A}$.

Obviously, majority implies marginal majority. In particular, it holds that $\mathcal{S}_{0} \subseteq$ $\mathcal{S}_{1} \subseteq \mathcal{S}_{2}$. In case the set of samples is a singleton, majority and marginal majority coincide, i.e., it holds that $\mathcal{S}_{1}=\mathcal{S}_{2}$ if $k=1$.

### 5.4.4. Monotonicity

We consider the setting where each trained panellist expresses a label for each of the samples in $\mathscr{A}$. Under the assumption that the closer a list of labels is to the 'true' labelling of the samples, the likelier it is that a trained panellist expresses this list of labels, we can expect the number of trained panellists expressing each list of labels to decrease whenever we deviate from the 'true' labelling of these samples.

Note that any list of labels $\mathbf{z} \in \mathscr{L}^{n}$ induces a natural order relation on $\mathscr{L}^{n}$ :

$$
\sqsubseteq_{\mathbf{z}}=\left\{\left(\mathbf{z}_{1}, \mathbf{z}_{2}\right) \in\left(\mathscr{L}^{n}\right)^{2} \mid(\forall j \in\{1, \ldots, n\})\left(\left(L_{j}, L_{j}^{1}, L_{j}^{2}\right) \in \mathcal{R}_{\leq_{\mathscr{L}}}\right)\right\} .
$$

The fact that $\left(\mathbf{z}_{1}, \mathbf{z}_{2}\right) \in \sqsubseteq_{\mathbf{z}}$ is denoted by $\mathbf{z}_{2} \sqsubseteq_{\mathbf{z}} \mathbf{z}_{1}$ and can be understood as $\mathbf{z}_{1}$ being closer to $\mathbf{z}$ than $\mathbf{z}_{2}$.

Remark For any $\mathbf{z} \in \mathscr{L}^{n}$, the relation $\sqsubseteq_{\mathbf{z}}$ defines an order relation on $\mathscr{L}^{n}$. Due to the fact that $\leq_{\mathscr{L}}$ is an order relation, $\sqsubseteq_{\mathbf{z}}$ trivially is reflexive and antisymmetric. Transitivity is proved easily in the following way. Consider $\mathbf{z}_{1}, \mathbf{z}_{2}, \mathbf{z}_{3} \in \mathscr{L}^{n}$ such that $\mathbf{z}_{2} \sqsubseteq_{\mathbf{z}} \mathbf{z}_{1}$ and $\mathbf{z}_{3} \sqsubseteq_{\mathbf{z}} \mathbf{z}_{2}$. Therefore, for any $j \in\{1, \ldots, n\}$, it holds that ( $L_{j} \leq \mathscr{L}$ $\left.L_{j}^{1} \leq \mathscr{L} L_{j}^{2} \vee L_{j}^{2} \leq_{\mathscr{L}} L_{j}^{1} \leq_{\mathscr{L}} L_{j}\right)$ and that $\left(L_{j} \leq_{\mathscr{L}} L_{j}^{2} \leq_{\mathscr{L}} L_{j}^{3} \vee L_{j}^{3} \leq_{\mathscr{L}} L_{j}^{2} \leq \mathscr{\mathscr { L }} L_{j}\right)$. We distinguish four cases:
(i) $L_{j} \leq_{\mathscr{L}} L_{j}^{1} \leq \mathscr{L} L_{j}^{2}$ and $L_{j} \leq \mathscr{L} L_{j}^{2} \leq \mathscr{L} L_{j}^{3}$. It trivially follows that

$$
L_{j} \leq_{\mathscr{L}} L_{j}^{1} \leq_{\mathscr{L}} L_{j}^{3}
$$

(ii) $L_{j} \leq_{\mathscr{L}} L_{j}^{1} \leq_{\mathscr{L}} L_{j}^{2}$ and $L_{j}^{3} \leq_{\mathscr{L}} L_{j}^{2} \leq_{\mathscr{L}} L_{j}$. It follows that $L_{j}^{1}=L_{j}^{2}$ and, therefore, that

$$
L_{j}^{3} \leq_{\mathscr{L}} L_{j}^{1} \leq_{\mathscr{L}} L_{j}
$$

(iii) $L_{j}^{2} \leq_{\mathscr{L}} L_{j}^{1} \leq_{\mathscr{L}} L_{j}$ and $L_{j} \leq_{\mathscr{L}} L_{j}^{2} \leq_{\mathscr{L}} L_{j}^{3}$. It follows that $L_{j}^{1}=L_{j}^{2}$ and, therefore, that

$$
L_{j} \leq \mathscr{L} L_{j}^{1} \leq \mathscr{L} L_{j}^{3}
$$

(iv) $L_{j}^{2} \leq \mathscr{L} L_{j}^{1} \leq \mathscr{L} L_{j}$ and $L_{j}^{3} \leq \mathscr{L} L_{j}^{2} \leq \mathscr{L} L_{j}$. It trivially follows that

$$
L_{j}^{3} \leq_{\mathscr{L}} L_{j}^{1} \leq_{\mathscr{L}} L_{j}
$$

We conclude that $\sqsubseteq_{\mathbf{z}}$ is transitive and, therefore, an order relation on $\mathscr{L}^{n}$.
Figure 5.2 displays a graphical representation of the Hasse diagram of $\sqsubseteq_{\mathbf{z}}$ for a set of two samples $\}^{3}$ and the qualitative scale $\mathscr{L}=\{B, A, G\}$, where $B$ means "Bad",
${ }^{3}$ Note that the cardinality of the set of samples usually is greater than two and this small set of samples is only considered for illustrating the Hasse diagram of $\sqsubseteq_{\mathbf{z}}$.

A means "Acceptable" and $G$ means "Good". Obviously, the considered (strict) total order relation is given by $B \prec A \prec G$.


Figure 5.2: Hasse diagram of $\sqsubseteq_{\mathbf{z}}$ for $\mathbf{z}=(G, G)$ (left) and $\mathbf{z}=(A, G)$ (right).

The number of trained panellists expressing each list of labels should be decreasing on the Hasse diagram of $\sqsubseteq_{\mathbf{z}}$ for the list of labels $\mathbf{z}$ associated with the 'true' labelling of the samples. In such a case, the matrix of labels expressed by the trained panellists is monotone w.r.t. the list of labels $\mathbf{z}$.

Example 5.1. Consider the set $\mathscr{A}=\left\{a_{1}, a_{2}\right\}$, the ordinal scale $\mathscr{L}=\{A, B, G\}$ and $n_{T}=12$ trained panellists, where the (transposed) matrix of labellings expressed by the trained panellists is the following:

$$
\mathbf{Z}^{\top}=\left(\begin{array}{lllllllllll}
G & G & G & A & A & G & G & B & A & G & A
\end{array}\right)
$$

For simplicity, we now consider the 'true' labelling of the samples to be $(G, G)$, and plot the Hasse diagram of $\sqsubseteq_{\mathbf{z}}$ for the list of labels $\mathbf{z}$ associated with $(G, G)$ in Figure 5.3. We can see that there is monotonicity w.r.t. the list of labels $(G, G)$, where the number of trained panellists expressing each list of labels is decreasing from top to bottom of the Hasse diagram.

Therefore, we conclude that the matrix of labels expressed by the trained panellists is called monotone w.r.t the list of labels $(G, G)$.

The set of all matrices of labels that are monotone w.r.t. at least one list of labels leads to the following natural consensus state:

$$
\mathcal{S}_{3}=\left\{\begin{array}{l|l}
\mathbf{Z} \in \mathscr{L}^{n_{T} \times n} & \begin{array}{l}
\left(\exists \mathbf{z} \in \mathscr{L}^{n}\right)\left(\forall \mathbf{z}_{1}, \mathbf{z}_{2} \in \mathscr{L}^{n}\right) \\
\left(\mathbf{z}_{2} \sqsubseteq_{\mathbf{z}} \mathbf{z}_{1} \Rightarrow n_{\mathbf{Z}}\left(\mathbf{z}_{2}\right) \leq n_{\mathbf{Z}}\left(\mathbf{z}_{1}\right)\right)
\end{array} \tag{5.8}
\end{array}\right\}
$$



Figure 5.3: Hasse diagram of $\sqsubseteq_{\mathbf{z}}$ for $\mathbf{z}=(G, G)$.
where, for any $\mathbf{z}^{\prime} \in \mathscr{L}^{n_{T}}, n_{\mathbf{Z}}\left(\mathbf{z}^{\prime}\right)$ denotes the number of times the list of labels $\mathbf{z}^{\prime}$ appears as row of $\mathbf{Z}$.

The underlying consensus label function $f: \mathcal{S}_{3} \rightarrow \mathcal{P}\left(\mathscr{L}^{n}\right)$ assigns to each matrix $\mathbf{Z}$ of labels in $\mathcal{S}_{3}$ the set of all lists of labels w.r.t. which it is monotone.

Example 5.2. Consider the set $\mathscr{A}=\left\{a_{1}, a_{2}\right\}$, the ordinal scale $\mathscr{L}=\left\{L_{1}, L_{2}, L_{3}, L_{4}, L_{5}\right\}$ and $n_{T}=20$ trained panellists, where the transposed matrix of labellings expressed by the trained panellists is the following:

$$
\mathbf{Z}^{\top}=\left(\begin{array}{llllllllllllllllll}
L_{1} & L_{1} & L_{2} & L_{2} & L_{2} & L_{3} & L_{3} & L_{3} & L_{3} & L_{4} & L_{4} & L_{4} & L_{4} & L_{4} & L_{4} & L_{4} & L_{5} & L_{5} \\
L_{5} & L_{5} \\
L_{3} & L_{4} & L_{3} & L_{4} & L_{4} & L_{3} & L_{4} & L_{4} & L_{5} & L_{2} & L_{3} & L_{4} & L_{4} & L_{4} & L_{4} & L_{5} & L_{3} & L_{4} \\
L_{4} & L_{5}
\end{array}\right)
$$

As can be seen in Fig. 5.4, the number of trained panellists expressing each list of labels decreases when we move away from the list of labels $\left(L_{4}, L_{4}\right)$. Therefore, the matrix of labels expressed by the trained panellists is monotone w.r.t. $\left(L_{4}, L_{4}\right)$.

### 5.4.5. Marginal monotonicity

In some settings, we may consider all the samples to be independent w.r.t. each other and we require monotonicity for each of the samples independently. Under the assumption that the closer a label is to the 'true' label of a sample, the likelier it is that an trained panellist expresses this label for the sample, we can expect the number of trained panellists expressing each label to decrease whenever we deviate from the 'true' label of this sample.


Figure 5.4: Example of a monotone matrix of labels w.r.t. $\left(L_{4}, L_{4}\right)$.

Note that any label $L \in \mathscr{L}$ induces a natural order relation on $\mathscr{L}$ :

$$
\sqsubseteq_{L}=\left\{\left(L^{1}, L^{2}\right) \in \mathscr{L}^{2} \mid\left(L, L^{1}, L^{2}\right) \in \mathcal{R}_{\leq \mathscr{L}}\right\},
$$

which can be seen as a particular case of $\sqsubseteq_{\mathbf{z}}$ for a singleton set of samples.
Figure 5.5 displays a graphical representation of the Hasse diagram of $\sqsubseteq_{L}$ for the qualitative scale $\mathscr{L}=\{B, A, G\}$.


Figure 5.5: Hasse diagram of $\sqsubseteq_{L}$ for $L=G$ (left) and $L=A$ (right).

The number of trained panellists expressing each label should be decreasing on the Hasse diagram of $\sqsubseteq_{L_{j}}$ for the 'true' label $L_{j}$ of each sample in the set of samples. In such a case, the matrix of labels expressed by the trained panellists is called marginally monotone w.r.t. the list of labels $\mathbf{z}$.

Remark Note that marginal monotonicity w.r.t. a list of labels $\mathbf{z}$ is equivalent to the fact that, for each sample $a_{j}$ in the set of samples, $L_{j}$ is a mode among all the labels expressed for sample $a_{j}$ and, at the same time, the number of times that a
label is expressed for sample $a_{j}$ decreases when we move away from $L_{j}$.

The set of all matrices of labels that are marginally monotone w.r.t. at least one list of labels leads to the following natural consensus state:

$$
\mathcal{S}_{4}=\left\{\begin{array}{l|l}
\mathbf{Z} \in \mathscr{L}^{n_{T} \times n} & \begin{array}{l}
\left(\exists \mathbf{z} \in \mathscr{L}^{n}\right)\left(\forall L^{1}, L^{2} \in \mathscr{L}\right)(\forall j \in\{1, \ldots, n\}) \\
\left(L^{2} \sqsubseteq_{L_{j}} L^{1} \Rightarrow \sum_{i=1}^{n_{T}} \mathbb{1}\left(\mathbf{Z}_{i j}=L^{1}\right) \leq \sum_{i=1}^{n_{T}} \mathbb{1}\left(\mathbf{Z}_{i j}=L^{2}\right)\right)
\end{array} \tag{5.9}
\end{array}\right\},
$$

where, for any $L^{\prime}, L^{\prime \prime} \in \mathscr{L}, \mathbb{1}\left(L^{\prime}=L^{\prime \prime}\right)$ equals one if $L^{\prime}=L^{\prime \prime}$ and zero otherwise.

The underlying consensus label function $f: \mathcal{S}_{4} \rightarrow \mathcal{P}\left(\mathscr{L}^{n}\right)$ assigns to each matrix $\mathbf{Z}$ of labels in $\mathcal{S}_{4}$ the set of all lists of labels w.r.t. which it is marginally monotone.

Obviously, monotonicity implies marginal monotonicity. In particular, it holds that $\mathcal{S}_{0} \subseteq \mathcal{S}_{3} \subseteq \mathcal{S}_{4}$. In case the set of samples is a singleton, monotonicity and marginal monotonicity coincide, i.e., it holds that $\mathcal{S}_{3}=\mathcal{S}_{4}$ if $k=1$.

Example 5.3. Consider the labellings of the samples given in Example 5.2. For sample $a_{1}$, the vector of frequencies $(2,3,4,7,4)$ (displayed in the left side of the diagram of Figure 5.6) is decreasing when we move away from the mode $L_{4}$. For sample $a_{2}$, the vector of frequencies $(0,1,5,11,3)$ (displayed in the right side of the diagram of Figure 5.6) is also decreasing when we move away from the mode $L_{4}$. Therefore, the given matrix of labels is marginally monotone w.r.t. $\left(L_{4}, L_{4}\right)$. This result was already expected due to the fact that the given matrix of labels is monotone w.r.t. $\left(L_{4}, L_{4}\right)$ in view of Example 5.2 and the fact that monotonicity implies marginal monotonicity.


Figure 5.6: Example of a marginally monotone matrix of labels w.r.t. $\left(L_{4}, L_{4}\right)$.

### 5.4.6. A general comparison

As previously discussed, unanimity is the smallest consensus state, and, therefore, it implies the consensus states of majority, marginal majority, monotonicity and marginal monotonicity. Moreover, the consensus states of majority and monotonicity respectively imply the consensus states of marginal majority and marginal monotonicity. In the following, we prove that there is no further relation between these five consensus states.

Consider the singleton $\mathscr{A}=\left\{a_{1}\right\}$, the ordinal scale $\mathscr{L}=\left\{L_{1}, L_{2}, L_{3}, L_{4}\right\}$ and $n_{T}=5$ trained panellists, where the transposed matrix of labellings expressed by the trained panellists is the following:

$$
\mathbf{Z}^{\top}=\left(\begin{array}{llll}
L_{1} & L_{1} & L_{1} & L_{3} \\
L_{4}
\end{array}\right)
$$

The matrix of labels expressed by the trained panellists belongs to the (marginal) majority consensus state, but not to the (marginal) monotonicity consensus state.

Consider now the following transposed matrix of labellings expressed by the trained panellists:

$$
\mathbf{Z}^{\top}=\left(\begin{array}{llll}
L_{1} & L_{1} & L_{2} & L_{3}
\end{array} L_{4}\right)
$$

The matrix of labels expressed by the trained panellists belongs to the (marginal) monotonicity consensus state, but not to the (marginal) majority consensus state.

The relations between the different consensus states are illustrated in Figure 5.7 . In this figure, an arrow indicates that the consensus state from which the arrow starts implies the consensus state to which the arrow points.


Figure 5.7: Relations between the different consensus states.

### 5.5. The optimization problem

The search for a closest matrix of labels in the unanimity or (marginal) majority consensus state is a well-known and easy task. Indeed, it is a separable problem that can be solved for each sample independently.

In this section, we address the search for a closest monotone matrix of labels ${ }^{4}$ by solving a transportation problem [123. In a transportation problem, there are several supply points and several demand points. At each supply point, a certain amount of product is produced, and it needs to be transported to the demand points. The required demands at each demand point need to be satisfied. Transporting a unit of product from a supply point to a demand point has an associated cost. The goal of the transportation problem is to find an optimal transportation distribution that minimizes the total transportation cost from sources to destinations.

Here, the supply points correspond to the different lists of labels appearing in the labellings expressed by the trained panellists. The quantity of product produced at each supply point is given by the number of times the list of labels appears in the labellings expressed by the trained panellists. The demand points correspond to all possible lists of labels that could be a result of a labelling of the samples. There are no demands at each demand point.

We define $p \times l^{n}$ variables $a_{u v}\left(u \in\{1, \ldots, p\}\right.$ and $\left.v \in\left\{1, \ldots, l^{n}\right\}\right)$ taking values in the set $\mathbb{Z}^{+}$of non-negative integers, $n$ being the number of samples in $\mathscr{A}, p$ being the number of different lists of labels in the labellings expressed by the trained panellists and $l$ being the number of labels in the ordinal scale (it obviously holds that $p \leq l^{n}$ ). $a_{u v}$ is the number of units shipped from the $u$-th source to the $v$-th destination. For any $u \in\{1, \ldots, p\}$ and any $v \in\left\{1, \ldots, l^{n}\right\}, a_{u v}=m$ means that $m$ units of the $u$-th list of labels are assigned to $m$ units of the $v$-th list of labels ${ }^{5}$. We have an initial matrix of labels where the $u$-th list of labels appears $s_{u}$ times. These values $s_{u}$ can be seen as the number of units of product that are produced at each supply point. The goal is to distribute these products satisfying the required monotonicity constraint. Unfortunately, the set of constraints corresponding to monotonicity is formed by the union of the monotonicity constraints associated with each list of labels. In order to obtain the solution of this optimization problem, $l^{n}$ parallel transportation problems need to be solved considering the monotonicity constraints associated with each possible list of labels. Finally, we consider the lowest cost resulting from the transportation problem.

[^16]For each possible list $\mathbf{z}$ of $n$ labels, the problem to be resolved is the following:

$$
\begin{array}{cl}
\text { Minimize } & \sum_{u=1}^{p} \sum_{v=1}^{l^{n}} \mathbf{C}_{u v} a_{u v} \text { w.r.t. }\left\{a_{u v}\right\} \\
\text { s.t. } & \sum_{v=1}^{l^{n}} a_{u v}=s_{u}, \text { for any } u \in\{1, \ldots, p\}, \\
& a_{u v} \in \mathbb{Z}^{+}, \text {for any } u \in\{1, \ldots, p\} \text { and any } v \in\left\{1, \ldots, l^{n}\right\}, \\
& \sum_{u=1}^{p} a_{u v_{2}} \leq \sum_{u=1}^{p} a_{u v_{1}}, \text { for any } v_{1}, v_{2} \in\left\{1, \ldots, l^{n}\right\} \text { s.t. } \mathbf{z}_{v_{2}} \sqsubseteq_{\mathbf{z}} \mathbf{z}_{v_{1}},
\end{array}
$$

where $\mathbf{C}_{u v}$ denotes the cost of changing the $u$-th list of labels in the matrix of labels into the $v$-th list of labels in the set of all possible lists of labels, and $\mathbf{z}_{v}$ represents the $v$-th list of labels in the set of all possible lists of labels.

Each optimization problem leads to the computation of a closest monotone matrix of labels w.r.t. the corresponding list of labels z. Transportation problems are solved in polynomial time [124]. However, here, this polynomial time is in terms of $p \times l^{n}$. Obviously, this is a computational drawback for sets of samples of large cardinality. Assignment problems have been deeply studied in combinatorial optimization, resulting in several methods to solve this problem in polynomial time, for instance, the Hungarian method [125] or the Auction algorithm [126].

### 5.6. Independence of the search for different consensus states

In this section, we prove the independence of the search for the five mentioned consensus states. First, we show that the search for unanimity, (marginal) majority and (marginal) monotonicity are independent w.r.t. each other. Second, we show that the search for majority and the search for marginal majority are independent w.r.t. each other. Third, we show that the search for monotonicity and the search for marginal monotonicity are independent w.r.t. each other.

### 5.6.1. Unanimity, (marginal) majority and (marginal) monotonicity

Consider the singleton $\mathscr{A}=\left\{a_{1}\right\}$, the ordinal scale $\mathscr{L}=\left\{L_{1}, \ldots, L_{7}\right\}$ and $n_{T}=10$ trained panellists, where the transposed matrix of labellings expressed by the trained panellists is the following:

$$
\mathbf{Z}^{\top}=\left(\begin{array}{lllllllll}
L_{1} & L_{2} & L_{2} & L_{3} & L_{5} & L_{5} & L_{5} & L_{5} & L_{7}
\end{array} L_{7}\right) .
$$

The frequencies corresponding to each of the labels are then represented by the following vector:

$$
(1,2,1,0,4,0,2) .
$$

First, the consensus labelling associated with the unanimity consensus state is ( $L_{4}$ ) due to the fact that any other list of labels will force to change at least one label into another label that is at absolute distance four.

Second, the consensus labelling associated with the (marginal) ${ }^{6}$ majority consensus state is $\left(L_{6}\right)$ due to the fact that it is the only list of labels that can reach a frequency of six just by changing labels into an adjacent labe ${ }^{7}$.

Third, we see that the consensus labelling associated with the (marginal) ${ }^{8}$ monotonicity consensus state is $\left(L_{5}\right) \sqrt{9}$.

We conclude that the consensus labellings of $\mathscr{A}$ according to the search for unanimity, (marginal) majority and (marginal) monotonicity are $\left(L_{4}\right),\left(L_{6}\right)$ and $\left(L_{5}\right)$, respectively. Therefore, the search for unanimity, for (marginal) majority and for (marginal) monotonicity are proved to be independent w.r.t. each other.

### 5.6.2. Marginal majority and majority

Consider the set $\mathscr{A}=\left\{a_{1}, a_{2}\right\}$, the ordinal scale $\mathscr{L}=\left\{L_{1}, L_{2}, L_{3}\right\}$ and $n_{T}=11$ trained panellists, where the transposed matrix of labellings expressed by the trained panellists is the following:

$$
\mathbf{Z}^{\top}=\left(\begin{array}{llllllllll}
L_{1} & L_{1} & L_{1} & L_{2} & L_{2} & L_{2} & L_{2} & L_{2} & L_{2} & L_{3} \\
L_{3} \\
L_{1} & L_{1} & L_{1} & L_{1} & L_{2} & L_{2} & L_{2} & L_{2} & L_{2} & L_{1}
\end{array} L_{1}\right)
$$

We note that, for sample $a_{1}$, the trained panellists have expressed label $L_{2}$ six times, and, for sample $a_{2}$, the trained panellists have expressed label $L_{1}$ six times. We conclude that the matrix of labels expressed by the trained panellists belongs to the marginal majority consensus state, resulting in $\left(L_{2}, L_{1}\right)$. However, one may note that $\left(L_{2}, L_{2}\right)$ appears five times as a row of the matrix of labellings expressed by the trained panellists, while $\left(L_{2}, L_{1}\right)$ appears only once, resulting in $\left(L_{2}, L_{2}\right)$ in case we consider the majority consensus state.

We conclude that the consensus labelling of $\mathscr{A}$ according to the search for marginal

[^17]majority is $\left(L_{2}, L_{1}\right)$, while the consensus labelling of $\mathscr{A}$ according to the search for majority is $\left(L_{2}, L_{2}\right)$. Therefore, the search for marginal majority and majority are proved to be independent w.r.t. each other.

### 5.6.3. Marginal monotonicity and monotonicity

Consider the set $\mathscr{A}=\left\{a_{1}, a_{2}\right\}$, the ordinal scale $\mathscr{L}=\left\{L_{1}, L_{2}, L_{3}\right\}$ and $n_{T}=10$ trained panellists, where the transposed matrix of labellings expressed by the trained panellists is the following:

$$
\mathbf{Z}^{\top}=\left(\begin{array}{lllllllll}
L_{1} & L_{1} & L_{1} & L_{2} & L_{2} & L_{2} & L_{3} & L_{3} & L_{3} \\
L_{3} \\
L_{1} & L_{1} & L_{1} & L_{3} & L_{3} & L_{3} & L_{2} & L_{2} & L_{2}
\end{array} L_{3}\right)
$$

We note that, for both samples, the trained panellists have expressed label $L_{1}$ three times, label $L_{2}$ three times and label $L_{3}$ four times. Therefore, for both samples $a_{1}$ and $a_{2}$, the frequencies corresponding to each of the labels are represented by the following vector:

$$
(3,3,4) .
$$

The frequencies are obviously decreasing when we move away from $L_{3}$. We conclude that the matrix of labels expressed by the trained panellists is marginally monotone w.r.t. the list of labels $\left(L_{3}, L_{3}\right)$.

On the other hand, the consensus labelling associated with the monotonicity consensus state is ( $L_{2}, L_{2}$ ), leading to the following closest monotone transposed matrix of labels:

$$
\mathbf{Z}^{\top}=\left(\begin{array}{lllllllll}
L_{1} & L_{1} & L_{2} & L_{2} & L_{2} & L_{2} & L_{2} & L_{3} & L_{3} \\
L_{3} \\
L_{1} & L_{2} & L_{1} & L_{2} & L_{3} & L_{3} & L_{2} & L_{2} & L_{2}
\end{array} L_{3}\right)
$$

As shown in Fig. 5.8, the cost associated with a closest monotone matrix of labels w.r.t. the list of labels $\left(L_{3}, L_{3}\right)$, which equals six, is greater than the cost associated with a closest monotone matrix of labels w.r.t. the list of labels $\left(L_{2}, L_{2}\right)$, which equals four.

We conclude that the consensus labelling of $\mathscr{A}$ according to the search for marginal monotonicity is ( $L_{3}, L_{3}$ ), while the consensus labelling of $\mathscr{A}$ according to the search for monotonicity is ( $L_{2}, L_{2}$ ). Therefore, the search for marginal monotonicity and monotonicity are proved to be independent w.r.t. each other.


Figure 5.8: Hasse diagram of $\sqsubseteq_{\mathbf{z}}$ for $\mathbf{z}=\left(L_{3}, L_{3}\right)$ (left) and $\mathbf{z}=\left(L_{2}, L_{2}\right)$ (right).

### 5.7. Case studies

This section is devoted to illustrate the method introduced in this chapter, using two real-life examples that describe different types of problems. The goal is to search for the consensus labelling of a given set of profiles of candidates and a set of movies.

### 5.7.1. Dating system

In this subsection, we consider an entirely different real-life problem, where we search for the consensus labelling of several dating profiles or candidates. Charles University maintains a dataset, named "the collaborative filtering dataset - dating agency" [127, which contains 17, 359, 346 labellings of 168,791 candidates by 135,359 users, where each user has evaluated at least 20 candidates, and users who provided constant ratings were excluded. We consider a part of the dataset that contains labellings by 526 users of three of the most evaluated candidates that we term $a_{1}, a_{2}$ and $a_{3}$.

The considered qualitative scale $\mathscr{L}$ consists of ten labels, where the labels represent a certain labelling of candidate by a user, as follows: $L_{1}$ "Very poor"; $L_{2}$ "Poor"; $L_{3}$ "Significantly below average"; $L_{4}$ "Below average"; $L_{5}$ "Average", $L_{6}$ "Above average"; $L_{7}$ "Significantly above average"; $L_{8}$ "Good"; $L_{9}$ "Very good"; $L_{10}$ "Best". In Table 5.1 we show the 1, 578 labels expressed for the candidates by the 526 users.

By considering the monometric defined by Eq. (5.4), we address the search for a closest monotone matrix of labels w.r.t. all possible lists of labels. After solving the optimization problem formalized in Section 5.5. we conclude that the list of

| Freq. | $a_{1}$ | $a_{2}$ | $a_{3}$ | Freq. | $a_{1}$ | $a_{2}$ | $a_{3}$ | Freq. | $a_{1}$ | $a_{2}$ | $a_{3}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 6 | 10 | 6 | 1 | 8 | 7 | 7 | 42 | 9 | 10 | 5 |
| 1 | 6 | 10 | 7 | 2 | 8 | 8 | 5 | 101 | 9 | 10 | 6 |
| 1 | 7 | 1 | 6 | 1 | 8 | 8 | 6 | 26 | 9 | 10 | 7 |
| 4 | 7 | 10 | 5 | 1 | 8 | 9 | 7 | 8 | 9 | 10 | 8 |
| 8 | 7 | 10 | 6 | 60 | 8 | 10 | 5 | 2 | 9 | 10 | 9 |
| 1 | 7 | 10 | 7 | 114 | 8 | 10 | 6 | 10 | 9 | 10 | 10 |
| 2 | 8 | 1 | 5 | 56 | 8 | 10 | 7 | 1 | 10 | 2 | 5 |
| 2 | 8 | 1 | 6 | 9 | 8 | 10 | 8 | 11 | 10 | 10 | 5 |
| 1 | 8 | 2 | 6 | 7 | 8 | 10 | 10 | 22 | 10 | 10 | 6 |
| 1 | 8 | 3 | 5 | 1 | 9 | 1 | 5 | 11 | 10 | 10 | 7 |
| 1 | 8 | 4 | 7 | 1 | 9 | 1 | 6 | 2 | 10 | 10 | 8 |
| 1 | 8 | 5 | 5 | 1 | 9 | 2 | 6 | 1 | 10 | 10 | 9 |
| 1 | 8 | 5 | 8 | 1 | 9 | 5 | 5 | 7 | 10 | 10 | 10 |
| 1 | 8 | 7 | 5 | 1 | 9 | 8 | 6 |  |  |  |  |

Table 5.1: Expressed lists of labels and their frequency for the three candidates.
labels for which the associated closest monotone matrix of labels leads to the lowest cost is $\left(L_{8}, L_{10}, L_{6}\right)$ (with a cost of $12,541,944$, where 1,525 expressed labels remained unchanged, 48 expressed labels were changed into an adjacent label in the scale $\mathscr{L}$ and five expressed labels were changed into a label at absolute distance two). We note that the consensus labelling of the set of candidates according to the search for unanimity, (marginal) majority and marginal monotonicity is also ( $L_{8}, L_{10}, L_{6}$ ).

### 5.7.2. Movie preferences

We now consider another real-life problem, where we search for the consensus labelling of several movies. The University of Minnesota maintains a dataset, named MovieLens [128], which contains over 100, 000 labellings of 1,682 movies by 943 users, where each user has evaluated at least 20 movies. We consider a part of the dataset that contains labellings by 201 users of the four most evaluated movies: $a_{1}$, Star Wars (1977); $a_{2}$, Contact (1997); $a_{3}$, Fargo (1996); and $a_{4}$, Return of the Jedi (1983).

The considered qualitative scale $\mathscr{L}$ consists of five labels, where $L_{1}, L_{2}, L_{3}, L_{4}$, and $L_{5}$ represent that the user evaluated the movie as "Awful", "Bad", "Not bad", "Good", and "Excellent", respectively. In Table 5.2, we show the 804 labels
expressed for the movies by the 201 users.

| Freq. | $a_{1}$ | $a_{2}$ | $a_{3}$ | $a_{4}$ | Freq. | $a_{1}$ | $a_{2}$ | $a_{3}$ | $a_{4}$ | Freq. | $a_{1}$ | $a_{2}$ | $a_{3}$ | $a_{4}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 1 | 2 | 4 | 1 | 2 | 4 | 3 | 4 | 4 | 3 | 5 | 3 | 4 | 3 |
| 1 | 2 | 4 | 3 | 3 | 2 | 4 | 3 | 5 | 3 | 5 | 5 | 3 | 4 | 4 |
| 1 | 2 | 4 | 4 | 2 | 4 | 4 | 3 | 5 | 4 | 2 | 5 | 3 | 4 | 5 |
| 1 | 2 | 4 | 5 | 2 | 1 | 4 | 4 | 2 | 4 | 1 | 5 | 3 | 5 | 3 |
| 1 | 2 | 5 | 5 | 2 | 3 | 4 | 4 | 3 | 3 | 4 | 5 | 3 | 5 | 4 |
| 1 | 3 | 1 | 4 | 4 | 1 | 4 | 4 | 3 | 4 | 5 | 5 | 3 | 5 | 5 |
| 2 | 3 | 2 | 4 | 3 | 1 | 4 | 4 | 4 | 2 | 1 | 5 | 4 | 1 | 3 |
| 1 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 3 | 4 | 5 | 4 | 1 | 4 |
| 1 | 3 | 3 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 2 | 5 | 4 | 1 | 5 |
| 1 | 3 | 4 | 3 | 3 | 1 | 4 | 4 | 5 | 3 | 2 | 5 | 4 | 2 | 4 |
| 1 | 3 | 4 | 4 | 2 | 7 | 4 | 4 | 5 | 4 | 1 | 5 | 4 | 3 | 3 |
| 2 | 3 | 4 | 4 | 3 | 3 | 4 | 5 | 3 | 3 | 3 | 5 | 4 | 3 | 4 |
| 1 | 3 | 4 | 4 | 4 | 1 | 4 | 5 | 3 | 5 | 8 | 5 | 4 | 3 | 5 |
| 1 | 3 | 4 | 5 | 4 | 2 | 4 | 5 | 4 | 4 | 3 | 5 | 4 | 4 | 4 |
| 1 | 3 | 5 | 4 | 3 | 1 | 4 | 5 | 5 | 2 | 9 | 5 | 4 | 4 | 5 |
| 2 | 3 | 5 | 5 | 3 | 2 | 4 | 5 | 5 | 3 | 3 | 5 | 4 | 5 | 3 |
| 1 | 4 | 1 | 1 | 3 | 4 | 4 | 5 | 5 | 4 | 13 | 5 | 4 | 5 | 4 |
| 1 | 4 | 1 | 5 | 4 | 1 | 5 | 1 | 2 | 5 | 12 | 5 | 4 | 5 | 5 |
| 1 | 4 | 2 | 2 | 3 | 2 | 5 | 2 | 3 | 4 | 1 | 5 | 5 | 3 | 4 |
| 1 | 4 | 2 | 3 | 2 | 1 | 5 | 2 | 4 | 4 | 4 | 5 | 5 | 3 | 5 |
| 1 | 4 | 2 | 3 | 3 | 1 | 5 | 2 | 4 | 5 | 4 | 5 | 5 | 4 | 4 |
| 1 | 4 | 2 | 4 | 2 | 1 | 5 | 2 | 5 | 4 | 4 | 5 | 5 | 4 | 5 |
| 2 | 4 | 2 | 4 | 3 | 3 | 5 | 2 | 5 | 5 | 1 | 5 | 5 | 5 | 3 |
| 1 | 4 | 2 | 5 | 3 | 1 | 5 | 3 | 1 | 5 | 6 | 5 | 5 | 5 | 4 |
| 1 | 4 | 3 | 1 | 4 | 1 | 5 | 3 | 2 | 3 | 13 | 5 | 5 | 5 | 5 |
| 1 | 4 | 3 | 2 | 3 | 1 | 5 | 3 | 3 | 5 |  |  |  |  |  |
| 1 | 4 | 3 | 2 | 4 | 1 | 5 | 3 | 4 | 2 |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Table 5.2: Expressed lists of labels and their frequency for the four movies.

By considering the monometric defined by Eq. (5.4), we address the search for a closest monotone matrix of labels w.r.t. all possible lists of labels. After solving the optimization problem formalized in Section 5.5, we conclude that the list of labels for which the associated closest monotone matrix of labels leads to the lowest cost is $\left(L_{5}, L_{4}, L_{5}, L_{4}\right)$ (with a cost of 63,516 , where 724 expressed labels remained unchanged and 80 expressed labels were changed into an adjacent label in the scale $\mathscr{L}$ ). We note that the consensus labelling of the set of movies according to the search for unanimity, (marginal) majority and marginal monotonicity is also

$$
\left(L_{5}, L_{4}, L_{5}, L_{4}\right)
$$

### 5.8. Application to sensory data

In this section, we apply the methods discussed in this chapter to the datasets gathered from labelling tests in Chapter 4 that measure the degree of freshness of Atlantic cod, chicken breasts, brown shrimp and Atlantic salmon samples.

### 5.8.1. Atlantic cod

In this subsection, we consider the results of experiments L4, L8 and H4 gathered in Table A. 4 where a number of panellists (eight or ten, depending on the experiment) assigned labels to the cod samples from company A described in Table 4.5, and we apply the aforementioned methods to determine the consensus labels of these samples. To illustrate the method introduced in this chapter, we first consider the results of storage experiment H4 in Table A.4.

To illustrate the method introduced in this chapter using the sensory data, we begin by considering the results of experiment H 4 (A) in Table A. 4 . We first summarize the results in Table 5.3 and show the 40 labels expressed for the cod samples $\mathscr{A}=\left\{a^{0}, a^{4}, a^{8}, a^{13}\right\}$ by the 10 trained panellists. In the following subsections, the considered qualitative scale $\mathscr{L}$ consists of five labels, where $L_{1}, L_{2}, L_{3}, L_{4}$, and $L_{5}$ represent that the degree of freshness of a sample is "Spoiled", "Marginal", "Satisfactory", "Fresh" and "Very Fresh", respectively.

| Trained panellist | $a^{0}$ | $a^{4}$ | $a^{8}$ | $a^{13}$ |
| :---: | :--- | :--- | :--- | :--- |
| 1 | $L_{5}$ | $L_{4}$ | $L_{3}$ | $L_{1}$ |
| 2 | $L_{5}$ | $L_{4}$ | $L_{2}$ | $L_{1}$ |
| 3 | $L_{5}$ | $L_{3}$ | $L_{3}$ | $L_{2}$ |
| 4 | $L_{5}$ | $L_{3}$ | $L_{2}$ | $L_{2}$ |
| 5 | $L_{5}$ | $L_{3}$ | $L_{1}$ | $L_{2}$ |
| 6 | $L_{5}$ | $L_{3}$ | $L_{1}$ | $L_{1}$ |
| 7 | $L_{5}$ | $L_{3}$ | $L_{1}$ | $L_{1}$ |
| 8 | $L_{5}$ | $L_{2}$ | $L_{4}$ | $L_{3}$ |
| 9 | $L_{4}$ | $L_{1}$ | $L_{3}$ | $L_{1}$ |
| 10 | $L_{3}$ | $L_{4}$ | $L_{2}$ | $L_{1}$ |

Table 5.3: Expressed list of labels for the four samples.

We see that, for sample $a^{0}$, eight trained panellists have expressed label $L_{5}$. Therefore, the label that minimizes the sum of the zero-one distances for sample $a^{0}$ is $L_{5}$ (it leads to a cost of two, while $L_{3}$ and $L_{4}$ lead to a cost of nine and $L_{1}$ and $L_{2}$ lead to a cost of ten). Analogously, $L_{3}$ is the label that minimizes the sum of the zero-one distances for sample $a^{4}$ (leading to a cost of five); $L_{1}, L_{2}$ and $L_{3}$ are the labels that minimize the sum of the zero-one distances for sample $a^{8}$ (leading to a cost of seven); and $L_{1}$ is the label that minimizes the sum of the zero-one distances for sample $a^{13}$ (leading to a cost of four). Therefore, the lists of labels that minimize the sum of the zero-one distances are ( $L_{5}, L_{3}, L_{1}, L_{1}$ ), $\left(L_{5}, L_{3}, L_{2}, L_{1}\right)$ and ( $\left.L_{5}, L_{3}, L_{3}, L_{1}\right)$.

In case we consider the absolute distance function, we see that the label that minimizes the sum of the absolute distances for sample $a^{0}$ is $L_{5}$ (it leads to a cost of three, while $L_{4}$ leads to a cost of nine, $L_{3}$ leads to a cost of $17, L_{2}$ leads to a cost of 27 and $L_{1}$ leads to a cost of 37). Analogously, $L_{3}$ is the label that minimizes the sum of the absolute distances for sample $a^{4}$ (leading to a cost of six); $L_{2}$ is the label that minimizes the sum of the absolute distances for sample $a^{8}$ (leading to a cost of eight); and $L_{1}$ is the label that minimizes the sum of the absolute distances for sample $a^{13}$ (leading to a cost of five). Therefore, the unique list of labels that minimizes the sum of the absolute distances is ( $L_{5}, L_{3}, L_{2}, L_{1}$ ).

Note that sample $a^{0}$ is assigned eight times label $L_{5}$ and one time label $L_{4}$ and label $L_{3}$. Obviously, the frequencies decrease when we move away from label $L_{5}$. For the second sample, the frequencies decrease when we move away from label $L_{3}$. For the third sample, the frequencies decrease when we move away from either label $L_{1}, L_{2}$, or $L_{3}$. Finally, for the fourth sample, the frequencies decrease when we move away from label $L_{1}$. Therefore, the matrix of labels expressed by the trained panellists already is marginally monotone w.r.t. the lists of labels ( $L_{5}, L_{3}, L_{1}, L_{1}$ ), $\left(L_{5}, L_{3}, L_{2}, L_{1}\right)$ and ( $\left.L_{5}, L_{3}, L_{3}, L_{1}\right)$. Unfortunately, monotonicity does not hold w.r.t. any of these three lists of labels, and the search for a closest monotone matrix of labels needs to be addressed.

By considering the monometric defined by Eq. (5.4), we address the search for a closest monotone matrix of labels w.r.t. all possible lists of labels. After solving the optimization problem formalized in Section 5.5. we conclude that the list of labels for which the corresponding closest monotone matrices of labels lead to the lowest cost are $\left(L_{5}, L_{3}, L_{3}, L_{1}\right)$ and $\left(L_{5}, L_{3}, L_{2}, L_{1}\right)$ (with a cost of 400 , where 30 out of the 40 labels expressed for the samples remained unchanged and 10 labels were changed into an adjacent label in the scale $\mathscr{L}$ ).

In Table 5.4 a closest monotone matrix of labels w.r.t. both lists of labels ( $L_{5}$, $L_{3}, L_{3}, L_{1}$ ) (left) and ( $L_{5}, L_{3}, L_{2}, L_{1}$ ) (right) is shown. The labels that have been changed into a different label in one of the obtained closest monotone matrices of labels in comparison with the matrix of labels expressed by the trained panellists (see Table 5.3) are highlighted.

| Trained panellist | $a^{0}$ | $a^{4}$ | $a^{8}$ | $a^{13}$ | Trained panellist | $a^{0}$ | $a^{4}$ | $a^{8}$ | $a^{13}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | $L_{5}$ | $L_{4}$ | $L_{3}$ | $L_{1}$ | 1 | $L_{5}$ | $L_{4}$ | $L_{3}$ | $L_{1}$ |
| 2 | $L_{5}$ | $L_{4}$ | $L_{2}$ | $L_{1}$ | 2 | $L_{5}$ | $L_{4}$ | $L_{2}$ | $L_{1}$ |
| 3 | $L_{5}$ | $L_{3}$ | $L_{3}$ | $L_{1}$ | 3 | $L_{5}$ | $L_{3}$ | $L_{3}$ | $L_{1}$ |
| 4 | $L_{5}$ | $L_{3}$ | $L_{2}$ | $L_{2}$ | 4 | $L_{5}$ | $L_{3}$ | $L_{2}$ | $L_{2}$ |
| 5 | $L_{5}$ | $L_{3}$ | $L_{1}$ | $L_{2}$ | 5 | $L_{5}$ | $L_{3}$ | $L_{1}$ | $L_{2}$ |
| 6 | $L_{5}$ | $L_{3}$ | $L_{2}$ | $L_{1}$ | 6 | $L_{5}$ | $L_{3}$ | $L_{2}$ | $L_{1}$ |
| 7 | $L_{5}$ | $L_{3}$ | $L_{1}$ | $L_{1}$ | 7 | $L_{5}$ | $L_{3}$ | $L_{1}$ | $L_{1}$ |
| 8 | $L_{5}$ | $L_{3}$ | $L_{3}$ | $L_{2}$ | 8 | $L_{5}$ | $L_{3}$ | $L_{3}$ | $L_{2}$ |
| 9 | $L_{5}$ | $L_{2}$ | $L_{3}$ | $L_{1}$ | 9 | $L_{5}$ | $L_{2}$ | $L_{2}$ | $L_{1}$ |
| 10 | $L_{4}$ | $L_{3}$ | $L_{3}$ | $L_{1}$ | 10 | $L_{4}$ | $L_{3}$ | $L_{2}$ | $L_{1}$ |

Table 5.4: Closest monotone matrix of labels w.r.t. ( $L_{5}, L_{3}, L_{3}, L_{1}$ ) (left) and ( $L_{5}, L_{3}, L_{2}, L_{1}$ ) (right) given in the form of a list of evaluations.

The consensus labelling of $\mathscr{A}$ according to the different methods are shown in Table 5.5

| Method | Consensus labelling |
| :---: | :---: |
| Min. sum of the zero-one distances | $\begin{array}{l}\left(L_{5}, L_{3}, L_{1}, L_{1}\right) \\ \left(L_{5}, L_{3}, L_{2}, L_{1}\right) \\ \\ \\ \left(L_{5}, L_{3}, L_{3}, L_{1}\right)\end{array}$ |
|  | $\left(L_{5}, L_{3}, L_{2}, L_{1}\right)$ |
|  | $\begin{array}{l}\left(L_{5}, L_{3}, L_{1}, L_{1}\right) \\ \left(L_{5}, L_{3}, L_{2}, L_{1}\right) \\ \left(L_{5}, L_{3}, L_{3}, L_{1}\right)\end{array}$ |
| Search for majority | $\begin{array}{l}\left(L_{5}, L_{3}, L_{1}, L_{1}\right) \\ \left(L_{5}, L_{3}, L_{2}, L_{1}\right)\end{array}$ |
| Search for marginal majority | $\left(L_{5}, L_{3}, L_{1}, L_{1}\right)$ |
| $\left(L_{5}, L_{3}, L_{2}, L_{1}\right)$ |  |
| $\left(L_{5}, L_{3}, L_{3}, L_{1}\right)$ |  |
| Search for monotonicity | $\left(L_{5}, L_{3}, L_{2}, L_{1}\right)$ |
| $\left(L_{5}, L_{3}, L_{3}, L_{1}\right)$ |  |
| Search for marginal monotonicity | $\left(L_{5}, L_{3}, L_{1}, L_{1}\right)$ |
| $\left(L_{5}, L_{3}, L_{2}, L_{1}\right)$ |  |
| $\left(L_{5}, L_{3}, L_{3}, L_{1}\right)$ |  |$\}$

Table 5.5: Consensus labelling of $\mathscr{A}$ for the different methods.

Note that both the search for majority and marginal majority lead to labels $L_{5}$,
$L_{3}$ and $L_{1}$ for samples $a^{0}, a^{4}$ and $a^{13}$, respectively. Nevertheless, majority leads to labels $L_{1}$ or $L_{2}$ for sample $a^{8}$, while marginal majority leads to labels $L_{1}, L_{2}$ or $L_{3}$ for sample $a^{8}$. Similarly, both the search for monotonicity and marginal monotonicity lead to consensus labels $L_{5}, L_{3}$ and $L_{1}$ for samples $a^{0}, a^{4}$ and $a^{13}$, respectively. Nevertheless, monotonicity leads to labels $L_{2}$ or $L_{3}$ for sample $a^{8}$, while marginal monotonicity leads to labels $L_{1}, L_{2}$ or $L_{3}$ for sample $a^{8}$. Since the monotonicity consensus state is a broader consensus state than the unanimity or majority consensus states, where it considers the understanding of all the labellings of the samples provided by the trained panellists, we conclude that the consensus labels of sample $a^{8}$ are $L_{2}$ and $L_{3}$.

Furthermore, we notice that majority and monotonicity (i.e., consensus labelling of multiple samples) result in different consensus labels than their marginal counterpart (i.e., consensus labelling of independent samples). When considering some subsets of the set of samples (for instance $\mathscr{A}^{\prime}=\left\{a^{0}, a^{4}, a^{13}\right\}$ ), the search for monotonicity leads to label $L_{4}$ as a possible consensus label for sample $a^{4}$ in addition to $L_{3}$. We notice that sample $a^{8}$ may have an influence on the consensus label of $a^{4}$. In subset $\mathscr{A}^{\prime}$, the consensus labels for $a^{4} \in \mathscr{A}$ according to the search for monotonicity are greater than or equal to the consensus label for $a^{4} \in \mathscr{A}^{\prime}$. We conclude that this case study raises the question whether the (cor)relations between the different samples play a meaningful role in the problem of obtaining the consensus labelling of multiple samples.

The consensus labellings of the cod samples in experiments L4, L8 and H4 according to the different methods are now summarized in Table 5.6 .

The consensus labelling of the set of cod samples $\left\{a^{0}, a^{4}, a^{8}, a^{13}\right\}$ from company A in experiment L4 according to the minimum sum of the zero-one distances and absolute distances and according to the search for unanimity, (marginal) majority and marginal monotonicity is (VF, S, S, M). We note that the consensus labellings of the set of these cod samples according to the search for monotonicity are (VF, S, S, M) and (VF, M, S, M). Therefore, we conclude that the possible consensus label for sample $a^{4}$ belongs to the set of (two consecutive) labels $\{\mathrm{S}, \mathrm{M}\}$. The result of two consecutive consensus labels implies a small disagreement among the panellists. We can see in Table 4.5 that there is a small disagreement among the panellists assigning labels to sample $a^{4}$ in experiment L4, where sample $a^{4}$ is assigned one time label F, two times labels M and SP, and three times label S. Therefore, by using monotonicity, which considers the labels expressed by the panellists for all the samples simultaneously, we conclude that the consensus labels for sample $a^{4}$ are S and M .

We note that the methods show multiple consensus labellings of the set of cod samples $\left\{a^{0}, a^{3}, a^{5}, a^{7}\right\}$ from company A in experiment L8. However, we notice that the consensus labels for samples $a^{0}, a^{5}$ and $a^{7}$ according to all the methods are VF, M and SP, respectively, and that the possible consensus label for sample

| Method | Consensus labelling |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Company A |  |  |  |  |  |  |  |  |  |  |  |
|  | L4 |  |  |  | L8 |  |  |  | H4 |  |  |  |
|  | $a^{0}$ | $a^{4}$ | $a^{8}$ | $a^{13}$ | $a^{0}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ | $a^{0}$ | $a^{4}$ | $a^{8}$ | $a^{13}$ |
| Min. sum of the zero-one distances | VF | S | S | M | VF | F | M | SP | VF | S | S | SP |
|  |  |  |  |  | VF | SP | M | SP | VF | S | M | SP |
|  |  |  |  |  |  |  |  |  | VF | S | SP | SP |
| Min. sum of the absolute distances | VF | S | S | M | VF | S | M | SP | VF | S | M | SP |
|  |  |  |  |  | VF | M | M | SP |  |  |  |  |
| Search for unanimity | VF | S | S | M | VF | F | M | SP | VF | S | S | SP |
|  |  |  |  |  | VF | SP | M | SP | VF | S | M | SP |
|  |  |  |  |  |  |  |  |  | VF | S | SP | SP |
| Search for majority | VF | S | S | M | VF | F | M | SP | VF | S | M | SP |
|  |  |  |  |  | VF | SP | M | SP | VF | S | SP | SP |
| Search for marginal majority | VF | S | S | M | VF | F | M | SP | VF | S | S | SP |
|  |  |  |  |  | VF | SP | M | SP | VF | S | M | SP |
|  |  |  |  |  |  |  |  |  | VF | S | SP | SP |
| Search for monotonicity | VF | S | S | M | VF | S | M | SP | VF | S | S | SP |
|  | VF | M | S | M | VF | M | M | SP | VF | S | M | SP |
| Search for marginal monotonicity | VF | S | S | M | VF | SP | M | SP | VF | S | S | SP |
|  |  |  |  |  |  |  |  |  | VF | S | M | SP |
|  |  |  |  |  |  |  |  |  | VF | S | SP | SP |

Table 5.6: Consensus labelling of the set of cod samples from company A described in Table 4.5 for each storage experiment by aggregating the labels gathered in Table A. 4 using the different methods.
$a^{3}$ belongs to the set of labels $\{\mathrm{F}, \mathrm{S}, \mathrm{M}, \mathrm{SP}\}$. Particularly, the consensus labels for sample $a^{3}$ are S and M . We conclude that the result of two non-consecutive labels or more than two (consecutive or non-consecutive) labels implies a large disagreement among the panellists. We can see in Table 4.5 that there is a large disagreement among the panellists assigning labels to sample $a^{3}$ in experiment L8 is assigned one time labels VF and S, two times label M, and three times labels F and SP. Therefore, by using monotonicity, which considers the labels expressed by the panellists for all the samples simultaneously, we conclude that the consensus labels for sample $a^{3}$ are S and M .

Since unanimity and majority focus too strongly on the need of the trained panellists to agree on their labellings of the cod samples, we can rely more on monotonicity to provide us with a better understanding on the consensus labels of the multiple cod samples.

### 5.8.2. Chicken breast

We consider the results of experiment H4 gathered in Table A.1, where 33 panellists assigned labels to the chicken samples described in Table 4.2, and we apply the aforedescribed methods to determine the consensus labelling of these samples.

The considered qualitative scale $\mathscr{L}$ consists of three labels F, S and SP that represent that the panellist evaluated the sample as "Fresh", "Satisfactory" and "Spoiled", respectively, with the (strict) total order relation SP $\prec \mathrm{S} \prec \mathrm{F}$. The consensus labelling of these chicken samples according to the different methods is shown in Table 5.7

|  | Consensus labelling |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Method |  |  |  |  |  |  | $a^{0}$ |
| $a^{5}$ | $a^{7}$ | $a^{8}$ | $a^{9}$ | $a^{11}$ | $a^{13}$ | $a^{15}$ |  |  |
| Min. sum of the zero-one distances | F | F | F | F | F | S | SP | SP |
| Min. sum of the absolute distances | F | F | F | F | F | S | SP | SP |
| Search for unanimity | F | F | F | F | F | S | SP | SP |
| Search for majority | F | F | F | F | F | S | SP | SP |
| Search for marginal majority | F | F | F | F | F | S | SP | SP |
| Search for monotonicity | F | F | F | F | F | S | SP | SP |
| Search for marginal monotonicity | F | F | F | F | F | S | SP | SP |

Table 5.7: Consensus labelling of the set of chicken samples described in Table 4.2 for experiment H 4 by aggregating the labels gathered in Table A.1 using the different methods.

The consensus labelling of the set of chicken samples in experiment H 4 according to the minimum sum of the zero-one distances and absolute distances and according to the search for unanimity, (marginal) majority and (marginal) monotonicity is (F, F, F, F, F, S, SP, SP). We conclude that all the methods yield the same result.

### 5.8.3. Brown shrimp

We consider the results of experiments L4 and H4 for every session (1 and 2) gathered in Table A.7. where a number of panellists (nine or ten, depending on the experiment) assigned labels to the brown shrimp samples described in Table 4.8 , and we apply the aforementioned methods to determine the consensus labelling of these samples. The consensus labellings of these shrimp samples according to the different methods are shown in Table 5.8

We note that the methods show multiple consensus labellings of the set of shrimp samples $\left\{a^{0}, a^{3}, a^{5}, a^{10}\right\}$ in session 1 of experiment L4. However, we conclude

| Method | Consensus labels |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Session 1 |  |  |  |  |  |  |  | Session 2 |  |  |  |  |  |  |  |
|  | L4 |  |  |  | H4 |  |  |  | L4 |  |  |  | H4 |  |  |  |
|  | $a^{0}$ | $a^{3}$ | $a^{5}$ | $a^{10}$ | $a^{0}$ | $a^{3}$ | $a^{7}$ | $a^{12}$ | $a^{0}$ | $a^{3}$ | $a^{5}$ | $a^{10}$ | $a^{0}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ |
| Min. sum of the zero-one distances | VF | VF | S | M | F | F | SP | SP | F | F | F | SP | VF | F | SP | M |
|  | VF | VF | S | SP |  |  |  |  | F | S | F | SP |  |  |  |  |
|  | VF | F | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | VF | F | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | VF | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | VF | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | F | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | F | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |
| Min. sum of the absolute distances | F | F | S | M | F | F | SP | SP | F | F | F | SP | VF | F | M | M |
|  |  |  |  |  |  |  |  |  | F | S | F | SP |  |  |  |  |
| Search for unanimity | VF | VF | S | M | F | F | SP | SP | F | F | F | SP | VF | F | SP | M |
|  | VF | VF | S | SP |  |  |  |  | F | S | F | SP |  |  |  |  |
|  | VF | F | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | VF | F | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | VF | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | VF | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | F | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | F | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |
| Search for majority | VF | VF | S | SP | F | F | SP | SP | F | F | F | SP | VF | F | SP | SP |
|  |  |  |  |  |  |  |  |  | F | S | F | SP |  |  |  |  |
| Search for marginal majority | VF | VF | S | M | F | F | SP | SP | F | F | F | SP | VF | F | SP | M |
|  | VF | VF | S | SP |  |  |  |  | F | S | F | SP |  |  |  |  |
|  | VF | F | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | VF | F | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | VF | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | VF | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | F | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | F | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |
| Search for monotonicity | F | VF | S | M | F | F | SP | SP | F | F | F | SP | VF | F | M | M |
|  | F | F | S | M |  |  |  |  | F | S | F | SP | VF | F | SP | M |
| Search for marginal monotonicity | VF | VF | S | M | F | F | SP | SP | F | F | F | SP | VF | F | SP | M |
|  | VF | VF | S | SP |  |  |  |  | F | S | F | SP |  |  |  |  |
|  | VF | F | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | VF | F | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | VF | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | VF | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | F | S | M |  |  |  |  |  |  |  |  |  |  |  |  |
|  | F | F | S | SP |  |  |  |  |  |  |  |  |  |  |  |  |

Table 5.8: Consensus labelling of the set of shrimp samples described in Table 4.8 for every session (1 and 2) and for each storage experiment by aggregating the labels gathered in Table A. 7 using the different methods.
that the consensus label of sample $a^{5}$ is S , and that the possible consensus label for samples $a^{0}, a^{3}$ and $a^{10}$ belongs to the sets of labels $\{\mathrm{VF}, \mathrm{F}\},\{\mathrm{VF}, \mathrm{F}\}$ and $\{\mathrm{M}, \mathrm{SP}\}$, respectively. We can see in Table A. 7 that there is a small disagreement among the panellists assigning labels to samples $a^{0}, a^{3}$ and $a^{10}$, Therefore, by using monotonicity, which considers the labels expressed by the panellists for all the samples simultaneously, we conclude that the consensus labels for samples $a^{0}$, $a^{5}$ and $a^{10}$ are $\mathrm{F}, \mathrm{S}$ and M, respectively, and that the possible consensus label for sample $a^{3}$ belongs to the set of labels $\{\mathrm{VF}, \mathrm{F}\}$.

The consensus labelling of the set of shrimp samples $\left\{a^{0}, a^{3}, a^{7}, a^{12}\right\}$ in session 1
of experiment H 4 according to the minimum sum of the zero-one distances and absolute distances and according to the search for unanimity, (marginal) majority and (marginal) monotonicity is (F, F, SP, SP). Similarly, the consensus labellings of the set of shrimp samples $\left\{a^{0}, a^{3}, a^{5}, a^{10}\right\}$ in session 2 of experiment L4 according to the minimum sum of the zero-one distances and absolute distances and according to the search for unanimity, (marginal) majority and (marginal) monotonicity are (F, F, F, SP) and (F, S, F, SP). We conclude that all the methods yield the same result.

We note that the methods show multiple consensus labellings of the set of shrimp samples $\left\{a^{0}, a^{3}, a^{5}, a^{7}\right\}$ in session 2 of experiment H4. However, we conclude that the consensus labels for samples $a^{0}$ and $a^{3}$ are VF and F, respectively, and that the possible consensus label of samples $a^{5}$ and $a^{7}$ belongs to the set of labels $\{\mathrm{M}, \mathrm{SP}\}$. Furthermore, we can see in Table A.7 that there is a small disagreement among the panellists assigning labels to sample $a^{5}$, where sample $a^{5}$ is assigned eight times a label greater than or equal to that of sample $a^{7}$. Therefore, by using monotonicity, which considers the labels expressed by the panellists for all the samples simultaneously, we conclude that the consensus labels for sample $a^{5}$ are M and SP.

Since unanimity and majority focus too strongly on the need of the trained panellists to agree on their labellings of the shrimp samples, we can rely more on monotonicity to provide us with a better understanding on the consensus labels of the multiple shrimp samples.

### 5.8.4. Atlantic salmon

Since the labelling test was carried out for salmon samples for 18 storage experiments, we will discuss the consensus labelling of sets of salmon samples with the most interesting results.

## Session 1

First, we consider the results of experiments H4, AN4, ANH4, A4, L4 and M4 for session 1 gathered in Table A.9, where a number of panellists (between eight and 12 , depending on the experiment) assigned labels to the salmon samples described in Table 4.11 for session 1, and we apply the aforementioned methods to determine the consensus labellings of these samples. The consensus labellings of these salmon samples according to the different methods are shown in Table 5.9.

The consensus labelling of the set of salmon samples $\left\{a^{1}, a^{5}, a^{9}, a^{11}\right\}$ in session 1 of experiment AN4 according to the minimum sum of the zero-one distances and according to the search for unanimity, (marginal) majority and marginal monotonicity is (VF, M, M, SP). In case we consider the consensus labellings according to the minimum sum of absolute distances, we have that the possible

| Method | Consensus labelling |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Session 1 |  |  |  |  |  |  |  |  |  |  |  |
|  | H4 |  |  |  | AN4 |  |  |  | ANH4 |  |  |  |
|  | $a^{1}$ | $a^{5}$ | $a^{7}$ | $a^{11}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ |
| Min. sum of the zero-one distances | VF | F | M | SP | VF | M | M | SP | VF | M | SP | SP |
| Min. sum of the absolute distances | VF | F | M | SP | VF | S | S | SP | F | M | SP | SP |
|  |  |  |  |  | VF | S | M | SP |  |  |  |  |
|  |  |  |  |  | VF | M | S | SP |  |  |  |  |
|  |  |  |  |  | VF | M | M | SP |  |  |  |  |
|  |  |  |  |  | F | S | S | SP |  |  |  |  |
|  |  |  |  |  | F | S | M | SP |  |  |  |  |
|  |  |  |  |  | F | M | S | SP |  |  |  |  |
|  |  |  |  |  | F | M | M | SP |  |  |  |  |
| Search for unanimity | VF | F | M | SP | VF | M | M | SP | VF | M | SP | SP |
| Search for majority | VF | F | M | SP | VF | M | M | SP | VF | M | SP | SP |
| Search for marginal majority | VF | F | M | SP | VF | M | M | SP | VF | M | SP | SP |
| Search for monotonicity | VF | F | M | SP | F | M | M | SP | F | M | SP | SP |
| Search for marginal monotonicity | VF | F | M | SP | VF | M | M | SP | VF | M | SP | SP |

Table 5.9: Consensus labelling of the set of salmon samples described in Table 4.11 for session 1 of each storage experiment by aggregating the labels gathered in Table A. 9 using the different methods.
consensus label of samples $a^{5}$ and $a^{9}$ belongs to the set of consecutive labels $\{\mathrm{S}, \mathrm{M}\}$. We note that the consensus labelling of the set of these salmon samples according to the search for monotonicity is (F, M, M, SP). Therefore, we conclude that the possible consensus label for sample $a^{1}$ belongs to the set of consecutive labels \{VF, F \}. Furthermore, we can see in Table A.9 that there is a small disagreement among the panellists assigning labels to samples $a^{1}, a^{5}$ and $a^{9}$. Therefore, by using monotonicity, which considers the labels expressed by the panellists for all the samples simultaneously, we conclude that the consensus labels for the set of salmon samples $\left\{a^{1}, a^{5}, a^{9}, a^{11}\right\}$ in session 1 of experiment AN4 are(F, M, M, SP).

The same conclusion can be reached for sample $a^{1}$ in the set of salmon samples $\left\{a^{1}, a^{5}, a^{9}, a^{11}\right\}$ in session 1 of experiment ANH4, where there is a small disagreement among the panellists assigning labels to sample $a^{1}$. Therefore, by using monotonicity, which considers the labels expressed by the panellists for all the samples simultaneously, we conclude that the consensus labels for the set of salmon samples $\left\{a^{1}, a^{5}, a^{9}, a^{11}\right\}$ in session 1 of experiment ANH4 are(F, M, SP, SP).

We note that the methods show multiple consensus labellings of the set of salmon samples $\left\{a^{1}, a^{5}, a^{9}, a^{11}\right\}$ in session 1 of experiment L4. However, we conclude that the consensus labels of samples $a^{1}, a^{9}$ and $a^{11}$ are $\mathrm{F}, \mathrm{S}$ and M , respectively, and that the possible consensus label of sample $a^{5}$ belongs to the set of labels $\{\mathrm{VF}, \mathrm{F}, \mathrm{S}, \mathrm{M}\}$. This implies a large disagreement among the trained panellists for sample $a^{5}$. We can see in Table A. 9 for session 1 of experiment L4 that sample $a^{5}$ is assigned six

|  |  |  |  |  |  | sensu | labe | eling |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  | Sess | on 1 |  |  |  |  |  |
|  |  |  | 4 |  |  |  |  |  |  |  |  |  |
| Method | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{13}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ |
|  | VF | S | SP | M | F | VF | S | M | F | VF | SP | S |
| Min. sum of the zero-one distances |  |  |  |  |  |  |  |  | F | VF | SP | M |
| Min. sum of the zero-one distances |  |  |  |  |  |  |  |  | F | F | SP | S |
|  |  |  |  |  |  |  |  |  | F | F | SP | M |
| Min. sum of the absolute distances | VF | S | SP | M | F | S | S | M | F | F | SP | M |
|  | VF | S | SP | M | F | VF | S | M | F | VF | SP | S |
| Search for unanimity |  |  |  |  |  |  |  |  | F | VF | SP | M |
| Search or unanimity |  |  |  |  |  |  |  |  | F | F | SP | S |
|  |  |  |  |  |  |  |  |  | F | F | SP | M |
|  | VF | S | SP | M | F | VF | S | M | F | VF | SP | S |
|  |  |  |  |  |  |  |  |  | F | VF | SP | M |
| Search for majority |  |  |  |  |  |  |  |  | F | F | SP | S |
|  |  |  |  |  |  |  |  |  | F | F | SP | M |
|  | VF | S | SP | M | F | VF | S | M | F | VF | SP | S |
| Search for marginal majority |  |  |  |  |  |  |  |  | F | VF | SP | M |
| Search for marginal majority |  |  |  |  |  |  |  |  | F | F | SP | S |
|  |  |  |  |  |  |  |  |  | F | F | SP | M |
|  | VF | S | SP | M | F | F | S | M | F | VF | M | S |
|  |  |  |  |  | F | S | S | M | F | VF | M | M |
|  |  |  |  |  |  |  |  |  | F | VF | SP | S |
| Search for monotonicity |  |  |  |  |  |  |  |  | F | F | M | S |
|  |  |  |  |  |  |  |  |  | F | F | M | M |
|  |  |  |  |  |  |  |  |  | F | F | SP | S |
| Search for marginal monotonicity | VF | S | SP | M | F | F | S | M | F | VF | SP | S |
|  |  |  |  |  | F | S | S | M | F | VF | SP | M |
|  |  |  |  |  | F | M | S | M | F | F | SP | S |
|  |  |  |  |  |  |  |  |  | F | F | SP | M |

Table 5.9: (Continued) Consensus labelling of the set of salmon samples described in Table 4.11 for session 1 of each storage experiment by aggregating the labels gathered in Table A.9 using the different methods.
times a label greater than or equal to that of sample $a^{1}$ and eight times a label less than or equal to that of sample $a^{9}$. Therefore, by using monotonicity, which considers the labels expressed by the panellists for all the samples simultaneously, we conclude that the consensus labels for sample $a^{5}$ are F and S .

Similarly, the methods show multiple consensus labellings of the set of salmon samples $\left\{a^{1}, a^{5}, a^{9}, a^{11}\right\}$ in session 1 of experiment M4. However, we conclude that the consensus label of sample $a^{1}$ is F. It is clear that the number of panellists (eight) is small and that there is a large disagreement among these panellists. Therefore, by using monotonicity, which considers the labels expressed by the panellists for all the samples simultaneously, we conclude that the consensus labels for samples $a^{5}, a^{9}$ and $a^{11}(\mathrm{VF}, \mathrm{F}),(\mathrm{M}, \mathrm{SP})$ and (S, M), respectively.

## Session 2

Second, we consider the results of experiments H4, AN4, ANH4 and A4 for session 2 gathered in Table A.9, where a number of panellists (between five and twelve, depending on the experiment) assigned labels to the salmon samples described in Table 4.11 for session 2, and we apply the aforementioned methods to determine the consensus labels of these samples. The consensus labellings of these salmon samples according to the different methods are shown in Table 5.10.

| Method | Consensus labelling |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Session 2 |  |  |  |  |  |  |  |
|  | H4 |  |  |  | AN4 |  |  |  |
|  | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ | $a^{3}$ | $a^{7}$ | $a^{9}$ | $a^{11}$ |
| Min. sum of the zero-one distances | VF | F | S | M | VF | M | M | SP |
| Min. sum of the absolute distances | VF | F | F | M | VF | M | M | SP |
|  | VF | F | S | M |  |  |  |  |
| Search for unanimity | VF | F | S | M | VF | M | M | SP |
| Search for majority | VF | F | S | M | VF | M | M | SP |
| Search for marginal majority | VF | F | S | M | VF | M | M | SP |
| Search for monotonicity | VF | F | F | M | VF | M | M | SP |
|  | VF | F | S | M |  |  |  |  |
| Search for marginal monotonicity | VF | F | S | M | VF | M | M | SP |
| Method | ANH4 |  |  |  | A4 |  |  |  |
|  | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ |
| Min. sum of the zero-one distances | F | F | F | M | VF | F | M | SP |
|  |  |  |  |  | F | F | S | M |
| Min. sum of the absolute distances | F | F | S | M | VF | F | M | SP |
| Search for unanimity | F | F | F | M | VF | F | M | SP |
|  |  |  |  |  | F | F | S | M |
| Search for majority | F | F | F | M | VF | F | M | SP |
| Search for marginal majority | F | F | F | M | VF | F | M | SP |
|  |  |  |  |  | F | F | S | M |
| Search for monotonicity | F | F | F | M | VF | F | M | SP |
|  | F | F | S | M | F | F | M | SP |
| Search for marginal monotonicity | F | F | F | M | VF | F |  | SP |
|  |  |  |  |  | F | F | S | M |

Table 5.10: Consensus labelling of the set of salmon samples described in Table 4.11 for session 2 of each storage experiment by aggregating the labels gathered in Table A. 9 using the different methods.

The consensus labelling of the set of salmon samples $\left\{a^{1}, a^{3}, a^{5}, a^{7}\right\}$ in session 2 of experiment H 4 according to the minimum sum of the zero-one distances and
absolute distances and according to the search for unanimity, (marginal) majority and marginal monotonicity is (VF, F, S, M). We note that the consensus labellings of the set of these salmon samples according to the search for monotonicity are (VF, F, F, M) and (VF, F, S, M). Furthermore, we can see in Table A.9 that there is a small disagreement among the panellists assigning labels to sample $a^{5}$. Therefore, by using monotonicity, which considers the labels expressed by the panellists for all the samples simultaneously, we conclude that the consensus labels for sample $a^{5}$ are F and S .

It can be concluded that the consensus labels of salmon samples $\left\{a^{3}, a^{5}, a^{7}\right\}$ in session 2 of experiment A4 are F, M and SP, respectively. However, from Table A.9 in session 2 of experiment A4, we see a small disagreement among the panellists whether sample $a^{1}$ is more fresh than or equally fresh to sample $a^{3}$. Therefore, by using monotonicity, we can conclude that the consensus labels of sample $a^{1}$ are VF and F.

## Session 3

Third, we consider the results of experiments H4, AN4, ANH4 and A4 for session 3 gathered in Table A.9. where a number of panellists (eight or nine, depending on the experiment) assigned labels to the salmon samples described in Table 4.11 for session 3, and we apply the aforementioned methods to determine the consensus labels of these samples. The consensus labellings of these salmon samples according to the different methods are shown in Table 5.11.

We can see in Table A. 9 that there is a small disagreement among the panellists assigning labels to samples $a^{1}, a^{5}$ and $a^{11}$. Keeping in mind that this disagreement is for a large number of samples, we conclude that, by using monotonicity, the consensus label of sample $a^{9}$ is M , and that the possible consensus label of samples $a^{1}, a^{5}$ and $a^{11}$ belongs to the sets of labels $\{\mathrm{VF}, \mathrm{F}\},\{\mathrm{F}, \mathrm{S}\}$ and $\{\mathrm{M}, \mathrm{SP}\}$, respectively.

We note that the methods show multiple consensus labellings of the set of salmon samples $\left\{a^{1}, a^{5}, a^{9}, a^{11}\right\}$ in session 3 of experiment A4. We can see in Table A. 9 that there is a small disagreement among the panellists assigning labels to sample $a^{11}$. The consensus labels of sample $a^{11}$ according to the search for marginal monotonicity are M and SP, whereas the consensus label according to the search for monotonicity is SP. Since monotonicity exploits the the labels expressed by the panellists for all the salmon samples, we conclude that the consensus label of sample $a^{11}$ is SP.

## Session 4

Fourth, we consider the results of experiments H4, AN4, ANH4 and A4 for session 4 gathered in Table A.9. where a number of panellists (eight or nine, depending on the experiment) assigned labels to the salmon samples described in Table 4.11 for session 4, and we apply the aforementioned methods to determine the consensus

| Method | Consensus labelling |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Session 3 |  |  |  |  |  |  |  |
|  | H4 |  |  |  | AN4 |  |  |  |
|  | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ |
| Min. sum of the zero-one distances | $\begin{gathered} \mathrm{VF} \\ \mathrm{~F} \end{gathered}$ | $\begin{aligned} & \text { VF } \\ & \text { VF } \end{aligned}$ | $\begin{aligned} & \mathrm{M} \\ & \mathrm{M} \end{aligned}$ | $\begin{aligned} & \mathrm{M} \\ & \mathrm{M} \end{aligned}$ | VF | F | M | SP |
| Min. sum of the absolute distances | F | VF | M | M | VF | F | M | SP |
| Search for unanimity | $\begin{gathered} \text { VF } \\ \text { F } \end{gathered}$ | $\begin{aligned} & \text { VF } \\ & \text { VF } \end{aligned}$ | $\begin{aligned} & \mathrm{M} \\ & \mathrm{M} \end{aligned}$ | $\begin{aligned} & \mathrm{M} \\ & \mathrm{M} \end{aligned}$ | VF | F | M | SP |
| Search for majority | F | VF | M | M | VF | F | M | SP |
| Search for marginal majority | $\begin{gathered} \text { VF } \\ \text { F } \end{gathered}$ | $\begin{aligned} & \text { VF } \\ & \text { VF } \end{aligned}$ | $\begin{aligned} & \mathrm{M} \\ & \mathrm{M} \end{aligned}$ | $\begin{aligned} & \mathrm{M} \\ & \mathrm{M} \end{aligned}$ | VF | F | M | SP |
| Search for monotonicity | $\begin{gathered} \text { VF } \\ \text { F } \end{gathered}$ | $\begin{aligned} & \text { VF } \\ & \text { VF } \end{aligned}$ | $\begin{aligned} & \mathrm{M} \\ & \mathrm{M} \end{aligned}$ | $\begin{aligned} & \mathrm{M} \\ & \mathrm{M} \end{aligned}$ | $\begin{gathered} \text { VF } \\ \text { F } \\ \text { F } \end{gathered}$ | S F S | M M M | SP M SP |
| Search for marginal monotonicity | $\begin{gathered} \mathrm{VF} \\ \mathrm{~F} \end{gathered}$ | $\begin{aligned} & \text { VF } \\ & \text { VF } \end{aligned}$ | $\begin{aligned} & \mathrm{M} \\ & \mathrm{M} \end{aligned}$ | $\begin{aligned} & \mathrm{M} \\ & \mathrm{M} \end{aligned}$ | VF | F | M | SP |
| Method | ANH4 |  |  |  | A4 |  |  |  |
|  | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ |
| Min. sum of the zero-one distances | VF | F | S | SP | VF | F | M | SP |
| Min. sum of the absolute distances | $\begin{aligned} & \text { VF } \\ & \text { VF } \end{aligned}$ | $\begin{aligned} & F \\ & F \end{aligned}$ | $\begin{gathered} \mathrm{S} \\ \mathrm{M} \end{gathered}$ | $\begin{aligned} & \mathrm{M} \\ & \mathrm{M} \end{aligned}$ | VF | F | M | M |
| Search for unanimity | VF | F | S | SP | VF | F | M | SP |
| Search for majority | VF | F | S | SP | VF | F | M | M |
| Search for marginal majority | VF | F | S | SP | VF | F | M | SP |
| Search for monotonicity | VF | F | S | SP | VF | F | M | SP |
| Search for marginal monotonicity | VF | F | S | SP | $\begin{aligned} & \text { VF } \\ & \text { VF } \end{aligned}$ | F |  | $\begin{gathered} \mathrm{M} \\ \mathrm{SP} \end{gathered}$ |

Table 5.11: Consensus labelling of the set of salmon samples described in Table 4.11 for session 3 of each storage experiment by aggregating the labels gathered in Table A. 9 using the different methods.
labels of these samples. The consensus labellings of these salmon samples according to the different methods are shown in Table 5.12.

We note that the methods show multiple consensus labellings of the set of salmon samples $\left\{a^{3}, a^{5}, a^{7}, a^{9}\right\}$ in session 4 of experiment H 4 . We can conclude that the consensus label for samples $a^{5}$ and $a^{7}$ is F. Furthermore, we can see in Table A. 9 that there is a large disagreement among the panellists assigning labels to samples $a^{3}$ and $a^{9}$. The consensus label of samples $a^{3}$ and $a^{9}$ according to majority is M. However, by using monotonicity, which does not focus too strongly as majority on the need of the trained panellists to agree on their labellings of the salmon samples,
we conclude that the consensus label for samples $a^{3}$ and $a^{9}$ is S .
We note that the methods show multiple consensus labellings of the set of salmon samples $\left\{a^{1}, a^{3}, a^{5}, a^{7}\right\}$ in session 4 of experiment AN4. We can see from Table A. 9 in session 4 of experiment AN4, that there is a small disagreement among the panellists for samples $a^{5}$ and $a^{7}$. Therefore, by using monotonicity, we conclude that the consensus label of samples $a^{5}$ and $a^{7}$ belongs to the set of labels $\{\mathrm{S}, \mathrm{M}\}$ and $\{\mathrm{M}, \mathrm{SP}\}$, respectively.

We note that the methods show multiple consensus labellings of the set of salmon samples $\left\{a^{1}, a^{3}, a^{5}, a^{7}\right\}$ in session 4 of experiment ANH4. We conclude that the consensus labels of samples $a^{3}$ and $a^{5}$ are VF and F , respectively. We can see from Table A. 9 in session 4 of experiment ANH4, that there is a small disagreement among the panellists for samples $a^{1}$ and $a^{7}$. Therefore, by using monotonicity, we conclude that the consensus label of samples $a^{1}$ and $a^{7}$ belongs to the set of labels $\{\mathrm{VF}, \mathrm{F}\}$.

Similarly, we can see from Table A.9 in session 4 of experiment A4, that there is a large disagreement among the panellists for sample $a^{5}$. Since the aforementioned methods focus too strongly on the need of the trained panellists to agree on their labellings of the salmon samples, we conclude that, by using monotonicity, the consensus label of samples $a^{5}$ belongs to the set of labels $\{\mathrm{F}, \mathrm{S}, \mathrm{M}\}$.

## Scoring

Finally, we consider the results of experiment AN4* gathered in Table A.11, where a number of panellists (nine or ten, depending on the day) assigned scores on a 5 -point scale (i.e. scores in the set $\{1,2,3,4,5\}$ ) to the salmon samples described in Table 4.11 in the order described in Table 4.14. Here, we are dealing with vectors of scores (rather than lists of labels), thus, we consider a scoring (rather than a labelling) of the samples to be a vector of $n$ scores where the $j$-th score corresponds to the sample $a_{j}$ in the set of samples. A scoring approach can be seen as a specific case of a labelling approach, in which the labels in $\mathscr{L}$ are equidistant. We apply the aforementioned methods to determine the consensus vector of scores of these samples according to the different methods are shown in Table 5.13 .

The consensus scoring of the set of salmon samples $\left\{\mathrm{A}^{1}, \mathrm{~B}^{2}, \mathrm{C}^{3}, \mathrm{D}^{4}\right\}$ on Tuesday (Group 1) according to the minimum sum of the zero-one distances and absolute distances and according to the search for unanimity, (marginal) majority and (marginal) monotonicity is $(5,4,3,4)$. We conclude that all the methods yield the same result.

We note that the methods show multiple consensus scorings of the set of salmon samples $\left\{\mathrm{A}^{2}, \mathrm{~B}^{3}, \mathrm{C}^{4}, \mathrm{D}^{5}\right\}$ on Thursday (Group 2). However, we conclude that the consensus score of samples $\mathrm{A}^{2}, \mathrm{C}^{4}$ and $\mathrm{D}^{5}$ is 5,4 and 2 , respectively, and that the possible consensus score of sample $\mathrm{B}^{3}$ belongs to the set of scores $\{3,4\}$ and $\{2,4\}$.

We note that the methods show multiple consensus scorings of the set of salmon samples $\left\{\mathrm{A}^{3}, \mathrm{~B}^{4}, \mathrm{C}^{5}, \mathrm{D}^{6}\right\}$ on Monday (Group 3). However, we conclude that the consensus score of samples $\mathrm{A}^{2}$ and $\mathrm{D}^{5}$ is 5 and 4 , respectively, and that the possible consensus score of samples $B^{3}$ and $C^{4}$ belongs to the set of scores $\{4,5\}$ and $\{2,4\}$, respectively. We note that the consensus scoring of the set of these salmon samples according to the minimum sum of the absolute distances and the search for monotonicity is $(5,4,2,4)$.

The consensus scoring of the set of salmon samples $\left\{\mathrm{A}^{5}, \mathrm{~B}^{6}, \mathrm{C}^{7}, \mathrm{D}^{8}\right\}$ on Friday (Group 5) according to the minimum sum of the zero-one distances and absolute distances and according to the search for unanimity, (marginal) majority and (marginal) monotonicity is $(4,3,2,2)$. We conclude that all the methods yield the same result.

### 5.9. Conclusions

In this chapter, we have presented a new method for obtaining the joint consensus labelling of multiple samples by better exploiting all the information expressed by the trained panellists. The use of this method is endorsed by the fact that it depends less on the choice of a monometric than any method based on the search for unanimity. It has been concluded that this method can be used to provide important information on the distribution of the labels assigned by the trained panellists, whereas the other methods are not always as reliable, such as in extreme cases where the labels are on opposite ends of the spectrum. The validity and potential applications of the presented method have been shown using three different real-world data sets and the sensory data gathered in Chapter 4.

It should be noted that our method is developed for settings where ordered labels are provided. For the setting where each untrained panellist expresses a ranking on the set of samples, some members of our research unit have already addressed the search for a consensus ranking based on the search for monotonicity [113]. We shed light on this method in the next chapter (Chapter 6).

However, in other settings, different kinds of information can be provided 129, 130, 131. For instance, trained panellists and untrained panellists may provide different kinds of information, such as absolute and relative information, respectively. These types of information can be combined in order to exploit the information expressed by both trained and untrained panellists. In Chapter 7, we introduce a first attempt at combining vectors of scores and rankings to improve the quality of the assessment of samples.

| Method | Consensus labelling |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Session 4 |  |  |  |  |  |  |  |
|  | H4 |  |  |  | AN4 |  |  |  |
|  | $a^{3}$ | $a^{5}$ | $a^{7}$ | $a^{9}$ | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ |
| Min. sum of the zero-one distances | F | F | F | F | VF | VF | S | SP |
|  | F | F | F | S |  |  |  |  |
|  | F | F | F | M |  |  |  |  |
|  | M | F | F | F |  |  |  |  |
|  | M | F | F | S |  |  |  |  |
|  | M | F | F | M |  |  |  |  |
| Min. sum of the absolute distances | F |  | F | S | VF | VF | M | M |
|  | S | F | F | S |  |  |  |  |
| Search for unanimity | F | F | F | F | VF | VF | S | SP |
|  | F | F | F | S |  |  |  |  |
|  | F | F | F | M |  |  |  |  |
|  | M | F | F | F |  |  |  |  |
|  | M | F | F | S |  |  |  |  |
|  | M | F | F | M |  |  |  |  |
| Search for majority | M | F | F | M | VF | VF | S | SP |
| Search for marginal majority | F | F | F | F | VF | VF | S | SP |
|  | F | F | F | S |  |  |  |  |
|  | F | F | F | M |  |  |  |  |
|  | M | F | F | F |  |  |  |  |
|  | M | F | F | S |  |  |  |  |
|  | M | F | F | M |  |  |  |  |
| Search for monotonicity | S | F | F | S | VF | VF | S | M |
|  |  |  |  |  | VF | VF | S | SP |
|  |  |  |  |  | VF | VF | M | M |
| Search for marginal monotonicity | F | F | F | S | VF | VF | S | SP |
|  | F | F | F | M |  |  |  |  |
|  | S | F | F | S |  |  |  |  |
|  | S | F | F | M |  |  |  |  |
| Method | ANH4 |  |  |  | A4 |  |  |  |
|  | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ |
| Min. sum of the zero-one distances | VF | VF | F | VF | F | F | S | SP |
| Min. sum of the absolute distances | VF | VF | F | F | F | F | S | SP |
| Search for unanimity | VF | VF | F | VF | F | F | S | SP |
| Search for majority | VF | VF | F | VF | F | F | S | SP |
| Search for marginal majority | VF | VF | F | VF | F | F | S | SP |
| Search for monotonicity | VF | VF | F | VF | F | F | F | SP |
|  | VF | VF | F | F | F | F | S | SP |
|  | F | VF | F | VF | F | F | M | SP |
| Search for marginal monotonicity | VF | VF | F | VF | F | F | S | SP |

Table 5.12: Consensus labelling of the set of salmon samples described in Table 4.11 for session 4 of each storage experiment by aggregating the labels gathered in Table A. 9 using the different methods.

| Method | Group 1 Tuesday |  |  |  | Group 2 <br> Thursday |  |  |  | Group 3 <br> Monday |  |  |  | Group 4 <br> Wednesday |  |  |  | Group 5 Friday |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{A}^{1}$ | $\mathrm{B}^{2}$ | $\mathrm{C}^{3}$ | $\mathrm{D}^{4}$ | $\mathrm{A}^{2}$ | $\mathrm{B}^{3}$ | $\mathrm{C}^{4}$ | $\mathrm{D}^{5}$ | $\mathrm{A}^{3}$ | $\mathrm{B}^{4}$ | $\mathrm{C}^{5}$ | $\mathrm{D}^{6}$ | $\mathrm{A}^{4}$ | $\mathrm{B}^{5}$ | $\mathrm{C}^{6}$ | $\mathrm{D}^{7}$ | $\mathrm{A}^{5}$ | $\mathrm{B}^{6}$ | $\mathrm{C}^{7}$ | $\mathrm{D}^{8}$ |
| Min. sum of the zero-one distances | 5 | 4 | 3 | 4 | 5 | 4 | 3 | 2 | 5 | 4 | 2 | 4 | 5 | 2 | 2 | 2 | 4 | 3 | 2 | 2 |
|  |  |  |  |  | 5 | 4 | 4 | 2 | 5 | 4 | 4 | 4 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | 5 | 5 | 2 | 4 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | 5 | 5 | 4 | 4 |  |  |  |  |  |  |  |  |
| Min. sum of the absolute distances | 5 | 4 | 3 | 4 | 5 | 4 | 3 | 2 | 4 | 4 | 3 |  | 4 | 3 | 2 | 2 | 4 | 3 | 2 | 2 |
|  |  |  |  |  | 5 | 4 | 4 | 2 |  |  |  |  |  |  |  |  |  |  |  |  |
| Search for unanimity | 5 | 4 | 3 | 4 |  | 4 | 3 | 2 | 5 | 4 | 2 | 4 | 5 | 4 | 2 | 4 | 4 | 3 | 2 | 2 |
|  |  |  |  |  | $5$ | 4 | 4 | 2 | 5 | 4 | 4 | 4 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | 5 | 5 | 2 | 4 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | 5 | 5 | 4 | 4 |  |  |  |  |  |  |  |  |
| Search for majority | 5 | 4 | 3 | 4 | 5 | 4 | 3 | 2 | 5 | 4 | 2 | 4 | 5 | 2 | 2 | 2 |  | 3 | 2 | 2 |
|  |  |  |  |  | 5 | 4 | 4 | 2 | 5 | 4 | 4 | 4 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | 5 | 5 | 2 | 4 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | 5 | 5 | 4 | 4 |  |  |  |  |  |  |  |  |
| Search for marginal majority | 5 | 4 | 3 | 4 | 5 | 4 | 3 | 2 | 5 | 4 | 2 | 4 | 5 | 2 | 2 | 2 | 4 | 3 | 2 | 2 |
|  |  |  |  |  | 5 | 4 | 4 | 2 | 5 | 4 | 4 | 4 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | 5 | 5 | 2 | 4 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | 5 | 5 | 4 | 4 |  |  |  |  |  |  |  |  |
| Search for monotonicity | 5 | 4 | 3 | 4 | 5 | 4 | 3 | 2 | 4 | 4 | 3 | 4 | 5 | 2 | 2 | 2 |  | 3 | 2 | 2 |
|  |  |  |  |  | 5 | 4 | 4 | 2 |  |  |  |  |  |  |  |  |  |  |  |  |
| Search for marginal monotonicity | 5 | 4 | 3 | 4 | 5 | 4 | 3 | 2 | $\begin{array}{llll}5 & 4 & 2 & 4 \\ 5 & 4 & 4 & 4 \\ 5 & 4 & 3 & 4 \\ 5 & 5 & 2 & 4 \\ 5 & 5 & 3 & 4 \\ 5 & 5 & 4 & 4\end{array}$ |  |  |  | 5 | 2 | 2 | 2 | 4 | 3 | 2 | 2 |
|  |  |  |  |  | 5 | 4 | 4 | 2 |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Table 5.13: Consensus vectors of scores of the sets of salmon samples described in Table 4.14 for experiment AN4* and for every group (1-5) by aggregating the scores gathered in Table A.11 using the different methods.

## 6 Aggregation of rankings

## Table of Contents

```
6.1 Introduction
6.2 Determining a consensus ranking
    6.2.1 The Borda count
    6.2.2 The method of Condorcet
    6.2.3 The method of Kemeny
    6.2.4 Monotonicity of the votrix
6.3 Consensus state problem
6.4 The optimization problem
6.5 Application of ranking rules to sensory data
    6.5.1 Chicken breast
    6.5.2 Atlantic cod
    6.5.3 Brown shrimp
    6.5.4 Atlantic salmon
6.6 Conclusions
```


### 6.1. Introduction

In this chapter we borrow ideas from the field of social choice theory to aggregate relative evaluations in the form of rankings. In social choice theory, the rankings expressed by several voters over a set of candidates are aggregated to reach a decision. Therefore, we will be elaborating on concepts of social choice theory, specifically the subdiscipline of ranking rules, hereafter.

Let us consider the problem of obtaining the consensus ranking of multiple food samples. A traditional approach for solving this problem starts with a data collection step, by asking a number of untrained panellists to express their personal preference on these samples. From here on, we call the rankings gathered from the untrained panellists a profile of rankings. A ranking rule is a function that, given a profile of rankings, decides which ranking on the set of samples is the winner.

In this chapter, we answer the following question:
Question III.2: How can we reach a consensus ranking?
We discuss a recent method for obtaining the consensus ranking of candidates based on the search for a consensus state. In that context, a consensus state is a
set of voting results, where (even when unanimity does not hold) the result of the election is still clear. In our context, the rankings expressed by the panellists are said to be in (or belong to) a consensus state if determining the consensus ranking is obvious. Monometrics [27], which are closely related to distance functions, will be a key tool for measuring how close the rankings expressed by the panellists are belonging to the chosen consensus state. In this chapter, three consensus states are discussed: unanimity, presence of a Condorcet ranking and monotonicity of the votrix. We advocate for the use of monotonicity as a tool for differentiating between truth and optimality in the aggregation of rankings [113], where we are interested in minimizing the cost (loss) of changing the given profile of rankings into another one in the required consensus state. Finally, the application of the prominent methods for aggregating rankings will be illustrated on the sensory data in Chapter 4

### 6.2. Determining a consensus ranking

Let $\mathscr{A}=\left\{a_{1}, \ldots, a_{n}\right\}$ be a set of $n$ samples. We consider the setting where $n_{U} \mathbb{1}^{1}$ untrained panellists have expressed their preference on the set of samples in the form of a strict order relation or ranking $\prec_{i}$ on $\mathscr{A}$. This list of $n_{U}$ rankings is called a profile of rankings and is denoted by $\mathbf{r}=\left(\prec_{i}\right)_{i=1}^{n_{U}}$.

Three of the most well-known methods for the aggregation of rankings are that of Borda [132], Condorcet [133] and Kemeny [134]. Typically, the methods of Kemeny and Condorcet are viewed as methods minimization a distance (measured in some way) between the profile of rankings and the postulated consensus ranking. What distinguishes both methods is the notion of consensus. In what follows, we will describe all three methods and introduce the notion of monotonicity of the votrix.

### 6.2.1. The Borda count

The Borda count, proposed by Jean-Charles de Borda [132, is a method where in a ranking of $n$ samples, the most preferred sample (ranked highest) receives ( $n-1$ ) points, the second most preferred sample receives $(n-2)$ points, and so on down to the least preferred sample (lowest ranked) which receives 0 points. For more than one ranking, the samples are sorted based on the total number of points they obtained from the rankings.

Definition 6.1 (Borda count). Let $\mathscr{A}=\left\{a_{1}, \ldots, a_{n}\right\}$ be a set of $n$ samples and $\mathbf{r}=\left(\prec_{i}\right)_{i=1}^{n_{U}}$ be a profile of $n_{U}$ rankings (on $\left.\mathscr{A}\right)$. Let $a_{j} \in \mathscr{A}$ be the $j$-th most

[^18]preferred sample in a ranking, then the score assigned to $a_{j}$ in the $i$-th ranking $\prec_{i}$ is $B_{\prec_{i}}\left(a_{j}\right)=n-j$. Thus, the Borda count $B_{\mathbf{r}}\left(a_{j}\right)$ of sample $a_{j}$ in the profile of rankings $\mathbf{r}$ is:
\[

$$
\begin{equation*}
B_{\mathbf{r}}\left(a_{j}\right)=\sum_{i=1}^{n_{U}} B_{\prec_{i}}\left(a_{j}\right) . \tag{6.1}
\end{equation*}
$$

\]

Intuitively, the winning ranking can be seen as the one that is the result of ordering the samples based on their Borda count ${ }^{2}$. Thus, the winning ranking can be characterized by the distance to the set of profiles of rankings where the first-ranked sample coincides for all rankings (135.

Note that even though the profile of rankings consists of all linear order relations, the winning ranking can be a ranking with ties in which two or more samples result in the same sum of Borda counts.

Example 6.1. Let us consider a set of four samples $\mathscr{A}=\{a, b, c, d\}$ and the profile $\mathbf{r}=\left(\prec_{i}\right)_{i=1}^{15}$ of fifteen rankings provided by the panellists shown in Table 6.1.

| $\# \prec_{i}$ | Ranking on $\mathscr{X}$ |
| :---: | :---: |
| 7 | $d \prec a \prec b \prec c$ |
| 5 | $c \prec b \prec d \prec a$ |
| 3 | $c \prec d \prec a \prec b$ |

Table 6.1: Profile of rankings on $X$ given by fifteen panellists.

Considering the Borda count, the respective points for each sample are:

$$
\begin{aligned}
& B_{\mathbf{r}}(a)=5 \cdot 3+3 \cdot 2+7 \cdot 1+0 \cdot 0=28, \\
& B_{\mathbf{r}}(b)=3 \cdot 3+7 \cdot 2+5 \cdot 1+0 \cdot 0=28, \\
& B_{\mathbf{r}}(c)=7 \cdot 3+0 \cdot 2+0 \cdot 1+8 \cdot 0=21, \\
& B_{\mathbf{r}}(d)=0 \cdot 3+5 \cdot 2+3 \cdot 1+7 \cdot 0=13 .
\end{aligned}
$$

A winning ranking on $\mathscr{A}$ is defined by sorting the samples according to their respective points. Therefore, considering the Borda count, the winning ranking is:

$$
d \prec c \prec b \sim a .
$$

It may not always be possible for panellists to differentiate between two samples, and the panellists might consider two or more samples to be similar. In this case,

[^19]every panellist should be allowed to provide a weak order relation or ranking with ties $\precsim_{i}$ on $\mathscr{A}$. There exist several extensions of the Borda count to weak orders, however, we follow the one where a weak order is first linearised and then the relative positions of tied samples are the average corresponding positions in the linear order [136, 137.

### 6.2.2. The method of Condorcet

The most important contribution to social choice theory subsequent to Borda is the notion nowadays known as the Condorcet winner [133. A Condorcet winner can be seen as the sample that would be preferred over any other sample by a simple majority (i.e., more than half of the panellists provide this preference) in pairwise comparison.
Definition 6.2 (Condorcet winner). Let $\mathscr{A}=\left\{a_{1}, \ldots, a_{n}\right\}$ be a set of $n$ samples and $\mathbf{r}=\left(\prec_{i}\right)_{i=1}^{n_{U}}$ be a profile of $n_{U}$ panellists. A profile $\mathbf{r}$ has a Condorcet winner $a_{w}$ if

$$
(\forall j \in\{1, \ldots, n\})\left(j \neq w \Longrightarrow\left(\#\left\{i \in\left\{1, \ldots, n_{U}\right\} \mid\left(a_{j} \prec_{i} a_{w}\right)\right\}>\frac{n_{U}}{2}\right)\right) .
$$

Analogously, a Condorcet loser can be seen as the sample that any other sample is preferred over it by a simple majority in pairwise comparison.
Definition 6.3 (Condorcet loser). Let $\mathscr{A}=\left\{a_{1}, \ldots, a_{n}\right\}$ be a set of $n$ samples and $\mathbf{r}=\left(\prec_{i}\right)_{i=1}^{n_{U}}$ be a profile of $n_{U}$ panellists. A profile $\mathbf{r}$ has a Condorcet loser $a_{l}$ if

$$
(\forall j \in\{1, \ldots, n\})\left(j \neq l \Longrightarrow\left(\#\left\{i \in\left\{1, \ldots, n_{U}\right\} \mid\left(a_{l} \prec_{i} a_{j}\right)\right\}>\frac{n_{U}}{2}\right)\right)
$$

In this chapter, we are more interested in determining a winning ranking of samples rather than a winning/losing sample. Another term associated with Condorcet is the Condorcet ranking. A Condorcet ranking is a ranking where every sample would be preferred over another sample ranked at a worse position than it by a simple majority in pairwise comparison. It must be noted that a Condorcet winner/ranking/loser does not always exist, but is unique if it does. When a Condorcet winner (respectively loser) and a Condorcet ranking exist, then the Condorcet winner (respectively loser) and the most preferred (respectively least preferred) sample of the Condorcet ranking coincide. It can be seen that the order of the samples in between the most preferred (Condorcet winner) and least preferred (Condorcet loser) samples is determined by finding the Condorcet winner of the set of samples and iteratively finding the next Condorcet winner of each of the resulting sets minus the previous Condorcet winners, as follows $a_{w_{j}} \in \mathscr{A} \backslash\left\{a_{w_{1}}, \ldots, a_{w_{(j-1)}}\right\}$
for $j \in\{2, \ldots, n-1\}$.
A weaker definition of the Condorcet ranking has been proposed, such that instead of a simple majority (i.e., more than half of the panellists provide the same preference), one requires a weak-majority (i.e., at least half of the panellists provide the same preference). As a result a weak Condorcet ranking is defined as a ranking where every sample would defeat any other sample ranked at a worse position than it by at least half of the panellists in pairwise comparison. It must be noted that the Condorcet ranking is always a weak Condorcet ranking, but the converse is only true when the number of panellists is odd. Moreover, the existence and the uniqueness of a weak Condorcet ranking are not assured.

Remark A key observation of Condorcet, which is now known as Condorcet's paradox, is that the collective preferences can be cyclic. For instance, the collective prefers a first sample over a second sample, prefers the second sample over a third, and prefers the third sample over the first.

The fact that a Condorcet ranking might not exist has led to the introduction of many different methods where the resulting ranking is the one that is the closest to being a Condorcet ranking. Several ways of measuring such closeness have been proposed. Here, we restrict our attention to two of the most prominent methods in social choice literature: Copeland's rule, which specifically attempts to satisfy the Condorcet criterion by looking at pairwise comparisons [138, and the method of Dodgson, who suggested a voting scheme that was reprinted at length in 139 .

The Copeland rule is the Condorcet method, in which samples are ordered by the number of their pairwise comparison wins, while subtracting the number of their pairwise comparison losses [140]. A variation of the Copeland rule exists, where instead of subtracting the number of their pairwise comparison losses, half of the number of ties is added. Even though Copeland's rule is based on the simple majority rule, the number of pairwise comparison losses and ties are taken into consideration when determining the Condorcet winner. Thus, when there is no Condorcet winner, Copeland's rule often leads to ties.

For example, suppose that there are 11 samples. To calculate the score for sample $a$, we look at how it performs against each of the other 10 samples, in a pairwise comparison. Thus, if a panellist prefers sample $a$ to sample $b$, we say that $a$ wins against $b$ in the pairwise comparison, and sample $a$ gets " 1 " point for each win. Otherwise, if sample $b$ is preferred to sample $a$, then we say that $a$ loses against $b$ in the pairwise comparison, and sample $a$ gets " -1 " point for each loss. Finally, if samples $a$ and $b$ are not preferred to each other, then we say that $a$ is tied to $b$ in the pairwise comparison, and " 0 " points for each tie. So, if $a$ won six pairwise comparisons against of the 10 other samples, and lost to the other four, then its score would be $6-4=2$.

The method of Dodgson is an extension of Condorcet, where it ends up with a Condorcet winner whenever there is one in the profile of rankings. If a Condorcet winner does not exist, the method of Dodgson looks for the sample that is closest to the Condorcet winner and is considered as a Condorcet winner after a minimum number of preference changes of panellist $3^{3}$. The method of Dodgson can be extended to the search for the Condorcet ranking.

Example 6.2. Let us consider a set of four samples $\mathscr{A}=\{a, b, c, d\}$ and the profile $\mathbf{r}=\left(\prec_{i}\right)_{i=1}^{15}$ of fifteen rankings provided by the panellists shown in Table 6.1.

Every profile of rankings is represented by a matri $\sqrt{7}^{4} \mathbf{V}$. This matrix is formed by the pairwise comparisons between each couple of samples. In that way, $\mathbf{V}_{i j}$ denotes the number of panellists that prefer the $i$-th sample to the $j$-th sample (by convention, values on the diagonal are zero).

Considering the Condorcet criterion, the respective number of times every sample defeats another sample in pairwise comparison is:

$$
\mathbf{V}=\begin{gathered}
\\
a \\
b \\
c \\
d
\end{gathered}\left(\begin{array}{cccc}
a & b & c & d \\
0 & 5 & 8 & 15 \\
10 & 0 & 8 & 10 \\
7 & 7 & 0 & 7 \\
0 & 5 & 8 & 0
\end{array}\right) .
$$

Sample $b$ is preferred to all the other samples by more than half of the panellists: ten panellists prefer sample $b$ to sample a, eight panellists prefer sample $b$ to sample $c$ and ten panellists prefer sample $b$ to sample $d$. Therefore, sample $b$ is the Condorcet winner. In addition, sample $a$ is preferred to sample $c$ by eight panellists and to sample $d$ by fifteen panellists and sample $d$ is preferred to sample $c$ by eight panellists. We conclude that sample $c$ is the Condorcet loser and that the Condorcet ranking is:

$$
c \prec d \prec a \prec b .
$$

However, the existence of the Condorcet ranking is not assured for every profile of rankings.

To illustrate the method of Dodgson, we now consider the profile of rankings where fifteen other rankings are provided by the panellists, as shown in Table 6.2.

[^20]| $\# \prec_{i}$ | Ranking on $\mathscr{A}$ |
| :---: | :---: |
| 6 | $d \prec c \prec b \prec a$ |
| 5 | $d \prec a \prec c \prec b$ |
| 4 | $b \prec a \prec d \prec c$ |

Table 6.2: Profile of rankings on $\mathscr{A}$ given by fifteen panellists.

The matrix $\mathbf{V}$ corresponding to this profile of rankings is:

$$
\mathbf{V}=\begin{gathered}
\\
a \\
b \\
c \\
d
\end{gathered}\left(\begin{array}{cccc}
a & b & c & d \\
0 & 10 & 6 & 11 \\
5 & 0 & 11 & 11 \\
9 & 4 & 0 & 15 \\
4 & 4 & 0 & 0
\end{array}\right) .
$$

Here, it is seen that there is no Condorcet winner (and therefore no Condorcet ranking) because none of the samples is preferred by more than half of the panellists.

Note that if we reverse the preference of two panellists w.r.t. samples a and c, then we obtain the profile of rankings in Table 6.3.

| $\# \prec_{i}$ | Ranking on $\mathscr{X}$ |
| :---: | :---: |
| 6 | $d \prec c \prec b \prec a$ |
| 3 | $d \prec a \prec c \prec b$ |
| 2 | $d \prec c \prec a \prec b$ |
| 4 | $b \prec a \prec d \prec c$ |

Table 6.3: Profile of rankings on $\mathscr{X}$ given by thirteen panellists.
Thus, the matrix $\mathbf{V}^{\prime}$ corresponding to the new profile of rankings is:

$$
\mathbf{V}^{\prime}=\begin{gathered}
\\
a \\
b \\
c \\
d
\end{gathered}\left(\begin{array}{cccc}
a & b & c & d \\
0 & 10 & 8 & 11 \\
5 & 0 & 11 & 11 \\
7 & 4 & 0 & 15 \\
4 & 4 & 0 & 0
\end{array}\right) .
$$

Now, the Condorcet winner, $a$, and the Condorcet ranking, $d \prec c \prec b \prec a$ exist. In Table 6.4, the number of reversals required to reach the closest Condorcet ranking is shown.

As we have previously discussed, it may not always be possible for panellists to

| Ranking | Rev. | Ranking | Rev. | Ranking | Rev. | Ranking | Rev. |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $d \prec c \prec b \prec a$ | $\mathbf{2}$ | $c \prec d \prec b \prec a$ | 10 | $b \prec d \prec c \prec a$ | 10 | $a \prec d \prec c \prec b$ | 7 |
| $d \prec c \prec a \prec b$ | 5 | $c \prec d \prec a \prec b$ | 13 | $b \prec d \prec a \prec c$ | 8 | $a \prec d \prec b \prec c$ | 11 |
| $d \prec b \prec c \prec a$ | 6 | $c \prec b \prec d \prec a$ | 14 | $b \prec c \prec d \prec a$ | 14 | $a \prec c \prec d \prec b$ | 15 |
| $d \prec b \prec a \prec c$ | 4 | $c \prec b \prec a \prec d$ | 14 | $b \prec c \prec a \prec d$ | 16 | $a \prec c \prec b \prec d$ | 19 |
| $d \prec a \prec c \prec b$ | 3 | $c \prec a \prec d \prec b$ | 17 | $b \prec a \prec d \prec c$ | 22 | $a \prec b \prec d \prec c$ | 15 |
| $d \prec a \prec b \prec c$ | 7 | $c \prec a \prec b \prec d$ | 20 | $b \prec a \prec c \prec d$ | 22 | $a \prec b \prec c \prec d$ | 23 |

Table 6.4: The number of reversals needed to make each ranking a unanimous ranking. The minimum number of reversals is shown in bold.
differentiate between two samples, and the panellists might consider two or more samples to be similar. In this case, every panellist should be allowed to provide a weak order relation or ranking with ties $\precsim_{i}$ on $\mathscr{A}$. Similar to the method of Dodgson, one looks for the sample that is closest to the Condorcet winner and is considered as a Condorcet winner after a minimum number of preference changes of panellists based on the Kemeny (rather than Kendall) $5^{5}$ distance function between rankings. This approach can be extended to the search for the Condorcet ranking in the case of having a profile of rankings with ties.

### 6.2.3. The method of Kemeny

The method of Kemeny is explicitly geared to find a consensus ranking. In the case where each panellist has the same ranking, then that is the obvious overall ranking. In the absence of such a consensus, one looks for a ranking that could be reached from the profile of rankings after a minimum number of preference changes.

By considering the method of Kemeny [134], one assigns a score to each ranking based on the Kendall distances [141] to the given profile of rankings.

Definition 6.4 (Kendall distance). For any two rankings $\prec_{1}$ and $\prec_{2}$, the Kendall distance between these rankings, denoted by $\partial_{K}\left(\prec_{1}, \prec_{2}\right)$, is the number of couples $(a, b) \in \mathscr{A}^{2}$ for which the relative order in the rankings $\prec_{1}$ and $\prec_{2}$ differ, i.e.,

$$
\partial_{K}\left(\prec_{1}, \prec_{2}\right)=\#\left\{(a, b) \in \mathscr{A}^{2} \mid a \prec_{1} b \wedge b \prec_{2} a\right\} .
$$

The distance between rankings leads to a natural definition of a distance between a ranking and a profile of rankings. The idea is to sum the distances between a ranking and each panellist's ranking in the profile. Thus, let $\mathbf{r}=\left(\prec_{i}\right)_{i=1}^{n_{U}}$ be a profile of $n_{U}$ panellists, then

[^21]\[

$$
\begin{equation*}
d_{K}(\prec, \mathbf{r})=\sum_{i=1}^{n_{U}} \partial_{K}\left(\prec, \prec_{i}\right) . \tag{6.2}
\end{equation*}
$$

\]

Let $\mathscr{R}$ be the set of all possible rankings. Thus, the ranking $\prec \in \mathscr{R}$ is a consensus ranking if it minimizes the sum of Kendall distances $d_{K}(\prec, \mathbf{r})$ to the profile $\mathbf{r}$ over all possible rankings in $\mathscr{R}$.

Example 6.3. Let us consider a set of four samples $\mathscr{A}=\{a, b, c, d\}$ and the profile $\mathbf{r}=\left(\prec_{i}\right)_{i=1}^{15}$ of fifteen rankings provided by the panellists shown in Table 6.1.

To determine the consensus ranking of these samples, we consider the problem defined by Eq. (6.2), and, for each of the 24 rankings in $\mathscr{R}$, we compute the sum of the Kendall distances to the rankings provided by the untrained panellists. In Table 6.5, the number of reversals required to reach each possible ranking is summarized.

| Ranking | Rev. | Ranking | Rev. | Ranking | Rev. | Ranking | Rev. |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $d \prec c \prec b \prec a$ | 37 | $c \prec d \prec b \prec a$ | 36 | $b \prec d \prec c \prec a$ | 43 | $a \prec d \prec c \prec b$ | 48 |
| $d \prec c \prec a \prec b$ | 32 | $c \prec d \prec a \prec b$ | $\mathbf{3 1}$ | $b \prec d \prec a \prec c$ | 44 | $a \prec d \prec b \prec c$ | 49 |
| $d \prec b \prec c \prec a$ | 38 | $c \prec b \prec d \prec a$ | 41 | $b \prec c \prec d \prec a$ | 42 | $a \prec c \prec d \prec b$ | 47 |
| $d \prec b \prec a \prec c$ | 39 | $c \prec b \prec a \prec d$ | 56 | $b \prec c \prec a \prec d$ | 57 | $a \prec c \prec b \prec d$ | 52 |
| $d \prec a \prec c \prec b$ | 33 | $c \prec a \prec d \prec b$ | 46 | $b \prec a \prec d \prec c$ | 59 | $a \prec b \prec d \prec c$ | 54 |
| $d \prec a \prec b \prec c$ | 34 | $c \prec a \prec b \prec d$ | 51 | $b \prec a \prec c \prec d$ | 58 | $a \prec b \prec c \prec d$ | 53 |

Table 6.5: The number of reversals needed to reach each possible ranking. The minimum number of reversals is shown in bold.

It may not always be possible for panellists to differentiate between two samples, and the panellists might consider two or more samples to be similar. In this case, every panellist should be allowed to provide a weak order relation or ranking with ties $\precsim_{i}$ on $\mathscr{A}$. The first distance function for rankings with ties was proposed by Kemeny and Snell [142]. This method considers all possible rankings and assigns points to each ranking based on the Kemeny distance to the given profile of rankings.

Definition 6.5 (Kemeny distance). For any two rankings with ties $\precsim_{1}$ and $\precsim_{2}$, the Kemeny distance, denoted by $d_{K}\left(\precsim_{1}, \precsim_{2}\right)$, between two rankings is

$$
\begin{align*}
d_{K}\left(\precsim_{1}, \precsim_{2}\right)= & 2 \#\left\{(a, b) \in \mathscr{A}^{2} \mid a \prec_{1} b \wedge b \prec_{2} a\right\} \\
& +\#\left\{(a, b) \in \mathscr{A}^{2} \mid a \prec_{1} b \wedge a \sim_{2} b\right\}  \tag{6.3}\\
& +\#\left\{(a, b) \in \mathscr{A}^{2} \mid a \sim_{1} b \wedge a \prec_{2} b\right\} .
\end{align*}
$$

The distance between two rankings leads to a natural definition of a distance between a ranking and a profile of rankings with ties. The idea is to sum the
distances between a ranking and each panellist's ranking in the profile. Thus let $\widetilde{\mathbf{r}}=\left(\precsim_{i}\right)_{i=1}^{n_{U}}$ be a profile of $n_{U}$ rankings, then

$$
\begin{equation*}
d_{K}(\precsim, \widetilde{\mathbf{r}})=\sum_{i=1}^{n_{U}} d_{K}\left(\precsim, \precsim_{i}\right) . \tag{6.4}
\end{equation*}
$$

Let $\widetilde{\mathscr{R}}$ be the set of all possible rankings with ties. Thus, the ranking $\precsim \in \widetilde{\mathscr{R}}$ is a consensus ranking if it minimizes the sum of Kemeny distances $d_{K}(\precsim, \widetilde{\mathbf{r}})$ over all possible rankings with ties in $\widetilde{\mathscr{R}}$. Note that when the rankings contain no ties, the Kemeny distance is equal to twice the Kendall distance [141].

In the study of sensory food quality, spoiled food samples can be considered as highly relevant samples, which have to be ranked with higher accuracy than other samples in the set of samples. In such decision problems, it is not the same to have differences in the top samples than in the bottom ones. The Kemeny distance is not sensitive to the positions at which the disagreements occur. A recent study has proposed the weighted Kemeny distance, where weights are introducted to distinguish where these differences occur [143].

Definition 6.6 (Weighted Kemeny distance). Let $\mathbf{w}=\left(w_{1}, \ldots, w_{n-1}\right) \in[0,1]^{n-1}$ be a weighting vector, such that $w_{1} \geq \ldots \geq w_{n-1}$ and $\sum_{i=1}^{n-1} w_{i}=1$. We denote by $\precsim^{\sigma}$ the ranking with ties obtained from $\precsim ~ b y ~ r a n k i n g ~ t h e ~ s a m p l e s ~ a c c o r d i n g ~ t o ~ a ~$ permutation $\sigma$, i.e., $a_{1} \precsim a_{2} \Leftrightarrow a_{1}^{\sigma} \precsim^{\sigma} a_{2}^{\sigma}$. For any two rankings with ties $\precsim_{1}$ and $\precsim_{2}$, the weighted Kemeny distance, denoted by $d_{K, \mathbf{w}}\left(\precsim_{1}, \precsim_{2}\right)$, between two rankings is defined as follows:

$$
\begin{align*}
d_{K, \mathbf{w}}\left(\precsim_{1}, \precsim_{2}\right)=\frac{1}{2}[ & \sum_{i, j=1, i<j}^{n} w_{i}\left|\operatorname{sgn}\left(p_{i}^{\sigma_{1}}-p_{j}^{\sigma_{1}}\right)-\operatorname{sgn}\left(q_{i}^{\sigma_{1}}-q_{j}^{\sigma_{1}}\right)\right|  \tag{6.5}\\
& \left.+\sum_{i, j=1, i<j}^{n} w_{i}\left|\operatorname{sgn}\left(q_{i}^{\sigma_{2}}-q_{j}^{\sigma_{2}}\right)-\operatorname{sgn}\left(p_{i}^{\sigma_{2}}-p_{j}^{\sigma_{2}}\right)\right|\right],
\end{align*}
$$

where $p_{i}$ is the position of sample $a_{i}$ in $\precsim_{1}, q_{i}$ is the position of sample $b_{i}$ in $\precsim_{2}$, $\sigma_{1}, \sigma_{2}$ are permutations such that $\precsim_{1}^{\sigma_{1}}=\precsim_{2}^{\sigma_{2}}$, and 'sgn' is the sign function:

$$
\operatorname{sgn}(x)=\left\{\begin{aligned}
1, & \text { if } x>0 \\
0, & \text { if } x=0 \\
-1, & \text { if } x<0
\end{aligned}\right.
$$

Example 6.4. Let us consider a set of four samples $\mathscr{A}=\{a, b, c, d\}$, the weighting vector $\mathbf{w}=\left(\frac{3}{6}, \frac{2}{6}, \frac{1}{6}\right)$, and the rankings provided by the panellists shown in Table 6.6. In Table 6.7, the Kemeny distance and the weighted Kemeny distance for each

| $\prec_{i}$ | Ranking on $\mathscr{X}$ |
| :---: | :---: |
| $\prec_{1}$ | $a \prec b \prec c \prec d$ |
| $\prec_{2}$ | $a \prec b \prec d \prec c$ |
| $\prec_{3}$ | $b \prec a \prec c \prec d$ |
| $\prec_{4}$ | $b \prec d \prec c \prec a$ |

Table 6.6: Rankings on $\mathscr{X}$ given by four panellists.
couple of rankings are summarized.

|  | $\left(\prec_{1}, \prec_{2}\right)$ | $\left(\prec_{1}, \prec_{3}\right)$ | $\left(\prec_{1}, \prec_{4}\right)$ | $\left(\prec_{2}, \prec_{3}\right)$ | $\left(\prec_{2}, \prec_{4}\right)$ | $\left(\prec_{3}, \prec_{4}\right)$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $d_{K}$ | 2 | 2 | 8 | 4 | 6 | 6 |
| $d_{K, \mathbf{w}}$ | $\frac{1}{3}$ | 1 | 3 | $\frac{4}{3}$ | $\frac{5}{2}$ | $\frac{5}{3}$ |

Table 6.7: Kemeny distance $d_{K}$ vs. weighted Kemeny distance $d_{K, \mathbf{w}}$.

Note that to sum the distances between a ranking and each panellist's ranking in the profile, the Kemeny distance $d_{K}\left(\precsim, \precsim_{i}\right)$ in Eq. 6.4 can be replaced with the weighted Kemeny distance $d_{K, \mathbf{w}}\left(\precsim, \precsim_{i}\right)$.

Remark Although the Borda count is undoubtedly one of the best-known ranking rules in social choice theory, it is really sensitive to manipulation [144], which was already criticized in the eighteenth century by Marquis de Condorcet. Interestingly, Condorcet's proposal has received criticism among the research community, in that a "Condorcet winner" is a "sometimes" concept (sometimes it is useful; sometimes it is not). This in itself suggests that the Condorcet winner should be critically re-examined. Young and Levenglick [145] showed that Kemeny's voting scheme is the unique voting scheme that is neutral, consistent and satisfies the Condorcet criteria. However, Bartholdi et al. [146] criticised against the method of Kemeny in that determining the winner is NP-hard.

### 6.2.4. Monotonicity of the votrix

In the field of social choice theory, the preferences of the voters are usually compressed into representations of votes gathering the most significant information. In this chapter, we restrict our attention to the most common representation, where each profile of rankings defines a matrix called the votrix [147, such that each row represents a sample in $\mathscr{A}$ and each column represents a position $j \in\{1, \ldots, n\}$. In this way, the element at the $i$-th row and $j$-th column equals the number of
times that the $i$-th sample has been preferred over the $j$-th sample in the profile of rankings given by the panellists.

Definition 6.7 (Votrix). Let $\mathscr{A}$ be a set of $n$ samples and $n_{U}$ be the number of untrained panellists. A matrix $\mathbf{V} \in\left\{0,1, \ldots, n_{U}\right\}^{n \times n}$ is called a votrix on $\mathscr{A}$ if there exists a profile $\mathbf{r}$ of $n_{U}$ rankings, such that for any two samples $a_{j_{1}}, a_{j_{2}} \in \mathscr{A}$, it holds that

$$
\mathbf{V}_{j_{1} j_{2}}=\#\left\{i \in\left\{1, \ldots, n_{U}\right\} \mid a_{j_{2}} \prec_{i} a_{j_{1}}\right\}
$$

Any ranking $\prec$ on the set $\mathscr{A}$ naturally induces a strict order relation $\sqsubset$ 148], such that for any couples of samples $\left(a_{i_{1}}, a_{j_{1}}\right),\left(a_{i_{2}}, a_{j_{2}}\right) \in \mathscr{A}^{2}$, where $a_{i_{1}} \neq a_{j_{1}}$ and $a_{i_{2}} \neq a_{j_{2}}$, it holds that $\left(a_{i_{2}}, a_{j_{2}}\right) \sqsubset\left(a_{i_{1}}, a_{j_{1}}\right)$, if

$$
\left(a_{i_{2}} \preceq a_{i_{1}}\right) \wedge\left(a_{j_{1}} \preceq a_{j_{2}}\right) \wedge\left(a_{i_{2}} \prec a_{i_{1}} \vee a_{j_{1}} \prec a_{j_{2}}\right) .
$$

This relation is graphically represented using the ranking $d \prec c \prec b \prec a$ in Figure 6.1


Figure 6.1: Hasse diagram of $\sqsubset$ for the ranking $d \prec c \prec b \prec a$.
The strict order relation $\sqsubset$ between couples of samples associated with a given ranking $\prec$ on $\mathscr{A}$ is used to define the monotonicity of the votrix w.r.t. this ranking. According to the proposal of Rademaker and De Baets [149], a votrix is called monotone w.r.t. to a ranking $c \prec b \prec a$ if the number of preferences of $a$ over $c$ is not less than both the number of preferences of $a$ over $b$ and the number of preferences of $b$ over $c$. Note that the values of a monotone votrix decrease from the top of the Hasse diagram of $\sqsubset$.

Definition 6.8 (Monotone votrix). A votrix is said to be monotone w.r.t a ranking
$\prec$ on the set $\mathscr{A}$ with corresponding order relation $\sqsubset i f$, for any $\left(a_{i_{1}}, a_{j_{1}}\right),\left(a_{i_{2}}, a_{j_{2}}\right) \in$ $\mathscr{A}^{2}$, such that $a_{i_{1}} \neq a_{j_{1}}, a_{i_{2}} \neq a_{j_{2}}$ and $\left(a_{i_{2}}, a_{j_{2}}\right) \sqsubset\left(a_{i_{1}}, a_{j_{1}}\right)$, it holds that

$$
\mathbf{V}_{i_{2} j_{2}} \leq \mathbf{V}_{i_{1} j_{1}}
$$

Example 6.5. Let us consider a set of four samples $\mathscr{A}=\{a, b, c, d\}$ and the profile $\mathbf{r}=\left(\prec_{i}\right)_{i=1}^{15}$ of fifteen rankings provided by the panellists shown in Table 6.1. This profile of rankings is not monotone w.r.t. any ranking on $\mathscr{A}$. However, we can see in Figure 6.2 that, with a small number of reversals (four), where ( $b, a$ ) is reversed to $(a, b)$ three times and $(c, d)$ is reversed to $(d, c)$ once, we can impose monotonicity w.r.t. the ranking $d \prec c \prec b \prec a$.


Figure 6.2: Votrix represented on the Hasse diagram of $\sqsubset$ for the ranking $d \prec c \prec b \prec a$.

The number of changes needed to impose monotonicity w.r.t. each possible ranking on $\mathscr{A}$ is summarized in Table 6.8.

| Ranking | Rev. | Ranking | Rev. | Ranking | Rev. | Ranking | Rev. |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $d \prec c \prec b \prec a$ | 4 | $c \prec d \prec b \prec a$ | 12 | $b \prec d \prec c \prec a$ | 15 | $a \prec d \prec c \prec b$ | 10 |
| $d \prec c \prec a \prec b$ | 8 | $c \prec d \prec a \prec b$ | 9 | $b \prec d \prec a \prec c$ | 16 | $a \prec d \prec b \prec c$ | 15 |
| $d \prec b \prec c \prec a$ | 7 | $c \prec b \prec d \prec a$ | 13 | $b \prec c \prec d \prec a$ | 14 | $a \prec c \prec d \prec b$ | 11 |
| $d \prec b \prec a \prec c$ | 13 | $c \prec b \prec a \prec d$ | 14 | $b \prec c \prec a \prec d$ | 15 | $a \prec c \prec b \prec d$ | 14 |
| $d \prec a \prec c \prec b$ | 9 | $c \prec a \prec d \prec b$ | 10 | $b \prec a \prec d \prec c$ | 17 | $a \prec b \prec d \prec c$ | 16 |
| $d \prec a \prec b \prec c$ | 12 | $c \prec a \prec b \prec d$ | 13 | $b \prec a \prec c \prec d$ | 16 | $a \prec b \prec c \prec d$ | 15 |

Table 6.8: The number of changes needed to impose monotonicity w.r.t. each ranking on $\mathscr{X}$.

The notion of monotonicity of weak votrices can be easily extended from a
votrix.
Definition 6.9 (Weak votrix). Let $\mathscr{A}$ be a set of $n$ samples and $n_{U}$ be the number of untrained panellists. A matrix $\widetilde{\mathbf{V}} \in\left\{0,1, \ldots, n_{U}\right\}^{n \times n}$ is called a weak votrix on $\mathscr{A}$ if there exists a profile $\widetilde{\mathbf{r}}$ of $n_{U}$ rankings with ties, such that for any two samples $a_{i_{1}}, a_{i_{2}} \in \mathscr{A}$, it holds that

$$
\widetilde{\mathbf{V}}_{j_{1} j_{2}}=\#\left\{i \in\left\{1, \ldots, n_{U}\right\} \mid a_{j_{2}} \precsim_{i} a_{j_{1}}\right\} .
$$

Definition 6.10 (Monotone weak votrix). A weak votrix is said to be monotone w.r.t. a ranking $\prec$ on the set $\mathscr{A}$ (with corresponding order relation $\sqsubset$ ) if, for any $\left(a_{i_{1}}, a_{j_{1}}\right),\left(a_{i_{2}}, a_{j_{2}}\right) \in \mathscr{A}^{2}$, such that $a_{i_{1}} \neq a_{j_{1}}, a_{i_{2}} \neq a_{j_{2}}$ and $\left(a_{i_{2}}, a_{j_{2}}\right) \sqsubset\left(a_{i_{1}}, a_{j_{1}}\right)$, it holds that

$$
\tilde{\mathbf{V}}_{i_{2} j_{2}} \leq \tilde{\mathbf{V}}_{i_{1} j_{1}}, \widetilde{\mathbf{V}}_{j_{1} i_{1}} \leq \tilde{\mathbf{V}}_{j_{2} i_{2}}
$$

It must be noted that there always exists the natural notion of a sample being in between two other samples. This notion is captured by the betweenness relation, as defined in Chapter 5 (Definition 5.2). Note that transitivity axioms are necessary conditions in order to guarantee the existence of an order relation that agrees with a betweenness relation. We refer to [117] for further details about order relations and their relationship with betweenness relations.

### 6.3. Consensus state problem

In most of the aforedescribed methods, deciding on the winning ranking is done by imposing certain conditions on the profile of rankings. As these important conditions allow to decide on the winner, they are called consensus states. A consensus state satisfies the following three properties:
(i) Anonymity (the principle that all panellists are treated equally, such that any permutation of (rankings provided by the) panellists should not affect the belonging to the consensus state).
(ii) Neutrality (the principle that all samples are treated equally, such that any permutation of samples in all the rankings provided by the panellists should result in the same permutation in the consensus state).
(iii) Unanimity (the principle of reaching the highest consensus state, such that if the same ranking is provided by all the panellists, then it belongs to the consensus state).

Note that every profile of rankings does not usually satisfy a consensus state due to the fact that is quite restrictive. Therefore, to aggregate a profile of rankings,
it is easier to search for the 'closest' profile of rankings in the chosen consensus state.

It has been argued that most ranking rules minimize the number of changes between the consensus state that needs to be reached from the profile of rankings [112, 150. As discussed in [27, the aggregation of rankings can generally be understood as a two-step procedure that measures the closeness to a desired consensus state in which determining the result of the aggregation of the given rankings is obvious, as follows:
(i) The 'closest' profile of rankings $\mathbf{r}^{*}$ in a consensus state is chosen based on a distance function.
(ii) The winning ranking $\prec^{*}$ is obtained by analysing the consensus state.

Note that the Borda count is one of the ranking rules that does not require any specific consensus state, since the winning ranking is determined by ordering the samples based on the total number of points they were awarded. The method of Kemeny uses the trivial notion of unanimity, where each panellist expresses the exact same ranking on the set of samples. Requiring a profile of rankings to be unanimous is a very restrictive property. Therefore, the presence of a Condorcet ranking, which is a less restrictive property, is commonly accepted. This consensus state holds when there exists a ranking such that every sample is preferred by at least $\frac{n_{U}}{2}$ panellists to all the samples ranked at a worse position. Another consensus state is that of monotonicity of the votrix, where the number of panellists providing each ranking should be decreasing on the Hasse diagram of $\sqsubset$ for the profile of rankings $\mathbf{r}$ associated with the 'true' ranking of the samples.

### 6.4. The optimization problem

In this chapter, we are interested in obtaining the 'closest' profile of rankings satisfying a certain property given a profile of rankings $\mathbf{r}$. Therefore, similar to the search for a closest matrix of labels in Chapter 5 (Section 5.5), the search for a closest profile of rankings (in the case of Condorcet ranking or monotone votrix) can also be done by solving a transportation problem [123].

Note that in our setting, each ranking in $\mathscr{R}$ is both a supply point and a demand point. The quantity of product produced at each supply point is given by $\mathbf{r}$ and, as the number of rankings needs to be preserved, all the produced units need to be transported to a demand point.

We define $(n!)^{2}$ variables $a_{u v}(u, v \in\{1, \ldots, n!\})$ taking values in the set $\mathbb{N} \cup\{0\}$, $n$ being the number of samples in $\mathscr{A}$. For any $u, v \in\{1, \ldots, n!\}, a_{u v}=m$ means that $m$ units of the $u$-th ranking in $\mathscr{R}$ are assigned to $m$ units of the $v$-th ranking in $\mathscr{R}$. We have an initial profile of rankings where the $u$-th ranking appears $s_{u}$
times. These values $s_{u}$ can be seen as the number of units of product that are produced at each supply point. The goal is to distribute these products satisfying the required monotonicity of the votrix constraint. Finally, we consider the lowest cost resulting from the transportation problem, as follows:

$$
\begin{array}{cl}
\text { Minimize } & \sum_{u=1}^{n!} \sum_{v=1}^{n!} \mathbf{C}_{u v} a_{u v} \text { w.r.t. }\left\{a_{u v}\right\} \\
\text { s.t. } & \sum_{v=1}^{n!} a_{u v}=s_{u}, \text { for any } u, v \in\{1, \ldots, n!\} \\
& a_{u v} \geq 0, \text { for any } u, v \in\{1, \ldots, n!\}, \\
& a_{u v} \in \mathbb{Z}, \text { for any } u, v \in\{1, \ldots, n!\}, \\
& P, \text { the property to be satisfied by the profile of rankings, }
\end{array}
$$

where $\mathbf{C}_{u v}$ denotes the cost of changing the $u$-th ranking in the profile of rankings $\mathbf{r}$ into the $v$-th ranking in the set of all possible rankings $\mathscr{R}$.

Thus, the optimization problem leads to the computation of a closest monotone profile of rankings w.r.t. the profile of rankings $\mathbf{r}$. This computation is done in polynomial time in terms of $n$ !. Obviously, this is a computational drawback for sets of samples of large cardinality, even though the number of food samples is typically quite small in sensory evaluation.

### 6.5. Application of ranking rules to sensory data

In this section, we apply the methods discussed in this chapter to the datasets gathered from the ranking tests in Chapter 4 where samples of chicken breasts, cod, brown shrimp and salmon samples are ranked in terms of freshness. Note that if the goal of an experiment is to obtain a consensus ranking with no ties (respectively, with ties), then the search for the 'closest' profile of rankings is measured on $\mathscr{R}$ (respectively, $\widetilde{\mathscr{R}}$ ).

### 6.5.1. Chicken breast

We consider the results of experiments L4 and H4 in each group (1, 2, 3 and 4) and for every session (1 and 2) in Table A.2, where two to four panellists ranked the chicken samples described in Table 4.3 , and we apply the aforedescribed methods to determine the consensus rankings of these samples. The consensus rankings of these samples according to the different methods are shown in Table 6.9.

Note that we will discuss the consensus rankings of sets of chicken samples with the most interesting results.

|  | Consensus ranking for group 1 |  |
| :---: | :---: | :---: |
| Method | Session 1 |  |
|  | L4 | H4 |
| Min. sum of Borda counts | $a^{9} \prec a^{5} \prec a^{7} \prec a^{0}$ | $a^{9} \prec a^{7} \sim a^{0} \prec a^{5}$ |
| Search for unanimity | $a^{9} \prec a^{5} \prec a^{7} \prec a^{0}$ | $a^{9} \prec a^{7} \prec a^{0} \prec a^{5}$ |
| Search for Condorcet ranking | $a^{9} \prec a^{5} \prec a^{7} \prec a^{0}$ | $a^{9} \prec a^{0} \prec a^{7} \prec a^{5}$ |
| Search for monotone votrix | $a^{9} \prec a^{5} \prec a^{7} \prec a^{0}$ | $a^{9} \prec a^{7} \prec a^{0} \prec a^{5}$ |
|  |  | $a^{9} \prec a^{0} \prec a^{7} \prec a^{5}$ |

Session 2

| Method | L 4 | H 4 |
| :---: | :---: | :---: |
| Min. sum of Borda counts | $a^{15} \prec a^{13} \prec a^{11}$ | $a^{11} \prec a^{13} \prec a^{15}$ |
| Search for unanimity | $a^{15} \prec a^{13} \prec a^{11}$ | $a^{11} \prec a^{13} \prec a^{15}$ |
| Search for Condorcet ranking | $a^{15} \prec a^{13} \prec a^{11}$ | $a^{11} \prec a^{13} \prec a^{15}$ |
| Search for monotone votrix | $a^{15} \prec a^{13} \prec a^{11}$ | $a^{11} \prec a^{13} \prec a^{15}$ |

Table 6.9: Consensus rankings of chicken samples described in Table 4.3 for experiments L4 and H4 in every session (1 and 2), in Table A.2 for the different methods. The symbol ${ }^{\text {(*) }}$ means that the ranking is the closest Condorcet ranking.

## Group 1

The consensus ranking of the set of chicken samples $\left\{a^{0}, a^{5}, a^{7}, a^{9}\right\}$ in session 1 of experiment L4 for group 1 according to the minimum sum of the Borda counts and according to the search for unanimity, Condorcet ranking, and monotone votrix is $a^{9} \prec a^{5} \prec a^{7} \prec a^{0}$. We conclude that all the methods yield the same result.

Based on the consensus rankings of the set of chicken samples $\left\{a^{0}, a^{5}, a^{7}, a^{9}\right\}$ in session 1 of experiment H4 for group 1 according to the different methods, we conclude that sample $a^{5}$ is most preferred, sample $a^{9}$ is least preferred, and that the order of samples $a^{0}$ and $a^{7}$ is unclear. This is due to the very small number of panellists (four) who have a disagreement on the ranking of the samples.

## Group 2

The consensus ranking of the set of chicken samples $\left\{a^{9}, a^{13}, a^{15}\right\}$ in session 2 of experiment H4 for group 1 according to the minimum sum of the Borda counts and according to the search for Condorcet ranking, and monotone votrix is $a^{15} \prec a^{13} \prec a^{9}$. We note that the consensus rankings according to the search for unanimity are $a^{15} \prec a^{13} \prec a^{9}$ and $a^{13} \prec a^{15} \prec a^{9}$. We conclude that sample $a^{9}$ is most preferred, and that it is unclear which of the samples $a^{13}$ and $a^{15}$ is least preferred. This is due to the very small number of panellists (four) who have a disagreement on the ranking of the samples.

| Method | Consensus ranking for group 2 |  |
| :---: | :---: | :---: |
|  | Session 1 |  |
|  | L4 | H4 |
| Min. sum of Borda counts | $a^{11} \prec a^{5} \prec a^{7} \prec a^{0}$ | $a^{11} \prec a^{7} \sim a^{5} \sim a^{0}$ |
| Search for unanimity | $a^{11} \prec a^{5} \prec a^{7} \prec a^{0}$ | $\begin{aligned} a^{11} & \prec a^{7} \prec a^{5} \prec a^{0} \\ a^{11} & \prec a^{7} \prec a^{0} \prec a^{5} \\ a^{11} & \prec a^{0} \prec a^{7} \prec a^{5} \\ a^{11} & \prec a^{5} \prec a^{0} \prec a^{7} \\ a^{11} & \prec a^{5} \prec a^{7} \prec a^{0} \end{aligned}$ |
| Search for Condorcet ranking | $a^{11} \prec a^{5} \prec a^{7} \prec a^{0}$ | $\begin{aligned} & a^{11} \prec a^{7} \prec a^{5} \prec a^{0 *} \\ & a^{11} \prec a^{7} \prec a^{0} \prec a^{5 *} \\ & a^{11} \prec a^{5} \prec a^{7} \prec a^{0 *} \\ & a^{11} \prec a^{5} \prec a^{0} \prec a^{7 *} \\ & a^{11} \prec a^{0} \prec a^{7} \prec a^{5 *} \\ & a^{11} \prec a^{0} \prec a^{5} \prec a^{7 *} \end{aligned}$ |
| Search for monotone votrix | $a^{11} \prec a^{5} \prec a^{7} \prec a^{0}$ | $\begin{aligned} & a^{11} \prec a^{7} \prec a^{5} \prec a^{0} \\ & a^{11} \prec a^{7} \prec a^{0} \prec a^{5} \\ & a^{11} \prec a^{5} \prec a^{7} \prec a^{0} \\ & a^{11} \prec a^{5} \prec a^{0} \prec a^{7} \\ & a^{11} \prec a^{0} \prec a^{7} \prec a^{5} \\ & a^{11} \prec a^{0} \prec a^{5} \prec a^{7} \end{aligned}$ |

Session 2

| Method | L 4 | H 4 |
| :---: | :---: | :---: |
| Min. sum of Borda counts | $a^{15} \prec a^{13} \prec a^{9}$ | $a^{13} \prec a^{15} \prec a^{9}$ |
| Search for unanimity | $a^{15} \prec a^{13} \prec a^{9}$ | $a^{13} \prec a^{15} \prec a^{9}$ |
|  | $a^{13} \prec a^{15} \prec a^{9}$ |  |
| Search for Condorcet ranking | $a^{15} \prec a^{13} \prec a^{9 *}$ | $a^{13} \prec a^{15} \prec a^{9}$ |
| Search for monotone votrix | $a^{15} \prec a^{13} \prec a^{9}$ | $a^{13} \prec a^{15} \prec a^{9}$ |

Table 6.9: (Continued) Consensus rankings of chicken samples described in Table 4.3 for experiments L4 and H4 in every session (1 and 2), in Table A.2 for the different methods. The symbol ${ }^{* *}$ means that the ranking is the closest Condorcet ranking.

## Group 3

Based on the consensus rankings of the set of chicken samples $\left\{a^{0}, a^{7}, a^{9}, a^{15}\right\}$ in session 1 of experiment L4 for group 3 according to the different methods, we conclude that sample $a^{9}$ is least preferred, followed by sample $a^{15}$, and that it is unclear which sample between samples $a^{0}$ and $a^{7}$ is most preferred. This is due to the very small number of panellists (three) who have a disagreement on the ranking of the samples.

Based on the consensus rankings of the set of chicken samples $\left\{a^{0}, a^{7}, a^{9}, a^{15}\right\}$ in

| Method | Consensus ranking for group 3 |  |
| :---: | :---: | :---: |
|  | Session 1 |  |
|  | L4 | H4 |
| Min. sum of Borda counts | $a^{9} \prec a^{15} \prec a^{7} \sim a^{0}$ | $a^{15} \sim a^{7} \prec a^{9} \prec a^{0}$ |
| Search for unanimity | $a^{9} \prec a^{15} \prec a^{0} \prec a^{7}$ | $\begin{aligned} & a^{15} \prec a^{9} \prec a^{7} \prec a^{0} \\ & a^{15} \prec a^{7} \prec a^{9} \prec a^{0} \\ & a^{15} \prec a^{7} \prec a^{0} \prec a^{9} \\ & a^{9} \prec a^{15} \prec a^{7} \prec a^{0} \\ & a^{9} \prec a^{7} \prec a^{15} \prec a^{0} \\ & a^{7} \prec a^{15} \prec a^{9} \prec a^{0} \\ & a^{7} \prec a^{15} \prec a^{0} \prec a^{9} \\ & a^{7} \prec a^{9} \prec a^{15} \prec a^{0} \end{aligned}$ |
| Search for Condorcet ranking | $a^{9} \prec a^{15} \prec a^{0} \prec a^{7}$ | $\begin{aligned} & a^{15} \prec a^{9} \prec a^{7} \prec a^{0 *} \\ & a^{9} \prec a^{15} \prec a^{7} \prec a^{0 *} \end{aligned}$ |
| Search for monotone votrix | $\begin{aligned} & a^{9} \prec a^{15} \prec a^{7} \prec a^{0} \\ & a^{9} \prec a^{15} \prec a^{0} \prec a^{7} \end{aligned}$ | $\begin{aligned} & a^{15} \prec a^{7} \prec a^{9} \prec a^{0} \\ & a^{7} \prec a^{15} \prec a^{9} \prec a^{0} \end{aligned}$ |

Session 2

| Method | L 4 | H 4 |
| :---: | :---: | :---: |
| Min. sum of Borda counts | $a^{11} \prec a^{13} \prec a^{5}$ | $a^{11} \prec a^{13} \prec a^{5}$ |
| Search for unanimity | $a^{11} \prec a^{13} \prec a^{5}$ | $a^{11} \prec a^{13} \prec a^{5}$ |
| Search for Condorcet ranking | $a^{11} \prec a^{13} \prec a^{5 *}$ | $a^{11} \prec a^{13} \prec a^{5}$ |
| Search for monotone votrix | $a^{11} \prec a^{13} \prec a^{5}$ | $a^{11} \prec a^{13} \prec a^{5}$ |

Table 6.9: (Continued) Consensus rankings of chicken samples described in Table 4.3 for experiments L4 and H4 in every session (1 and 2), in Table A.2 for the different methods. The symbol ${ }^{* *}$, means that the ranking is the closest Condorcet ranking.
session 1 of experiment H4 according to the different methods, the overall ranking of these samples is unclear. This is due to the very small number of panellists (two) who have a disagreement on the ranking of the samples.

## Group 4

Based on the consensus rankings of the set of chicken samples $\left\{a^{0}, a^{9}, a^{15}\right\}$ in session 2 of experiment L4 for group 4 according to the different methods, we conclude that sample $a^{0}$ is most preferred, and that it is unclear which of the samples $a^{9}$ and $a^{15}$ is least preferred. This is due to the very small number of panellists (four) who have a disagreement on the ranking of the samples.

We now consider the results of experiments L8 and H8 in Table A.3, where fourteen panellists ranked the chicken samples described in Table 4.3, and we apply the aforedescribed methods to determine the consensus rankings of these samples. The consensus rankings of these samples according to the different methods are shown

|  | Consensus ranking for group 4 |  |  |
| :---: | :---: | :---: | :---: |
| Method | Session 1 |  |  |
|  | L4 |  |  |
| Min. sum of Borda counts | $a^{13} \prec a^{11} \prec a^{5} \prec a^{7}$ | $a^{11} \prec a^{13} \prec a^{7} \prec a^{5}$ |  |
| Search for unanimity | $a^{13} \prec a^{11} \prec a^{5} \prec a^{7}$ | $a^{11} \prec a^{13} \prec a^{7} \prec a^{5}$ |  |
| Search for Condorcet ranking | $a^{13} \prec a^{11} \prec a^{5} \prec a^{7}$ | $a^{11} \prec a^{13} \prec a^{7} \prec a^{5}$ |  |
| Search for monotone votrix | $a^{13} \prec a^{11} \prec a^{5} \prec a^{7}$ | $a^{11} \prec a^{13} \prec a^{7} \prec a^{5}$ |  |

Session 2

| Method | L4 | H4 |
| :---: | :---: | :---: |
| Min. sum of Borda counts | $a^{15} \sim a^{9} \prec a^{0}$ | $a^{9} \prec a^{15} \prec a^{0}$ |
| Search for unanimity | $a^{15} \prec a^{9} \prec a^{0}$ | $a^{9} \prec a^{15} \prec a^{0}$ |
|  | $a^{9} \prec a^{15} \prec a^{0}$ |  |
| Search for Condorcet ranking | $a^{15} \prec a^{9} \prec a^{0 *}$ <br> $a^{9} \prec a^{15} \prec a^{0 *}$ | $a^{9} \prec a^{15} \prec a^{0}$ |
|  | $a^{15} \prec a^{9} \prec a^{0}$ <br> $a^{9} \prec a^{15} \prec a^{0}$ | $a^{9} \prec a^{15} \prec a^{0}$ |

Table 6.9: (Continued) Consensus rankings of chicken samples described in Table 4.3 for experiments L4 and H4 in every session (1 and 2), in Table A.2 for the different methods. The symbol '*' means that the ranking is the closest Condorcet ranking.
in Table 6.10.

|  | Consensus ranking |  |
| :---: | :---: | :---: |
| Method | L8 | H8 |
| Min. sum of Borda counts | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ | $a^{6} \prec a^{5} \prec a^{4} \prec a^{0} \prec a^{2}$ |
| Search for unanimity | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ | $a^{6} \prec a^{5} \prec a^{4} \prec a^{0} \prec a^{2}$ |
|  | $a^{5} \prec a^{6} \prec a^{4} \prec a^{2} \prec a^{0}$ | $a^{6} \prec a^{5} \prec a^{0} \prec a^{4} \prec a^{2}$ |
|  |  | $a^{6} \prec a^{5} \prec a^{0} \prec a^{2} \prec a^{4}$ |
| Search for Condorcet ranking | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ | $a^{6} \prec a^{5} \prec a^{4} \prec a^{0} \prec a^{2 *}$ |
|  |  | $a^{6} \prec a^{5} \prec a^{0} \prec a^{4} \prec a^{2 *}$ |
| Search for monotone votrix | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ |

Table 6.10: Consensus rankings of chicken samples described in Table 4.3 for experiments L8 and H8, in Table A.3 for the different methods. The symbol ${ }^{* *}$ means that the ranking is the closest Condorcet ranking.

The consensus ranking of the set of chicken samples $\left\{a^{0}, a^{2}, a^{4}, a^{5}, a^{6}\right\}$ in experiment L8 according to the minimum sum of the Borda counts and according to the search for Condorcet ranking, and monotone votrix is $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$. We note that the consensus rankings according to the search for unanimity are
$a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ and $a^{5} \prec a^{6} \prec a^{4} \prec a^{2} \prec a^{0}$. Therefore, we conclude that sample $a^{0}$ is most preferred, followed by samples $a^{2}$ and $a^{4}$, in that order, and that it is unclear which of the samples $a^{5}$ and $a^{6}$ is least preferred. This is due to a slight disagreement among the panellists, where one half of the panellists has preferred sample $a^{5}$ over $a^{6}$, while the other half of the panellists has preferred sample $a^{6}$ over $a^{5}$, as shown in Table A.3.

Based on the consensus rankings of the set of chicken samples $\left\{a^{0}, a^{2}, a^{4}, a^{5}, a^{6}\right\}$ in experiment H8 according to the different methods, we conclude that sample $a^{6}$ is least preferred, followed by sample $a^{5}$, and that the order of samples $\left\{a^{0}, a^{2}, a^{4}\right\}$ according to preference is unclear. This is due to a large disagreement among the panellists.

### 6.5.2. Atlantic cod

We consider the results of experiments L4, H4 and H8 for every session (1 and 2) and experiments L8 and A4 in Table A.6. where eight to ten panellists ranked the cod samples described in Table 4.6, and we apply the aforedescribed methods to determine the consensus rankings of these samples. The consensus rankings of these samples according to the different methods are shown in Table 6.11.

Note that we will discuss the consensus rankings of sets of cod samples with the most interesting results.

The consensus ranking of the set of cod samples $\left\{a^{0}, a^{4}, a^{8}, a^{13}\right\}$ in session 1 of experiment L4 according to the minimum sum of the Borda counts and according to the search for unanimity, Condorcet ranking, and monotone votrix is $a^{13} \prec a^{8} \prec a^{4} \prec a^{0}$. We conclude that all the methods yield the same result.

Based on the consensus rankings of the set of cod samples $\left\{a^{0}, a^{3}, a^{5}, a^{7}\right\}$ in session 1 of experiment H 8 according to the different methods, we conclude that sample $a^{0}$ is most preferred, followed by sample $a^{3}$, and that it is unclear which of the samples $a^{5}$ and $a^{7}$ is least preferred. This is due to a slight disagreement among the panellists, where one half of the panellists has preferred sample $a^{5}$ over $a^{7}$, while the other half of the panellists has preferred sample $a^{7}$ over $a^{5}$, as shown in Table A. 6

Based on the consensus rankings of the set of cod samples $\left\{a^{3}, a^{4}, a^{5}, a^{6}\right\}$ in session 2 of experiment H8 according to the different methods, we conclude that sample $a^{5}$ is least preferred, followed by sample $a^{6}$, and that it is unclear which of the samples $a^{3}$ and $a^{4}$ is most preferred. This is due to a slight disagreement among the panellists, as shown in Table A. 6 .

Based on the consensus rankings of the set of cod samples $\left\{a^{0}, a^{3}, a^{5}, a^{7}\right\}$ in experiment L8 according to the different methods, we conclude that sample $a^{0}$ is most preferred, followed by sample $a^{3}$, and that it is unclear which of the samples $a^{5}$

|  | Consensus ranking |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Method | Session 1 |  |  |  |
|  | L4 | H4 |  |  |
| Min. sum of Borda counts | $a^{13} \prec a^{8} \prec a^{4} \prec a^{0}$ | $a^{8} \prec a^{13} \prec a^{6} \prec a^{4}$ | $a^{5} \prec a^{7} \prec a^{3} \prec a^{0}$ |  |
| Search for unanimity | $a^{13} \prec a^{8} \prec a^{4} \prec a^{0}$ | $a^{8} \prec a^{13} \prec a^{6} \prec a^{4}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |  |
| $a^{5} \prec a^{7} \prec a^{3} \prec a^{0}$ |  |  |  |  |

Session 2

|  | Session 2 |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Method | L4 | H 4 | H 8 |  |
| Min. sum of Borda counts | $a^{8} \prec a^{7} \prec a^{6} \prec a^{5}$ | $a^{11} \prec a^{8} \prec a^{7} \prec a^{6}$ | $a^{5} \prec a^{6} \prec a^{4} \prec a^{3}$ |  |
| Search for unanimity | $a^{8} \prec a^{7} \prec a^{6} \prec a^{5}$ | $a^{11} \prec a^{8} \prec a^{7} \prec a^{6}$ | $a^{5} \prec a^{6} \prec a^{4} \prec a^{3}$ |  |
|  |  |  | $a^{5} \prec a^{6} \prec a^{3} \prec a^{4}$ |  |
| Search for Condorcet ranking | $a^{8} \prec a^{7} \prec a^{6} \prec a^{5}$ | $a^{11} \prec a^{8} \prec a^{7} \prec a^{6}$ | $a^{5} \prec a^{6} \prec a^{4} \prec a^{3 *}$ |  |
|  |  |  | $a^{5} \prec a^{6} \prec a^{3} \prec a^{4 *}$ |  |
| Search for monotone votrix | $a^{8} \prec a^{7} \prec a^{5} \prec a^{6}$ | $a^{11} \prec a^{8} \prec a^{7} \prec a^{6}$ | $a^{5} \prec a^{6} \prec a^{4} \prec a^{3}$ |  |
|  |  |  |  |  |
| Method |  |  |  |  |


| Min. sum of Borda counts | $a^{7} \sim a^{5} \prec a^{3} \prec a^{0}$ | $a^{3} \prec a^{2} \sim a^{0} \prec a^{1}$ |
| :---: | :--- | :--- |
| Search for unanimity | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{3} \prec a^{2} \prec a^{0} \prec a^{1}$ |
| Search for Condorcet ranking | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{3} \prec a^{2} \prec a^{0} \prec a^{1}$ |
| Search for monotone votrix | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{3} \prec a^{2} \prec a^{0} \prec a^{1}$ |
|  | $a^{5} \prec a^{7} \prec a^{3} \prec a^{0}$ | $a^{3} \prec a^{0} \prec a^{2} \prec a^{1}$ |

Table 6.11: Consensus rankings of cod samples described in Table 4.6 for experiments L4, H4 and H8 in every session (1 and 2) and experiments L8 and A4, in Table A.6. for the different methods. The symbol ${ }^{* *}$, means that the ranking is the closest Condorcet ranking.
and $a^{7}$ is least preferred. This is due to a slight disagreement among the panellists, where five panellists have preferred sample $a^{5}$ over $a^{7}$, while the remaining four panellists have preferred sample $a^{7}$ over $a^{5}$, as shown in Table A. 6 .

Based on the consensus rankings of the set of cod samples $\left\{a^{0}, a^{1}, a^{2}, a^{3}\right\}$ in experiment A4 according to the different, we conclude that sample $a^{1}$ is most preferred, sample $a^{3}$ is least preferred, and that the order of samples $a^{0}$ and $a^{2}$ is unclear. This is due to a slight disagreement among the panellists, where five panellists have preferred sample $a^{0}$ over $a^{2}$, while the remaining three panellists have preferred sample $a^{2}$ over $a^{0}$, as shown in Table A. 6 .

### 6.5.3. Brown shrimp

We consider the results of experiments L 4 and H 4 for every session (1 and 2) in Table A.8, where eight to ten panellists ranked the shrimp samples described in Table 4.9, and we apply the aforedescribed methods to determine the consensus rankings of these samples. The consensus rankings of these samples according to the different methods are shown in Table 6.12,

|  | Consensus ranking |  |
| :---: | :---: | :---: |
| Method | Session 1 |  |
|  | L4 | H4 |
| Min. sum of Borda counts | $a^{10} \prec a^{5} \prec a^{3} \sim a^{0}$ | $a^{12} \prec a^{7} \prec a^{3} \prec a^{0}$ |
| Search for unanimity | $a^{10} \prec a^{5} \prec a^{0} \prec a^{3}$ | $a^{12} \prec a^{7} \prec a^{3} \prec a^{0}$ |
| Search for Condorcet ranking | $a^{10} \prec a^{5} \prec a^{0} \prec a^{3}$ | $a^{12} \prec a^{7} \prec a^{3} \prec a^{0}$ |
| Search for monotone votrix | $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{12} \prec a^{7} \prec a^{3} \prec a^{0}$ |

Session 2

| Method | L4 | H4 |
| :---: | :---: | :---: |
| Min. sum of Borda counts | $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{7} \prec a^{5} \prec a^{3} \sim a^{0}$ |
| Search for unanimity | $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |
| Search for Condorcet ranking | $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |
| Search for monotone votrix | $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{7} \prec a^{5} \prec a^{0} \prec a^{3}$ |

Table 6.12: Consensus rankings of shrimp samples described in Table 4.9 for experiments L4 and H4 in every session (1 and 2), in Table A.8 for the different methods.

We note that the methods show multiple consensus rankings of the set of shrimp samples $\left\{a^{0}, a^{3}, a^{5}, a^{10}\right\}$ in session 1 of experiment L4. However, we deduce that sample $a^{1} 0$ is least preferred, followed by sample $a^{5}$, and that it is unclear which of the samples $a^{0}$ and $a^{3}$ is most preferred. This is due to a slight disagreement among the panellists, where five panellists have preferred sample $a^{0}$ over $a^{3}$, while the remaining four panellists have preferred sample $a^{3}$ over $a^{0}$, as shown in Table A.8.

We conclude that all the methods yield the consensus ranking $a^{12} \prec a^{7} \prec a^{3} \prec a^{0}$ for the set of shrimp samples $\left\{a^{0}, a^{3}, a^{7}, a^{12}\right\}$ in session 1 of experiment H4, and yield the consensus ranking $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ of the set of shrimp samples $\left\{a^{0}, a^{3}, a^{5}, a^{10}\right\}$ in session 2 of experiment L4.

Based on the consensus rankings of the set of shrimp samples $\left\{a^{0}, a^{3}, a^{5}, a^{7}\right\}$ in session 2 of experiment H 4 according to the different methods, we conclude that
sample $a^{7}$ is least preferred, followed by sample $a^{5}$, and that it is unclear which of the samples $a^{0}$ and $a^{3}$ is most preferred. This is due to a slight disagreement among the panellists, where six panellists have preferred sample $a^{0}$ over $a^{3}$, while the remaining four panellists have preferred sample $a^{3}$ over $a^{0}$, as shown in Table A. 8 .

### 6.5.4. Atlantic salmon

## Session 1

First, we consider the results of experiments H4, AN4, ANH4, A4, L4 and M4 for session 1 in Table A.10, where eight to twelve panellists ranked the salmon samples described in Table 4.12, and we apply the aforedescribed methods to determine the consensus rankings of these samples. The consensus rankings of these samples according to the different methods are shown in Table 6.13 .

Note that we will discuss the consensus rankings of sets of salmon samples with the most interesting results.

| Method | Consensus ranking |  |  |
| :---: | :---: | :---: | :---: |
|  | Session 1 |  |  |
|  | H4 | AN4 | ANH4 |
| Min. sum of Borda counts | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{13} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| Search for unanimity | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $\begin{aligned} & a^{13} \prec a^{9} \prec a^{5} \prec a^{1} \\ & a^{9} \prec a^{13} \prec a^{5} \prec a^{1} \end{aligned}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| Search for Condorcet ranking | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $\begin{aligned} & a^{13} \prec a^{9} \prec a^{5} \prec a^{1 *} \\ & a^{9} \prec a^{13} \prec a^{5} \prec a^{1 *} \end{aligned}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| Search for monotone votrix | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{9} \prec a^{13} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| Method | A4 | L4 | M4 |
| Min. sum of Borda counts | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{7} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| Search for unanimity | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{7} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| Search for Condorcet ranking | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{7} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| Search for monotone votrix | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{7} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |

Table 6.13: Consensus rankings of salmon samples described in Table 4.12 for experiments H4, AN4, ANH4, A4, L4 and M4 in session 1 in Table A.10 for the different methods. The symbol ${ }^{* *}$ means that the ranking is the closest Condorcet ranking.

The consensus ranking of the set of salmon samples $\left\{a^{1}, a^{5}, a^{9}, a^{11}\right\}$ in session 1 of experiment H4 according to the minimum sum of the Borda counts and according to the search for unanimity, Condorcet ranking, and monotone votrix is $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$. We conclude that all the methods yield the same result.

Based on the consensus rankings of the set of salmon samples $\left\{a^{1}, a^{5}, a^{9}, a^{13}\right\}$ in session 1 of experiment AN4 according to the different methods, we conclude that sample $a^{1}$ is most preferred, followed by sample $a^{5}$, and that it is unclear which of the samples $a^{9}$ and $a^{13}$ is most preferred. This is due to a slight disagreement among the panellists, where one half of the panellists has preferred sample $a^{9}$ over $a^{13}$, while the other half of the panellists has preferred sample $a^{13}$ over $a^{9}$, as shown in Table A. 10 .

## Session 2

Second, we consider the results of experiments H4, AN4, ANH4 and A4 for the other three sessions (2, 3 and 4) in Table A.10. where five to twelve panellists ranked the salmon samples described in Table 4.12, and we apply the aforedescribed methods to determine the consensus rankings of these samples. The consensus rankings of these samples according to the different methods are shown in Table 6.14

|  | Consensus ranking |  |
| :---: | :--- | :---: |
| Method | Session 2 |  |
|  | H4 |  |
| Min. sum of Borda counts | $a^{7} \prec a^{5} \prec a^{3} \sim a^{1}$ | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ |
| Search for unanimity | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ |
| Search for Condorcet ranking | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ |
| Search for monotone votrix | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ |
| Method |  |  |
| ANH4 | A4 |  |
| Min. sum of Borda counts | $a^{7} \prec a^{3} \sim a^{1} \prec a^{5}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| Search for unanimity | $a^{7} \prec a^{3} \prec a^{1} \prec a^{5}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| Search for Condorcet ranking | $a^{7} \prec a^{3} \prec a^{1} \prec a^{5}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| Search for monotone votrix | $a^{7} \prec a^{3} \prec a^{1} \prec a^{5}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| $a^{7} \prec a^{1} \prec a^{3} \prec a^{5}$ |  |  |

Table 6.14: Consensus rankings of salmon samples described in Table 4.12 for experiments H4, AN4, ANH4 and A4 in sessions 2, 3 and 4 in Table A.10, for the different methods. The symbol ${ }^{* *}$, means that the ranking is the closest Condorcet ranking.

The consensus ranking of the set of salmon samples $\left\{a^{1}, a^{3}, a^{5}, a^{7}\right\}$ in session 2 of experiment H 4 according to the search for unanimity, Condorcet ranking, and monotone votrix is $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$. We note that the consensus ranking according to the minimum sum of the Borda counts is $a^{7} \prec a^{5} \prec a^{3} \sim a^{1}$. We conclude that sample $a^{7}$ is least preferred, followed by sample $a^{5}$, and that it is unclear which of the samples $a^{3}$ and $a^{1}$ is most preferred. This is due to a slight disagreement among the panellists, where five panellists have preferred sample $a^{3}$ over $a^{1}$, while the remaining three panellists have preferred sample $a^{1}$ over $a^{3}$, as shown in

|  | Consensus ranking |  |
| :---: | :---: | :---: |
| Method | Session 3 |  |
|  | H4 | AN4 |
| Min. sum of Borda counts | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| Search for unanimity | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| Search for Condorcet ranking | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| Search for monotone votrix | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
|  |  |  |
| Method | ANH4 |  |
| Min. sum of Borda counts | $a^{11} \prec a^{5} \prec a^{9} \prec a^{1}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ |
| Search for unanimity | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ |
|  | $a^{11} \prec a^{5} \prec a^{9} \prec a^{1}$ |  |
| $a^{11} \prec a^{5} \prec a^{1} \prec a^{9}$ |  |  |
| Search for Condorcet ranking | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1 *}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ |
|  | $a^{11} \prec a^{5} \prec a^{9} \prec a^{1 *}$ |  |
| Search for monotone votrix | $a^{11} \prec a^{5} \prec a^{9} \prec a^{1}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ |

Table 6.14: Consensus rankings of salmon samples described in Table 4.12 for experiments H4, AN4, ANH4 and A4 in sessions 2, 3 and 4 in Table A. 10 for the different methods. The symbol ${ }^{* *}$ means that the ranking is the closest Condorcet ranking.

|  | Consensus ranking |  |
| :---: | :---: | :---: |
| Method | Session 4 |  |
| H4 | AN4 |  |
| Min. sum of Borda counts | $a^{9} \prec a^{7} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ |
| Search for unanimity | $a^{9} \prec a^{7} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ |
| Search for Condorcet ranking | $a^{9} \prec a^{7} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ |
| Search for monotone votrix | $a^{9} \prec a^{7} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ |
| Method | ANH4 |  |
| Min. sum of Borda counts | $a^{7} \prec a^{5} \sim a^{1} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| Search for unanimity | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| Search for Condorcet ranking | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| Search for monotone votrix | $a^{7} \prec a^{1} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |

Table 6.14: (Continued) Consensus rankings of salmon samples described in Table 4.12 for experiments H4, AN4, ANH4 and A4 in sessions 2, 3 and 4 in Table A. 10 for the different methods. The symbol ${ }^{* *}$ means that the ranking is the closest Condorcet ranking.

Table A. 10
Based on the consensus rankings of the set of salmon samples $\left\{a^{1}, a^{3}, a^{5}, a^{7}\right\}$ in session 2 of experiment ANH4 according to the different methods. However, we deduce that sample $a^{5}$ is most preferred, sample $a^{7}$ is least preferred, and that the order of the samples $a^{1}$ and $a^{3}$ is unclear. This is due to the small number of panellists (five) who have slight disagreement, as shown in Table A. 10 .

## Session 3

Based on the consensus rankings of the set of salmon samples $\left\{a^{1}, a^{5}, a^{9}, a^{11}\right\}$ in session 3 of experiment ANH4 according to the different methods, we conclude that sample $a^{11}$ is least preferred, and that the order of the samples $a^{1}, a^{5}$ and $a^{9}$ is unclear. This is due to a large disagreement among the panellists, as shown in Table A. 10

## Session 4

The consensus ranking of the set of salmon samples $\left\{a^{1}, a^{3}, a^{5}, a^{7}\right\}$ in session 4 of experiment ANH4 according to the search for unanimity, Condorcet ranking, and monotone votrix is $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$. We note that the consensus ranking according to the minimum sum of the Borda counts and according to is $a^{7} \prec a^{5} \sim a^{1} \prec a^{3}$. We deduce that sample $a^{3}$ is most preferred, sample $a^{7}$ is least preferred, and that the order of samples $a^{1}$ and $a^{5}$ is unclear. This is due to a slight disagreement among the panellists, where five panellists have preferred sample $a^{1}$ over $a^{5}$, while the other four panellists have preferred sample $a^{5}$ over $a^{1}$, as shown in Table A. 10

## Ranking with ties

Based on the consensus rankings of the set of salmon samples in every group, we conclude that sample $\mathrm{D}^{5}$ is preferred the least followed by sample $\mathrm{C}^{4}$ in group 2, sample $\mathrm{C}^{5}$ is preferred the least followed by sample $\mathrm{D}^{6}$ in group 3 , sample $\mathrm{D}^{7}$ is preferred the least and sample $\mathrm{A}^{4}$ is preferred the most in group 4, and sample $\mathrm{C}^{7}$ is preferred the least followed by sample $\mathrm{D}^{8}$ in group 5.

### 6.6. Conclusions

In this chapter, we have discussed methods that are characterized by a distance function and a consensus state. In addition, we have discussed the use of monometrics that are endorsed by the fact that they better fit the nature of the problem of looking for the 'closest' profile of rankings in the chosen consensus state. We have presented the search of these 'closest' profiles in a consensus state as an optimization problem. Finally, the discussed methods have been illustrated using

| Method | Group 1 | Group 2 | Troup 3 |
| :---: | :---: | :---: | :---: |
| Thursday | Monday |  |  |
| Min. sum of Borda counts | $\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~A}^{1} \prec \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~A}^{3} \prec \mathrm{~B}^{4}$ |
|  | $\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~B}^{2} \sim \mathrm{~A}^{1}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \sim \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \sim \mathrm{~A}^{3} \prec \mathrm{~B}^{4}$ |
| Search for unanimity | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \sim \mathrm{~B}^{3} \sim \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~B}^{4} \sim \mathrm{~A}^{3}$ |  |
|  | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \sim \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ |  |  |
|  | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ |  |  |


| Search for Condorcet weak order | $\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~A}^{1} \prec \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2} \quad \mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~A}^{3} \prec \mathrm{~B}^{4}$ |  |
| :--- | :--- | :--- | :--- |
| Search for weak monotone votrix | $\mathrm{D}^{4} \prec \mathrm{C}^{3} \prec \mathrm{~A}^{1} \prec \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~A}^{3} \prec \mathrm{~B}^{4}$ |
|  | $\mathrm{D}^{4} \prec \mathrm{C}^{3} \prec \mathrm{~B}^{2} \sim \mathrm{~A}^{1}$ |  |  |

Group $4 \quad$ Group 5

| Method | Wednesday | Friday |
| :---: | :--- | :--- |
| Min. sum of Borda counts | $\mathrm{D}^{7} \prec \mathrm{~B}^{5} \prec \mathrm{C}^{6} \prec \mathrm{~A}^{4}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| Search for unanimity | $\mathrm{D}^{7} \prec \mathrm{C}^{6} \sim \mathrm{~B}^{5} \prec \mathrm{~A}^{4}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| Search for Condorcet weak order | $\mathrm{D}^{7} \prec \mathrm{~B}^{5} \prec \mathrm{C}^{6} \prec \mathrm{~A}^{4}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| Search for weak monotone votrix | $\mathrm{D}^{7} \prec \mathrm{~B}^{5} \prec \mathrm{C}^{6} \prec \mathrm{~A}^{4}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
|  | $\mathrm{D}^{7} \prec \mathrm{C}^{6} \sim \mathrm{~B}^{5} \prec \mathrm{~A}^{4}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~B}^{6} \sim \mathrm{~A}^{5}$ |

Table 6.15: The consensus rankings of salmon samples (A, B, C, D) in each group (1-4).
the sensory data gathered in Chapter 4 to obtain the consensus ranking of each ranking experiment.

As we have previously mentioned, trained panellists and untrained panellists may provide different kinds of information, such as, absolute and relative information, respectively. These types of information can be combined in order to exploit the information expressed by both trained and untrained panellists. In Chapter 7, we introduce a first attempt at integrating absolute information, in the form of vectors of scores, with rankings to improve the accuracy of the consensus ranking.

## 7 Combining scores and rankings

## Table of Contents

| 7.1 | Introduction |
| :---: | :---: |
| 7.2 | Obtaining consensus scores |
| 7.3 | Obtaining consensus rankings |
| 7.4 | Integrating scores and rankings |
|  | 7.4.1 Improving the quality of the assessment of |
|  | a consensus vector of scores |
|  | 7.4.2 Improving the quality of the assessment of |
|  | a consensus ranking |
| 7.5 | Information about samples |
|  | 7.5.1 Knowledge of storage days |
|  | 7.5.2 Results of clustering analysis |
|  | 7.5.3 Other sensory evaluation tests |
|  | 7.5.4 The constrained mode, median and mean |
| 7.6 | Application to sensory data |
| 7.7 | Integrating rankings for assigning consensus scores |
|  | 7.7.1 Integrating scores for determining consensus rankings |
|  | 7.7.2 Incorporating knowledge of storage days for |
|  | assigning joint consensus scores |
|  | 7.7.3 Incorporating results of a clustering analysis for |
|  | assigning joint consensus scores |
|  | 7.7.4 Incorporating consensus rankings for |
|  | assigning joint consensus scores |
|  | 7.7.5 Comparing the consensus vectors of scores |
| 7.8 | Conclusions |

### 7.1. Introduction

So far, we have discussed sensory evaluation performed by either trained (labelling) or untrained (ranking) panels, however, it has been shown that both approaches have their limitations [151, 21]. On the one hand, trained panellists are a limited and (in some cases) very expensive source of (absolute) information. As a result, there usually is a limited amount of data available to obtain a reasonable consensus evaluation and determine the overall quality of a sample. On the other hand, Rodrigue et al. [151] have suggested that one should instead consider untrained panellists, with the obvious limitation of requiring a much greater number of
panellists. Interestingly, untrained panellists could provide a cost-efficient source of additional information. It is common that untrained panellists rank (two or more) samples according to their personal degree of appreciation, yielding relative information [24]. Conceptually, ranking of samples is recognized as a simpler task than assigning scores [31] and is thus preferred for untrained panellists. Obviously, absolute and relative information have different properties [152], and the methods for gathering them have different limitations.

In this chapter, we focus on absolute and relative information in the form of scores and rankings, respectively. Scores are typically discrete, equally spaced and non-decimal values on an ordinal scale, thus, allowing for straightforward analysis to determine the overall quality of a sample.

Studies in the field of social science on values and preference have recommended the use of both scoring and ranking methods [153, 154] to provide a complete understanding of the appreciation of samples. Recently, there has been a simultaneous adoption of scoring methods for determining the quality of food samples and ranking methods for determining the existence of a significant difference between the samples [52, 155, 156, 157]. However, the data resulting from these methods were dealt with separately. Therefore, we advocate for a combined scoring and ranking approach, resulting in a method for simultaneously exploiting the scores and rankings, something that has not been previously developed.

In this chapter, we first propose an approach for integrating ${ }^{1}$ rankings provided by untrained panellists with scores provided by trained panellists to improve the assessment of the quality of the consensus score. We propose a second approach for integrating scores with rankings to improve the assessment of the quality of the consensus ranking. We make a first attempt at a combined scoring and ranking approach to determine the consensus score that describes the overall quality of a food sample or to determine the consensus ranking that describes the overall ranking of multiple food samples in terms of preference by considering the median and the Kemeny median because of their similarities (both of them minimize an $\ell_{1}$-type of distance).

On the one hand, quality control, where trained panellists could probably never be replaced by untrained panellists [158], is a great example of where the developed method for determining a consensus score of a sample would be of most use. Research has shown that untrained panellists may be not be well-equiped in accurately detecting spoilage, however, they are capable of detecting differences among samples [21. In a real-life setting, the number of trained panellists is limited, thus, additional information obtained from untrained panellists can be exploited. On the other hand, product development, where consumers (untrained panellists) could never be represented by trained panellists, is another great example of where

[^22]the developed method for determining a consensus ranking of samples would be of use. However, in a real-life setting, the number of untrained panellists can be limited at times. Additional information obtained from a smaller number of trained panellists would be of higher quality and utmost importance, and, thus, such information must be exploited alongside.

In food evaluation, some unexploited information about the samples, retrieved from earlier sensory and other non-sensory sources, is known beforehand. Typical examples of such information include: the storage days of the food sample, previously performed sensory evaluation tests, such as ranking, discriminative, or threshold detection tests, or previously performed clustering analysis of chemical data of the samples. These types of information are usually relative and hint at some relations between the consensus scores of the food samples. In this chapter, we make use of such additional information of several food samples by incorporating ${ }^{2}$ them into the scores provided by trained panellists to jointly find the consensus scores of these samples.

In this chapter, we will answer the following question:
Question III.3: How can we combine scores and rankings to reach an improved consensus evaluation?

We will discuss our proposed methods for combining scores assigned by trained panellists and rankings provided by untrained panellists to reach an improved consensus evaluation (score or ranking). The application of these methods will be illustrated on the sensory data in Chapter 4 .

### 7.2. Obtaining consensus scores

We start with a description of the methods that constitute the main building blocks of our approach. We denote by $a_{j}$ the $j$-th sample in a set $\mathscr{A}=\left\{a_{1}, \ldots, a_{n}\right\}$ of $n$ samples.

We consider the setting where $n_{T}$ trained panellists each have assigned a score on a 5 -point scale to each of the samples. Note that other $k$-point scales can also be used, such that $k \in \mathbb{N}$, with a preference for an odd number of points allowing for a neutral response in case of bipolar scales [159]. Typically, labels are placed on the scale to specify the sensory attribute to be evaluated, such as intensity, preference, overall quality, etc. The scale we use throughout this chapter is shown in Figure 4.2. We denote by $s_{i j}$ the score assigned by the $i$-th panellist to sample $a_{j}$, where $i \in\left\{1, \ldots, n_{T}\right\}$ and $j \in\{1, \ldots, n\}$. We denote by $s_{i}$ the vector of scores $\left(s_{i 1}, \ldots, s_{i n}\right)$ assigned by the $i$-th panellist.

[^23]The goal is to agree on the consensus score that should be assigned to each sample in $\mathscr{A}$. We denote by $\partial\left(s, s^{\prime}\right)$ the distance between two scores and by $d\left(\mathbf{s}, \mathbf{s}^{\prime}\right)$ the distance between two constant vectors of scores. Since the scale used in this chapter allows for distances, the distance $d$ can be defined as the sum of distances $\partial$ for all the $n$ scores. One could note that several examples of this procedure are commonly used in practice. The simplest method for assigning a consensus score given a list of scores is the mode, i.e., the score that appears with the highest frequency in the scores provided by the panellists. This amounts to computing the sum of zero-one distances. Another common technique is that of the median, i.e., the score that separates the lower half from the higher half of the scores provided by the panellists. This amounts to computing the sum of absolute distances. Arguably, the third most common method is based on the notion of (arithmetic) mean, i.e., the average of the scores provided by the panellists. In case the considered scale is a continuous segment of the real line, this amounts to computing the root of the sum of squared differences between two scores.

Repeating the preceding procedure for all samples is equivalent to the following. In order to describe the three methods at the same time, we consider the more general setting in which, for each possible vector of scores, we compute the vector of scores $\mathbf{s}^{*}$ that minimizes the sum of distances $d$ (for a fixed distance function $\partial$ ) to the vectors of scores assigned by the trained panellists, as follows:

$$
\begin{equation*}
\mathbf{s}^{*}=\underset{\mathbf{s} \in\{1, \ldots, k\}^{n}}{\arg \min } \sum_{i=1}^{n_{T}} d\left(\mathbf{s}, \mathbf{s}_{i}\right) \tag{7.1}
\end{equation*}
$$

The minimizer $\mathbf{s}^{*}$ is thus assigned as the consensus vector of scores (note that there can be multiple minimizers $\mathbf{s}^{*}$ ). To determine $\mathbf{s}^{*}$, the distance $d$ is computed for all possible vectors of scores, where the number of $n$-tuples of a set of 5 scores (also known as arrangements with repetition) is equal to $5^{n} \quad 160$.

Example 7.1. Consider a simple setting where three panellists each assign a score to four samples $a_{1}, a_{2}, a_{3}$ and $a_{4}$ on the 5-point scale fixed in Figure 4.2. The scores assigned by the panellists are summarized in Table 7.1.

|  | $a_{1}$ | $a_{2}$ | $a_{3}$ | $a_{4}$ |
| :---: | :---: | :---: | :---: | :---: |
| $\mathbf{s}_{1}$ | 2 | 1 | 3 | 3 |
| $\mathbf{s}_{2}$ | 2 | 2 | 5 | 4 |
| $\mathbf{s}_{3}$ | 1 | 2 | 3 | 4 |

Table 7.1: The scores assigned to the samples $a_{1}, a_{2}, a_{3}$ and $a_{4}$ by the trained panellists in Example 7.1.

To determine the consensus score that should be assigned to each of these samples,
we consider the problem defined by Eq. 7.1), and, for each of the 625 possible vectors of scores, we compute the zero-one distance, absolute distance and squared difference to the vectors of scores $\mathbf{s}_{1}, \mathbf{s}_{2}$ and $\mathbf{s}_{3}$ provided by the trained panellists. For simplicity, we only show the computation for the vector $\mathbf{s}=(2,2,3,4)$ and summarize the results in Table 7.2. We see that, for this vector of scores, the sum of absolute distances is equal to five.

| $\mathbf{s}$ | $(2,2,3,4)$ | $d_{0}\left(\mathbf{s}, \mathbf{s}_{i}\right)$ | $d_{1}\left(\mathbf{s}, \mathbf{s}_{i}\right)$ | $d_{2}\left(\mathbf{s}, \mathbf{s}_{i}\right)$ |
| :---: | :---: | :---: | :---: | :---: |
| $\mathbf{s}_{1}$ | $(2,1,3,3)$ | 2 | 2 | 2 |
| $\mathbf{s}_{2}$ | $(2,2,5,4)$ | 1 | 2 | 4 |
| $\mathbf{s}_{3}$ | $(1,2,3,4)$ | 1 | 1 | 1 |

Table 7.2: The zero-one distance $d_{0}\left(\mathbf{s}, \mathbf{s}_{i}\right)$, absolute difference $d_{1}\left(\mathbf{s}, \mathbf{s}_{i}\right)$ and squared difference $d_{2}\left(\mathbf{s}, \mathbf{s}_{i}\right)$ between $\mathbf{s}=(2,2,3,4)$ and each vector of scores $\mathbf{s}_{1}, \mathbf{s}_{2}$ and $\mathbf{s}_{3}$ provided by the trained panellists.

For each of the 625 possible vectors of scores, we compute the sum of absolute distances and conclude that $\mathbf{s}^{*}=(2,2,3,4)$ is the vector of scores with the smallest sum of distances. As a result, we assign $\mathbf{s}^{*}=(2,2,3,4)$ as the consensus vector of scores of the four samples.

### 7.3. Obtaining consensus rankings

We consider the setting where $n_{U}$ untrained panellists each have compared all samples in the considered set $\mathscr{A}=\left\{a_{1}, \ldots, a_{n}\right\}$ of $n$ samples. The untrained panellists are asked to provide a complete ranking of the samples, however, they are allowed to express ties in case they consider two or more samples to be equally suitable. This results in $n_{U}$ rankings with ties, which we simply refer to as rankings. We denote by $\precsim_{i}$ the ranking with ties provided by the $i$-th panellist. Throughout this chapter, we refer to rankings with ties as rankings. Recall that a ranking of $n$ samples corresponds to information for $n(n-1) / 2$ pairs of samples. For instance, a ranking of three samples, say $a_{1} \prec a_{2} \sim a_{3}$, corresponds to information for three pairs $\left\{a_{1}, a_{2}\right\},\left\{a_{1}, a_{3}\right\}$ and $\left\{a_{2}, a_{3}\right\}$ and expresses that $a_{2}$ is ranked above $a_{1}$, that $a_{3}$ is ranked above $a_{1}$, and that $a_{2}$ and $a_{3}$ are tied. The fact that a sample is ranked above another sample means that the former is preferred over the latter.

The goal is to attain a consensus ranking of the samples in $\mathscr{A}$. We identify the method of Kemeny [134, which determines the ranking that minimizes the sum of the Kemeny distances to the rankings provided by the untrained panellists, as a particularly desirable approach. We recall that the Kemeny distance (Definition 6.5) between two rankings is computed as follows. Simply stated, for each pair of samples $\left\{a_{u}, a_{v}\right\}$, if both rankings agree on the order of the samples, we write down 0 ;
if in one ranking, $a_{u}$ is ranked above $a_{v}$ (or $a_{v}$ is ranked above $a_{u}$ ) and, in the other ranking, $a_{u}$ and $a_{v}$ are tied, we write down 1 ; and, if in one ranking, $a_{u}$ is ranked above $a_{v}$ and, in the other ranking, $a_{v}$ is ranked above $a_{u}$, we write down 2 . After writing down the numbers for all $n(n-1) / 2$ possible pairs, the Kemeny distance between the two rankings equals the sum of these numbers. Note that when the rankings contain no ties, the Kemeny distance is equal to the double of the Kendall distance [141]. We denote by $\widetilde{\mathscr{R}}$ the set of all possible rankings with ties. Consequently, for each possible ranking in $\widetilde{\mathscr{R}}$, we compute the ranking $\precsim^{*}$ that minimizes the sum of Kemeny distances to the rankings provided by the untrained panellists, as follows:

$$
\begin{equation*}
\precsim^{*}=\underset{\precsim \epsilon \widetilde{\mathscr{R}}}{\arg \min } \sum_{i=1}^{n_{U}} d_{K}\left(\precsim, \precsim_{i}\right), \tag{7.2}
\end{equation*}
$$

where $d_{K}\left(\precsim^{1}, \precsim^{2}\right)$ denotes the Kemeny distance between any two rankings $\precsim^{1}$ and $\precsim^{2} 3$. The minimizer $\precsim^{*}$, called the Kemeny median, is thus assigned as the consensus ranking (note that there can be multiple minimizers $\precsim^{*}$ ). To determine $\precsim^{*}$, the distance $d_{K}$ is computed for all possible rankings, where, for $n \leq 15$, the number of possible rankings can be estimated as the nearest integer to $n!/ 2(\log 2)^{n+1}$ [161.

Example 7.2. Consider a simple setting where five panellists each compare four samples $a_{1}, a_{2}, a_{3}$ and $a_{4}$, resulting in five rankings. The rankings are gathered in Table 7.3.

$$
\begin{array}{ll}
\precsim_{1} & a_{1} \sim a_{2} \prec a_{3} \sim a_{4} \\
\precsim_{2} & a_{1} \prec a_{2} \sim a_{3} \sim a_{4} \\
\precsim_{3} & a_{1} \sim a_{2} \prec a_{3} \prec a_{4} \\
\precsim_{4} & a_{1} \prec a_{2} \sim a_{3} \prec a_{4} \\
\precsim_{5} & a_{1} \sim a_{2} \prec a_{3} \sim a_{4} \\
\hline
\end{array}
$$

Table 7.3: The rankings of samples $a_{1}, a_{2}, a_{3}$ and $a_{4}$ expressed by the untrained panellists in Example 7.2 .

To determine the consensus ranking of these samples, we consider the problem defined by Eq. 7.2), and, for each of the 75 rankings, we compute the sum of the Kemeny distances to the rankings $\precsim_{1}, \precsim_{2}, \precsim_{3}, \precsim_{4}$ and $\precsim_{5}$ provided by the untrained panellists. For simplicity, we only show the computation for the ranking $\precsim=a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ and summarize the results in Table 7.4. We see that, for this ranking, the sum of the Kemeny distances is equal to seven.

[^24]| $\precsim$ | $a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ | $d_{K}\left(\precsim, \precsim_{i}\right)$ |
| :---: | :---: | :---: |
| $\precsim_{1}$ | $a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ | 0 |
| $\precsim_{2}$ | $a_{1} \prec a_{2} \sim a_{3} \sim a_{4}$ | 3 |
| $\precsim_{3}$ | $a_{1} \sim a_{2} \prec a_{3} \prec a_{4}$ | 1 |
| $\precsim_{4}$ | $a_{1} \prec a_{2} \sim a_{3} \prec a_{4}$ | 3 |
| $\precsim_{5}$ | $a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ | 0 |

Table 7.4: The Kemeny distance $d_{K}\left(\precsim, \precsim_{i}\right)$ between $\prec=a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ and each ranking $\precsim_{1}, \preceq_{2}, \preceq_{3}, \preceq_{4}$ and $\precsim_{5}$ provided by the untrained panellists.

After computing all the sums of Kemeny distances for each of the 75 rankings, we conclude that $\precsim^{*}=a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ is the ranking with the smallest sum of Kemeny distances. As a result, we assign $\precsim^{*}=a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ as the consensus ranking of the four samples.

### 7.4. Integrating scores and rankings

So far, we have described the median for scores and the Kemeny median for rankings. In this section, we develop two methods: the first to assign a consensus score to each sample by integrating rankings with scores, and the second to determine a consensus ranking of multiple samples by integrating scores with rankings.

### 7.4.1. Improving the quality of the assessment of a consensus vector of scores

We consider the setting where $n_{T}$ trained panellists each have assigned a score to each of the $n$ samples in $\mathscr{A}$ resulting in the vectors of scores $\mathbf{s}_{1}, \ldots, \mathbf{s}_{n_{T}}$. Due to limitations in the number of trained panellists, any additional source of information could potentially improve the quality of the assessment of a sample. Thus, assuming that the number of trained panellists is small, we consider that $n_{U}$ untrained panellists are each asked to rank the $n$ samples in $\mathscr{A}$ resulting in the rankings $\precsim_{1}, \ldots, \precsim_{n_{U}}$. Typically, the number of untrained panellists is quite large $\left(n_{U}>75\right)$ 1].

The goal is to agree on the consensus score that should be assigned to each sample in $\mathscr{A}$, while integrating rankings with scores. We propose to consider a combination of the median and Kemeny median associated with the vectors of scores provided by the trained panellists and the rankings provided by the untrained panellists. We denote by $\widetilde{\mathscr{R}}$ the set of all possible rankings of $n$ samples. To compute the
'distance $4_{4}^{4}$ between each possible vector of scores $\mathbf{s}$ and the rankings provided by the untrained panellists, we first define the set $\vartheta_{\text {s }}$ of all possible rankings $\precsim \in \widetilde{\mathscr{R}}$ that do not contradict the vector of scores $\mathbf{s}$ (i.e., the set of all rankings in which samples that are assigned larger scores in $s$ are ranked above samples that are assigned smaller scores in s), as follows:

$$
\begin{equation*}
\vartheta_{\mathbf{s}}=\left\{\precsim \in \widetilde{\mathscr{R}} \mid(\forall i, j \in\{1, \ldots, n\})\left(s_{i}<s_{j} \Rightarrow a_{i} \prec a_{j}\right)\right\} . \tag{7.3}
\end{equation*}
$$

We say that a vector of scores $\mathbf{s}$ is not contradicted by a ranking $\precsim$ if $\precsim \in \vartheta_{\mathbf{s}}$. Note that the set $\vartheta_{\mathbf{s}}$ is never empty. Therefore, we determine the ranking $\precsim^{*}$ among those in $\vartheta_{\mathbf{s}}$ that minimizes the sum of Kemeny distances to the rankings provided by the untrained panellists, as follows:

$$
\begin{equation*}
\precsim^{*}=\underset{\precsim \in \vartheta_{\mathrm{s}}}{\arg \min } \sum_{i=1}^{n_{U}} d_{K}\left(\precsim, \precsim_{i}\right) . \tag{7.4}
\end{equation*}
$$

Combining the sum of absolute distances to the vectors of scores provided by the trained panellists and the minimal sum of Kemeny distances to the rankings provided by the untrained panellists requires the definition of a cost function. Thus, we define $C_{\alpha}(\mathbf{s})$ as a convex combination of the distances associated with the vectors of scores provided by the trained panellists and the rankings provided by the untrained panellists, as follows:

$$
\begin{equation*}
C_{\alpha}(\mathbf{s})=\frac{\alpha}{B_{T}} \sum_{i=1}^{n_{T}} d_{1}\left(\mathbf{s}, \mathbf{s}_{i}\right)+\frac{(1-\alpha)}{B_{U}} \min _{\precsim \in \vartheta_{\mathbf{s}}} \sum_{i=1}^{n_{U}} d_{K}\left(\precsim, \precsim_{i}\right), \tag{7.5}
\end{equation*}
$$

where $B_{T}=n_{T} \cdot n \cdot(k-1)$ and $B_{U}=n_{U} \cdot n \cdot(n-1)$ are normalizing constants $5^{5}$, and $\alpha \in[0,1]$ is a weight that controls the influence of the ranking information. Larger values of $\alpha$ give more importance to the trained panellists, whereas smaller values of $\alpha$ give more importance to the untrained panellists. At one extreme, when $\alpha=1$, only the scores provided by the trained panellists are used to obtain the vector of scores that minimizes $C_{\alpha}(\mathbf{s})$ with no impact from the rankings provided by the untrained panellists. At the other extreme, when $\alpha=0$, only the rankings provided by the untrained panellists are used to assign the vector of scores that minimizes $C_{\alpha}(\mathbf{s})$ with no impact from the scores provided by the trained panellists.

[^25]We determine the vector of scores $\mathbf{s}_{\alpha}^{*}$ that minimizes Eq. 77.5), as follows:

$$
\begin{equation*}
\mathbf{s}_{\alpha}^{*}=\underset{\mathbf{s} \in\{1, \ldots, k\}^{n}}{\arg \min } C_{\alpha}(\mathbf{s}) \tag{7.6}
\end{equation*}
$$

Note that there can be multiple minimizers $\mathbf{s}_{\alpha}^{*}$ for the same $\alpha$. In decision making theory, specifically in the area of multiple criteria decision making, the optimization problem (7.6) is seen as solving multiple objective problems by turning them into a single objective problem [162, 163, 164].

Since $\alpha \in[0,1]$, it is impossible to compute $\mathbf{s}_{\alpha}^{*}$ for each $\alpha$. Therefore, bearing in mind that, for any fixed vector of scores $\mathbf{s}, f(\alpha):=C_{\alpha}(\mathbf{s})$ can be visualized as a (straight) line, we compare the lines of each possible pair of vectors of scores. For each possible pair of lines, we distinguish three cases: there are no points of intersection, there is exactly one point of intersection, or both lines coincide. These facts can then be used to analytically compute $\mathbf{s}^{*}$ as a function of $\alpha$.

Remark The optimization problems $(7.6$ and 7.10 can also be solved as multiobjective optimization problems [165, 166, where there is a trade-off between two (conflicting) objectives: (1) minimizing the distance associated with the vector of scores and (2) minimizing the distance associated with the rankings. Solving multi-objective optimization problems can allow researchers to find a representative set of (Pareto) optimal solutions, and quantify the trade-offs in satisfying the different objectives. However, solving a multi-objective optimization problem is not as straightforward as a single-objective optimization problem, as there may exist, nor non-trivial problems, a large number of Pareto optimal solutions. We refer to the texts of Multicriteria Optimization by Ehrgott 90 and Evolutionary Algorithms for Solving Multi-Objective Problems by Coello et al. [167.

The total number of possible vectors of scores is typically larger than the total number of possible rankings. Note that $k^{n}>n!/ 2(\log 2)^{n+1}$, for $k \geq n$ (i.e., when the number of points on a $k$-point scale is greater than or equal to the number of samples). Thus, there can be multiple vectors of scores that are not contradicted by the Kemeny median. Therefore, in the special case when $\alpha=0$ (i.e., only the rankings provided by the untrained panellists are used), there will always be multiple minimizers $\mathbf{s}_{0}^{*}$. As will be discussed in the following example, this results in the minimizers forming a fan-shaped pattern starting at $\alpha=0$.

Example 7.3. Consider the scores expressed by trained panellists in Example 7.1 and the rankings expressed by untrained panellists in Example 7.2 for the samples $a_{1}, a_{2}, a_{3}$ and $a_{4}$.

To determine the consensus score that should be assigned to each of these samples, we consider the problem defined by Eq. (7.6) by computing $C_{\alpha}(\mathbf{s})$ for each of the 625 vectors of scores. For simplicity, we show the computation for the vector of scores $\mathbf{s}=(2,2,3,4)$, which was identified as the consensus vector of scores in

Example 7.1. with $\sum_{i=1}^{5} d_{1}\left(\mathbf{s}, \mathbf{s}_{i}\right)=5$. The sum of absolute distances associated with the vectors of scores is upper bounded by $B_{T}=3 \cdot 4 \cdot 4=48$.

To compute the minimal sum of Kemeny distances $\sum_{i=1}^{5} d_{K}\left(\precsim, \precsim_{i}\right)$, we first determine the set $\vartheta_{\mathbf{s}}$ of all possible rankings that do not contradict $\mathbf{s}$. Since the score of $a_{4}$ is the largest, $a_{4}$ is ranked above the other samples. Similarly, $a_{3}$ is ranked above $a_{1}$ and $a_{2}$. Since the scores of $a_{1}$ and $a_{2}$ are equal, any of the following cases applies: $a_{1}$ is ranked above $a_{2}, a_{1}$ and $a_{2}$ are tied, and $a_{2}$ is ranked above $a_{1}$, as follows:

$$
\vartheta_{(2,2,3,4)}=\left\{\begin{array}{c}
a_{1} \prec a_{2} \prec a_{3} \prec a_{4}, \\
a_{1} \sim a_{2} \prec a_{3} \prec a_{4}, \\
a_{2} \prec a_{1} \prec a_{3} \prec a_{4}
\end{array}\right\} .
$$

We compute the sum of the Kemeny distances between each $\precsim \in \vartheta_{\mathbf{s}}$ and the rankings provided by the untrained panellists. The results are summarized in Table 7.5. The sum of Kemeny distances associated with the rankings is upper bounded by $B_{U}=5 \cdot 4 \cdot 3=60$. Therefore, $\precsim=a_{1} \sim a_{2} \prec a_{3} \prec a_{4}$ is the ranking that minimizes the sum of Kemeny distances, with $\sum_{i=1}^{5} d_{K}\left(\precsim, \precsim_{i}\right)=8$.

| $\precsim$ | $\sum_{i=1}^{5} \tilde{d}_{K}\left(\precsim, \precsim_{i}\right)$ |
| :---: | :---: |
| $a_{1} \prec a_{2} \prec a_{3} \prec a_{4}$ | 9 |
| $a_{1} \sim a_{2} \prec a_{3} \prec a_{4}$ | $\mathbf{8}$ |
| $a_{2} \prec a_{1} \prec a_{3} \prec a_{4}$ | 13 |

Table 7.5: Sum of Kemeny distances between each ranking $\prec$ that does not contradict $\mathbf{s}=(2,2,3,4)$ and the rankings provided by the untrained panellists.

It follows that

$$
C_{\alpha}(\mathbf{s})=\frac{5}{45} \alpha+\frac{8}{60}(1-\alpha)=\frac{8}{60}-\frac{4}{180} \alpha
$$

After computing $C_{\alpha}(\mathbf{s})$ for each of the 625 possible vectors of scores $\mathbf{s}$, we illustrate in Figure 7.1 all the $\mathbf{s}_{\alpha}^{*}$ that minimize $C_{\alpha}(\mathbf{s})$ for at least one value of $\alpha \in[0,1]$. As discussed above, for $\alpha=0$, there will always be multiple minimizers $\mathbf{s}_{0}^{*}$ associated with all vectors of scores that are not contradicted by the Kemeny median. Since we know from Example 7.2 that $\precsim^{*}=a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ is the Kemeny median, we illustrate (in black) all the vectors of scores $\mathbf{s}_{0}^{*}$ that are not contradicted by this $\precsim^{*}$ in Figure 7.1. These vectors of scores form a fan-shaped pattern starting at $\alpha=0$ since, at the left end, they all result in the same value $C_{0}(\mathbf{s})$, and, at the right end, they result in (mostly) different values $C_{1}(\mathbf{s})$. In addition, we highlight only the vectors of scores that are minimizers for $\alpha \in] 0,1]$.

The obtained minimizers $\mathbf{s}_{\alpha}^{*}$ are summarized as follows:


Figure 7.1: Illustrating the vectors of scores $\mathbf{s}_{\alpha}^{*}$ that minimize $C_{\alpha}(\mathbf{s})$ for $\alpha \in[0,1]$ in Example 7.3

$$
\mathbf{s}_{\alpha}^{*}=\left\{\begin{array}{ccrr}
\{(1,1,2,2),(1,1,3,3),(1,1,4,4),(1,1,5,5), & & \\
(2,2,3,3),(2,2,4,4),(2,2,5,5),(3,3,4,4), & , \text { if } & \alpha=0, \\
(3,3,5,5),(4,4,5,5),(1,1,1,1),(2,2,2,2), & & \\
(3,3,3,3),(4,4,4,4),(5,5,5,5)\} & , \text { if } & 0<\alpha<\frac{4}{9}, \\
\{(2,2,3,3),(2,2,4,4)\} & , \text { if } & \alpha=\frac{4}{9}, \\
\{(2,2,3,3),(2,2,3,4),(2,2,4,4)\} & , \text { if } & \alpha>\frac{4}{9} .
\end{array}\right.
$$

Since we do not intend to rely solely on the rankings provided by the untrained panellists, we ignore the minimizers for $\alpha=0$. We conclude that the consensus vector of scores is $(2,2,3,4)$ for $\alpha>\frac{4}{9},(2,2,3,3)$ and $(2,2,4,4)$ for $0<\alpha<\frac{4}{9}$, and all the three vectors of scores for $\alpha=\frac{4}{9}$. We visualize these vectors of scores in Figure 7.2.

We deduce on the basis of the vectors of scores provided by the trained panellists that $a_{4}$ is of higher quality than $a_{3}$, since the former is assigned a higher score than the latter in the consensus vector of scores (see $\alpha>\frac{4}{9}$ ). However, integrating the rankings provided by the untrained panellists hints that these samples are similar in terms of quality, since they were both assigned the same score in the consensus vectors of scores (see $\alpha<\frac{4}{9}$ ). We conjecture that the low number of trained panellists might have limited the quality of the initial assessment of the samples.


Figure 7.2: A visualization of the scores assigned to each sample $a_{1}, a_{2}, a_{3}$ and $a_{4}$ in Example 7.3 for (a) $\alpha>\frac{4}{9}$ and for (b) and (c) $0<\alpha<\frac{4}{9}$.

Remark It is clear that the choice of $\alpha$ is difficult. To explain a potential rationale behind the choice of $\alpha$, we consider the setting where a small number of trained panellists provide scores and a large number of untrained panellists provide rankings. Thus, we can rely more on the rankings, however, not completely since rankings only provide relative information. Therefore, we just consider the minimizers $\mathbf{s}_{\alpha}^{*}$ for small values of $\alpha$ close to, but not equal to zero.

For instance, in Example 7.3, if we consider the setting where the above results are obtained given a very large number of untrained panellists, which is a typical setting where consumers are usually considered as untrained panellists, then we would conclude that the consensus vectors of scores are $(2,2,3,3)$ and $(2,2,4,4)$ since they are the solution for any $\alpha \in] 0, \frac{4}{9}[$.

### 7.4.2. Improving the quality of the assessment of a consensus ranking

Typically, quite a large number of untrained panellists $\left(n_{U}>75\right)$ is required [1]. We now consider the setting where the number of untrained panellists is limited, and any additional source of information could potentially improve the quality of the assessment of the samples. Thus, assuming that the number of untrained
panellists is not as large as required, we consider that $n_{T}$ trained panellists are each asked to assign a score to each of the $n$ samples in $\mathscr{A}$ resulting in the vecotrs of scores $\mathbf{s}_{1}, \ldots, \mathbf{s}_{n_{T}}$.

The goal is to agree on the consensus ranking on the samples in $\mathscr{A}$, while integrating scores with rankings. We propose a similar approach to the aforementioned method for improving the quality of the assessment of a consensus vector of scores. To compute the 'distance' between each possible ranking $\precsim$ and the vectors of scores provided by the trained panellists, we first define the set $\varphi \precsim$ of all possible vectors of scores that do not contradict the ranking $\precsim$ (i.e., the set of all vectors of scores in which samples that are ranked similar to or above other samples are assigned larger or same scores in s), as follows:

$$
\begin{equation*}
\varphi_{\precsim}=\left\{\mathbf{s} \in\{1, \ldots, k\}^{n} \mid(\forall i, j \in\{1, \ldots, n\})\left(a_{i} \precsim a_{j} \Rightarrow s_{i} \leq s_{j}\right)\right\} . \tag{7.7}
\end{equation*}
$$

We say that a ranking $\precsim$ is not contradicted by a vector of scores $\mathbf{s}$ if $\mathbf{s} \in \varphi_{\precsim}$. Note that the set $\varphi_{\precsim}$ is never empty. Therefore, we determine the vector of scores $\mathbf{s}^{*}$ among those in $\varphi_{\precsim}$ that minimizes the sum of absolute distances to the vectors of scores provided by the trained panellists, as follows:

$$
\begin{equation*}
\mathbf{s}^{*}=\underset{\mathbf{s} \in \varphi_{\precsim}}{\arg \min } \sum_{i=1}^{n_{T}} d_{1}\left(\mathbf{s}, \mathbf{s}_{i}\right) . \tag{7.8}
\end{equation*}
$$

Combining the minimal sum of absolute distances to the vectors of scores provided by the trained panellists and the sum of Kemeny distances to the rankings provided by the untrained panellists requires the definition of a cost function. Thus, we define $D_{\alpha}(\precsim)$ as a convex combination of the distances associated with the vectors of scores provided by the trained panellists and the rankings provided by the untrained panellists, as follows:

$$
\begin{equation*}
D_{\alpha}(\precsim)=\frac{\alpha}{B_{T}} \min _{\mathbf{s} \in \varphi_{\precsim}} \sum_{i=1}^{n_{T}} d_{1}\left(\mathbf{s}, \mathbf{s}_{i}\right)+\frac{(1-\alpha)}{B_{U}} \sum_{i=1}^{n_{U}} d_{K}(\precsim, \precsim i) . \tag{7.9}
\end{equation*}
$$

Finally, we determine the ranking $\precsim_{\alpha}^{*}$ that minimizes Eq. $\sqrt{7.9}$, as follows:

$$
\begin{equation*}
\precsim_{\alpha}^{*}=\underset{\precsim \in \widetilde{\mathscr{R}}}{\arg \min } D_{\alpha}(\precsim) . \tag{7.10}
\end{equation*}
$$

Note that there can be multiple minimizers $\precsim_{\alpha}^{*}$ for the same $\alpha$.
Similar to the aforementioned approach, since $\alpha \in[0,1]$, it is impossible to compute $\precsim_{\alpha}^{*}$ for each $\alpha$. Therefore, bearing in mind that, for any fixed ranking $\precsim$, $g(\alpha):=D_{\alpha}(\precsim)$ can be visualized as a (straight) line, we compare the lines of each
possible pair of rankings. For each possible pair of lines, we distinguish three cases: there are no points of intersection, there is exactly one point of intersection, or both lines coincide. These facts can then be used to analytically compute $\precsim^{*}$ as a function of $\alpha$.

In addition, as the total number of possible vectors of scores is typically larger than the total number of possible rankings, there can be multiple rankings that are not contradicted by the median. Therefore, in the special case when $\alpha=1$ (i.e., only the vectors of scores by the trained panellists are used), there will always be multiple minimizers $\precsim_{1}^{*}$. As will be discussed in the following example, this results in the minimizers forming a fan-shaped pattern starting at $\alpha=1$.

Example 7.4. We continue with the data from Example 7.3. To determine the consensus ranking of the samples, we consider the problem defined by Eq. 7.10) by computing $D_{\alpha}(\precsim)$ for each of the 75 rankings. For simplicity, we show the computation for the ranking $\precsim^{*}=a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$, which was identified as the consensus ranking in Example 7.2, with $\sum_{i=1}^{5} d_{K}\left(\precsim, \precsim_{i}\right)=7$. The sum of Kemeny distances associated with the rankings is upper bounded by $B_{U}=5 \cdot 4 \cdot 3=60$.

Now, we consider the set $\varphi_{\precsim}$ of all possible vectors of scores that do not contradict $\precsim$, as follows:

$$
\varphi_{a_{1} \sim a_{2} \prec a_{3} \sim a_{4}}=\left\{\mathbf{s} \in\{1, \ldots, 5\}^{4} \mid\left(s_{1}=s_{2} \leq s_{3}=s_{4}\right)\right\} .
$$

We compute the sum of the absolute distances between each $\mathbf{s} \in \varphi_{\precsim}$ and the vectors of scores provided by the trained panellists. The results are summarized in Table 7.6. The sum of absolute distances associated with the vectors of scores is upper bounded by $B_{T}=3 \cdot 4 \cdot 4=48$. Therefore, $\mathbf{s}=\{(2,2,3,3),(2,2,4,4)\}$ are the vectors of scores that minimize the sum of absolute distances, with $\sum_{i=1}^{5} d_{1}\left(\mathbf{s}, \mathbf{s}_{i}\right)=6$. Finally, we obtain:

$$
D_{\alpha}(\precsim)=\frac{\alpha}{120}+\frac{7}{60} .
$$

After computing $D_{\alpha}(\precsim)$ for each of the 75 possible rankings $\precsim$, we illustrate in Figure 7.3 all the $\precsim_{\alpha}^{*}$ that minimize $D_{\alpha}(\precsim)$ for at least one value of $\alpha \in[0,1]$. As discussed above, for $\alpha=1$, there will always be multiple minimizers $\precsim_{1}^{*}$ associated with all rankings that are not contradicted by the median. These vectors of scores form a fan-shaped pattern starting at $\alpha=1$ since, at the right end, they all result in the same value $D_{1}(\precsim)$, and, at the left end, they result in (mostly) different values $D_{0}(\precsim)$.

| $\mathbf{s}$ | $\sum_{i=1}^{5} d_{1}\left(\mathbf{s}, \mathbf{s}_{i}\right)$ | $\mathbf{s}$ | $\sum_{i=1}^{5} d_{1}\left(\mathbf{s}, \mathbf{s}_{i}\right)$ |
| :---: | :---: | :---: | :---: |
| $(1,1,1,1)$ | 20 | $(2,2,5,5)$ | 10 |
| $(1,1,2,2)$ | 14 | $(3,3,3,3)$ | 12 |
| $(1,1,3,3)$ | 8 | $(3,3,4,4)$ | 12 |
| $(1,1,4,4)$ | 8 | $(3,3,5,5)$ | 16 |
| $(1,1,5,5)$ | 12 | $(4,4,4,4)$ | 18 |
| $(2,2,2,2)$ | 12 | $(4,4,5,5)$ | 22 |
| $(2,2,3,3)$ | 6 | $(5,5,5,5)$ | 28 |
| $(2,2,4,4)$ | $\mathbf{6}$ |  |  |

Table 7.6: Sum of absolute distances between each vector of scores $\mathbf{s}$ that does not contradict $\precsim=a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ and the vectors of scores provided by the trained panellists.

The obtained minimizers $\precsim_{\alpha}^{*}$ are summarized as follows:

$$
\precsim_{\alpha}^{*}=\left\{\begin{array}{rlrl}
\left\{a_{1} \sim a_{2} \prec a_{3} \sim a_{4}\right\} \\
\left\{\begin{aligned}
& a_{1} \sim a_{2} \prec a_{3} \sim a_{4} \\
& a_{1} \sim a_{2} \\
& \prec a_{3} \prec a_{4}
\end{aligned}\right\} & , \text { if } & \alpha<\frac{4}{9}, \\
\left\{a_{1} \sim a_{2} \prec a_{3} \prec a_{4}\right\} \\
\text {,if } & , \text { if } & \frac{4}{9}<\alpha<\frac{4}{9} \\
\left\{\begin{array}{l}
a_{1}
\end{array} \sim a_{2} \prec a_{3} \prec a_{4}\right. \\
a_{1} \prec a_{2} \prec a_{3} \prec a_{4} \\
a_{2} \prec a_{1} \prec a_{3} \prec a_{4}
\end{array}\right\}, \text { if } \quad \alpha=1
$$

Since we do not intend to rely solely on the scores provided by the trained panellists, we ignore the minimizers for $\alpha=1$. We conclude that the consensus ranking is $a_{1} \sim a_{2} \prec a_{3} \sim a_{4}$ for $\alpha<\frac{4}{9}, a_{1} \sim a_{2} \prec a_{3} \prec a_{4}$ for $\frac{4}{9}<\alpha<1$, and both these rankings for $\alpha=\frac{4}{9}$.

We deduce on the basis of the rankings provided by the untrained panellists that $a_{4}$ is similar to $a_{3}\left(\right.$ see $\left.\alpha<\frac{4}{9}\right)$. However, integrating the scores provided by the trained panellists hints that sample $a_{4}$ might be ranked above sample $a_{3}$ (see $\frac{4}{9}<\alpha<1$ ). We conjecture that the low number of untrained panellists might have limited the quality of the initial assessment of the samples.

### 7.5. Information about samples

Typically, gathering together several panellists is a challenging and expensive exercise and providing a score is a task that can be fulfilled quickly. Therefore, it


Figure 7.3: Illustrating the rankings $\precsim_{\alpha}^{*}$ that minimize $D_{\alpha}(\precsim)$ for $\alpha \in[0,1]$ in Example 7.4
is common to provide the panellists with multiple food samples during the same experiment. In general, the scores assigned to each sample are considered to be independent, and the assessment of a consensus score to each sample is assumed to be an independent task. Note that it is often difficult to gather the same number of panellists for different experiments.

Consider the problem setting where several panellists provide scores for each of the $n$ food samples. These scores are not necessarily gathered in any particular order, however, for simplicity reasons, they have been represented in an increasing order ${ }^{6}$. For example, consider a simple setting where nine panellists each assign a score to a given food sample on the 5 -point scale fixed in Fig. 4.2. The scores provided by the panellists are $4,1,2,1,4,3,3,4,3$ and are represented in increasing order as $1,1,1,2,3,3,3,4,4,4$. Denote by $m_{j}$ the number of scores assigned to the $j$-th sample and by $\mathbf{s}_{i}(j)$ the $i$-th lowest score assigned by a panellist to sample $a_{j}$, where $j \in\{1, \ldots, n\}$ and $i \in\left\{1, \ldots, m_{j}\right\}$.

The goal is to agree on the consensus score that should be assigned to each sample in $\mathscr{A}$. Obtaining the median score for each sample is equivalent to directly computing the vector $\mathbf{s}^{*}$, as follows:

[^26]\[

$$
\begin{equation*}
\mathbf{s}^{*}=\underset{\mathbf{s} \in\{1, \ldots, k\}^{n}}{\arg \min } \sum_{j=1}^{n} \sum_{i=1}^{m_{j}} \partial\left(\mathbf{s}(j), \mathbf{s}_{i}(j)\right), \tag{7.11}
\end{equation*}
$$

\]

In case $\partial=\partial_{0}$, we refer to $\mathbf{s}^{*}(j)$ as the mode of the $j$-th sample; and in case $\partial=\partial_{1}$, we refer to $\mathbf{s}^{*}(j)$ as the median of the $j$-th sample. Note that to compute the mean of the $j$-th sample, we consider $\partial=\partial_{2}$ and compute the root of the sum of distances. Note that, in case $m_{j}$ is odd for all $j$, the median (and, thus, the minimizer of Eq. (7.11) for $\partial=\partial_{1}$ ) is unique. In case $m_{j}$ is even for at least one $j$, there can be multiple medians (and, thus, multiple minimizers of Eq. 7.11) for $\partial=\partial_{1}$ ).

One could note that some relations between the scores of the different samples could also be known. We provide a non-exhaustive list of some potential real-life situations here after.

### 7.5.1. Knowledge of storage days

Researchers are often interested in studying the evolution of attributes of perishable food. This is typically done by asking panellists to provide a score to food samples that come from different time spans of the shelf life of the same food product. A common method used in this situation is the time-intensity (TI) method [1, Chapter 8]. In general, it is expected that food samples should be less fresh as time goes by. Thus, it is expected that the less fresh the sample is, the lower the score should be. One example is evaluating the freshness or tenderness of meats, where the score may only decrease with time [168].

Consider the setting where samples coming from the same food product are indexed in increasing order of storage days. Thus, the potential consensus scores should naturally reduce to those that satisfy the following constraints:

$$
\mathbf{s}(1) \geq \cdots \geq \mathbf{s}(n)
$$

Note that the overall trend of the scores should be decreasing, however, different decreasing patterns are possible. One possible pattern of scores is illustrated in Figure 7.4. Note that multiple consecutive samples could be assigned the same score, as illustrated in Figure 7.4 for points 0,1 and 2. Typically, this occurs when the number $k$ of points on the scale is small.

In studies on acceptability of beverages, the evolution of certain attributes of beverages is of interest. Typically, beverage samples are studied at different time spans of the air exposure of the same beverage. In general, it is expected that beverage samples should have an increasing acceptance at first, eventually


Figure 7.4: Example of scores describing freshness of food that decreases over time.
decreasing afterwards. Typical examples include the evaluation of the astringency and flavour of beer and wine, where these attributes increase in intensity at first and eventually decrease with time [169, 170.

Consider the setting where samples coming from the same beverage are indexed in increasing order of time spans of air exposure. Thus, the potential consensus scores should naturally reduce to those that satisfy the following constraints:

$$
\mathbf{s}(1) \leq \cdots \leq \mathbf{s}(a) \text { and } \mathbf{s}(a) \geq \cdots \geq \mathbf{s}(n), \text { for } a \in\{1, \ldots, n\}
$$

This means that there should be a unimodal pattern. One possible pattern of scores is illustrated in Figure 7.5. Note that if one considers a short duration, say $t \in\{0, \ldots, 3\}$, then the overall trend is only increasing. Similarly, if one considers a long duration, however, at a later time, say $t \in\{2, \ldots, 8\}$, then the overall trend is only decreasing.

### 7.5.2. Results of clustering analysis

In many studies, food samples are stored at different (temperature and atmospheric) conditions or represent the same food product but originate from a different initial batch, manufacturer or season. In addition, the initial contamination of the samples (i.e., initial microbial load) and the similarity of the samples, in terms of dimensions and composition, play a big role in the spoilage rate of every sample, and, thus, the decreasing pattern of the scores may not always hold. Thus, the storage days could not be used as the only tool to compare these samples. For instance, it is not always the case that samples that have been stored at different conditions for the same duration of time will be similar. Similarly, it is not always the case


Figure 7.5: Example of scores describing the intensity of wine flavour that is unimodal.
that a sample is always preferred over another sample that is stored at different conditions and has been stored for longer.

It is well established that microbial growth is the most important cause of food spoilage [22, producing volatile organic compounds (VOCs) and, subsequently, off-odours and off-flavours. These odours and flavours result in an olfactory impact that is associated with the spoilage of food. Therefore, the relation between the VOC profiles and the quality of food has caught the attention of many researchers in food science. Recently, the composition of the VOC profiles has been successfully used to evaluate the quality of food, such as seafood [61, 65] and meat 66.

To establish a relation between the VOC profiles of the samples and their resulting consensus scores, clustering analysis, a method for merging similar groups of samples based on the similarity of their VOC profile, can be used. Hierarchical agglomerative clustering [67] is a commonly used clustering analysis tool that has been recently used by researchers in food science 38 ].

In general, it is expected that samples clustered together should be quite similar, and, thus, their scores should not be very different. Therefore, the absolute difference of the scores of these samples should not exceed a certain threshold. Note that we prefer not to impose that samples in the same cluster should have strictly the same scores because this might be too restrictive. However, this is still a possibility, as will be further explained below.

The considered setting may naturally reduce the potential consensus scores to those that satisfy the following constraints:

$$
|\mathbf{s}(i)-\mathbf{s}(j)| \leq \epsilon, \text { for } i, j \in I_{b},
$$

where $I_{b}$ are the indices corresponding to the $b$-th cluster and $\epsilon$ is a threshold on
the absolute difference of the scores of samples in the same cluster. Note that the value of $\epsilon$ may depend on the number of points $k$ on the scale used for scoring. The special case where $\epsilon=0$ amounts to restricting with equality constraints only. For instance, consider the case where samples $\left\{a_{1}, a_{2}\right\}$ are found in one cluster and samples $\left\{a_{3}, a_{4}, a_{5}\right\}$ are found in another cluster. It is expected that the absolute difference of the scores of every pair of samples in each cluster should be less than or equal to one. This process is illustrated in Figure 7.6. It can be seen that the absolute difference of scores for each couple of samples in the same cluster is less than or equal to one, thus, satisfying the constraints.


Figure 7.6: Example of a dendrogram, where samples $\left\{a_{1}, a_{2}\right\}$ are in the first (in blue) cluster and samples $\left\{a_{3}, a_{4}, a_{5}\right\}$ are in the second (in orange) cluster. The scores are consistent with the clusters, equivalently, satisfying the constraints.

### 7.5.3. Other sensory evaluation tests

## Ranking test

Recently, researchers have been adopting scoring methods for determining the quality of food along with ranking methods to determine the order of the samples according to their quality [52, 155, 156, 157. Ranking tests involve several panellists providing rankings (with ties) on samples. Typically, these rankings are aggregated to obtain a consensus ranking that describes an underlying order of the samples, thus, it is expected that the scores agree with this consensus ranking of the samples. For example, rankings have been previously used to study the desirability of
different meats [171, 172. In this setting, these rankings can be useful information as a reference for relative desirability of meats.

In general, it is expected that samples ranked higher are preferred over samples ranked lower, thus, it is expected that the higher the sample is ranked the greater the score should be. Note that we do not impose that a sample ranked higher than another sample should have a strictly greater score, instead, we allow their scores to be equal as well. This is due to the fact that the considered scale is typically not rich enough for allowing to distinguish similar samples. In case two samples are tied, it is expected that their scores should be similar ${ }^{77}$. The considered setting may naturally reduce the potential consensus scores to those that satisfy the following constraints:

$$
a_{i} \precsim a_{j} \Rightarrow \mathbf{s}(i) \leq \mathbf{s}(j), \text { for } i, j \in\{1, \ldots, n\}
$$

For instance, consider the ranking $a_{1} \prec a_{2} \sim a_{3} \prec a_{4}$. It is expected that $a_{1}$ should be assigned a score smaller than or equal to that of sample $a_{2}$, which should be assigned a score equal to that of sample $a_{3}$, which should be assigned a score smaller than or equal to that of sample $a_{4}$. More formally, the resulting constraints are $\mathbf{s}(1) \leq \mathbf{s}(2)=\mathbf{s}(3) \leq \mathbf{s}(4)$. This process is illustrated in Figure 7.7.


Figure 7.7: Example of scores describing the ranking $a_{1} \prec a_{2} \sim a_{3} \prec a_{4}$.

## Discrimination test

Many discrimination tests can be seen as a special case of a ranking test. For

[^27]instance, in an A-notA test, panellists are provided with one sample and asked whether or not it is similar to a reference sample A [173, Chapter 4]. Based on the responses of the panellists, if there is no significant difference between the samples, then it is expected that they should be assigned a similar score. Therefore, the absolute difference of the scores of these samples should not exceed a certain threshold $\epsilon$. This process is illustrated in Figure 7.8.


Figure 7.8: Example of scores describing that there is no significant difference between sample $a_{2}$ and the reference sample $a_{1}$, and where a threshold $\epsilon=1$ is considered.

Another instance is a duo-trio test, where panellists are provided with two samples and a reference sample that is identical to one of the two samples and are asked to match one of the two samples to the reference sample [173, Chapter 4]. It is expected that the reference sample and the sample identical to it should be scored equally. Moreover, if a large number of panellists are not able to distinguish the identical samples, then it is expected that the third sample should be assigned a similar score to that of the identical samples. Therefore, the absolute difference of the scores of the non-identical samples should not exceed a certain threshold $\epsilon$. This process is illustrated in Figure 7.9

Another instance is a two-out-of-five test, where panellists are given five samples and are asked to distinguish two identical samples from the other three samples [173, Chapter 4]. It is expected that the reference sample and the sample identical to it should be scored equally. Moreover, if a large number of panellists are not able to distinguish the identical samples from the other three, then it is expected that there is no significant difference among the five samples and that all the samples should be assigned similar scores. Therefore, the absolute difference of the scores



Figure 7.9: Example of scores describing that there is no significant difference between $a_{3}$ and the reference sample $a_{1}$ (equivalently, its identical sample $a_{2}$ ), and where a threshold $\epsilon=1$ is considered.
of these samples should not exceed a certain threshold $\epsilon$. For instance, if a large number of panellists matched one sample with another sample, then it is expected that these samples should be assigned equal scores. This process is illustrated in Figure 7.10 .

## Threshold test

In threshold tests, panellists are asked to determine a threshold of noticing a certain stimulus [1, Chapter 6]. Different versions of the threshold test have been proposed, the differential threshold test and the absolute threshold test being the most prominent examples. In the former, the aim is to determine the threshold at which an increase in a noticed stimulus can be perceived, whereas in the latter, the aim is to determine the lowest threshold at which a stimulus can be noticed. Note that the case where there is a decrease in stimulus can also be considered. One example is determining the (consumer) rejection of chocolate bitterness [174].

In differential threshold tests, it is expected that the sample where an increase in stimulus is not noticed should have a quite similar score to the previous sample that has one increment less of the stimulus, and, thus, their scores should not be very different. Therefore, the absolute difference of the scores of these samples


Figure 7.10: Example of scores describing that samples $a_{1}$ and $a_{2}$ are not distinguished from the other samples, and a threshold $\epsilon=1$ is considered.
should not exceed a certain threshold $\epsilon$. However, it is expected that the samples where an increase in stimulus is noticed should have a score greater than or equal to the score of the previous sample. Therefore, the absolute difference of the scores of these samples should exceed this threshold.

Note that the scores should be either increasing $\mathbf{s}(i) \leq \mathbf{s}(i+1)$ or decreasing $\mathbf{s}(i) \geq$ $\mathbf{s}(i+1)$ for $i \in\{1, \ldots, n\}$. The considered setting may naturally reduce the potential consensus scores to those that satisfy the following additional constraints:

$$
\begin{aligned}
& |\mathbf{s}(i)-\mathbf{s}(i+1)| \leq \epsilon, \quad \text { for } i \in \kappa_{1}, \\
& |\mathbf{s}(i)-\mathbf{s}(i+1)| \geq \epsilon, \quad \text { for } i \in \kappa_{2},
\end{aligned}
$$

where $\kappa_{1}$ are the indices corresponding to the samples where a stimulus is not noticed, $\kappa_{2}$ are the indices corresponding to the samples where a stimulus is noticed, and $\epsilon$ is a threshold on the absolute difference of the scores of consecutive samples. For instance, consider that the stimulus is first noticed at sample $a_{4}$. It is expected that the absolute differences of the scores of consecutive samples $a_{1}$ and $a_{2}$, and samples $a_{2}$ and $a_{3}$ should be smaller than or equal to $\epsilon=1$ and that the absolute difference of the scores of samples $a_{3}$ and $a_{4}$ should be greater than or equal to $\epsilon=1$. Similarly, given that the stimulus is noticed a second time at sample $a_{6}$, it is expected that the absolute difference of the scores of samples $a_{4}$ and $a_{5}$ should be smaller than or equal to one and that the absolute difference of the scores of samples $a_{5}$ and $a_{6}$ should be greater than or equal to one. This process is illustrated in Figure 7.11.


Figure 7.11: Example of scores in a differential detection test describing no detection of a stimulus (in green) and the detection of a stimulus (in orange) at samples $a_{4}$ and $a_{6}$, and where a threshold $\epsilon=1$ is considered.

In absolute threshold tests, samples with a stimulus are only compared to a reference sample. Thus, if an increase in stimulus is not noticed in a sample, then it is expected that this sample should be assigned a score similar to that of the reference sample. Therefore, the absolute difference of their scores should not exceed a certain threshold $\epsilon$. However, it is expected that the samples where an increase in stimulus is noticed should be assigned a different score than that assigned to the reference sample. Therefore, the absolute difference of their scores should exceed this threshold.

Note that the scores should be either increasing $\mathbf{s}(i) \leq \mathbf{s}(i+1)$ or decreasing $\mathbf{s}(i) \geq \mathbf{s}(i+1)$ for $i \in\{1, \ldots, n\}$. We consider the first sample $(i=1)$ to be the reference sample. The considered setting may naturally reduce the potential consensus scores to those that satisfy the following additional constraints:

$$
\begin{aligned}
& |\mathbf{s}(1)-\mathbf{s}(i)| \leq \epsilon, \quad \text { for } i \in\{1, \ldots, c-1\}, \\
& |\mathbf{s}(1)-\mathbf{s}(i)| \geq \epsilon, \quad \text { for } i \in\{c, \ldots, n\}
\end{aligned}
$$

where $c$ is the sample at which an increase in stimulus is noticed and $\epsilon$ is a threshold on the absolute difference of the scores of each sample with the reference sample.

For instance, consider that a stimulus is first noticed at sample $a_{4}$. It is expected that the absolute difference of the scores of samples $a_{1}$ and $a_{4}$ should be greater
than or equal to one. Now, consider that sample $a_{4}$, where the first stimulus is noticed, is the new reference and that a second stimulus is noticed at sample $a_{6}$. It is expected that the absolute difference of the scores of samples $a_{4}$ and $a_{6}$ should be greater than or equal to one. This process is illustrated in Figure 7.12,



Figure 7.12: Example of scores that are increasing (dashed line) in an absolute detection test describing no detection of a stimulus (in green) and the detection of a first stimulus (in orange) at sample $a_{4}$ when compared to the reference sample $a_{1}$ and a second stimulus (in orange) at sample $a_{6}$ when compared to the new reference sample $a_{4}$, and where a threshold $\epsilon=1$ is considered.

### 7.5.4. The constrained mode, median and mean

As we have previously discussed, the considered settings may naturally reduce the set of potential consensus scores from $\{1, \ldots, 5\}^{n}$ to a non-empty subset $\mathcal{S} \subseteq\{1, \ldots, 5\}^{n}$. We conjecture that in most real-life situations it seems natural for $\mathcal{S}$ to be defined as the conjunction of some (in)equality constraints on the components of $\mathbf{s}$. However, this condition should not be a requirement if it does not comply with the characteristics of the considered problem.

Thus, the consensus scores should be the ones given by the vector that minimizes the sum of distances while satisfying the constraints of $\mathcal{S}$, as follows:

$$
\begin{equation*}
\mathbf{s}^{*}=\underset{\mathbf{s} \in \mathcal{S}}{\arg \min } \sum_{j=1}^{n} \sum_{i=1}^{m_{j}} \partial\left(\mathbf{s}(j), \mathbf{s}_{i}(j)\right) \tag{7.12}
\end{equation*}
$$

In case $\partial=\partial_{0}$, we refer to $\mathrm{s}^{*}$ as a constrained mode; and in case $\partial=\partial_{1}$, we refer to $\mathbf{s}^{*}$ as a constrained median. In case $\partial=\partial_{2}$, we compute the root of the sum of distances $\partial_{2}$ and refer to the minimizer as a constrained mean. These concepts are illustrated in the following example.

Example 7.5. Consider a simple setting where nine panellists each assign a score to two given food samples on the 5-point scale fixed in Figure 4.2. The scores assigned to each sample are represented in increasing value in Table 7.7.

| The $i$-th lowest score | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Score of sample $a_{1}\left(\mathbf{s}_{i}(1)\right)$ | 1 | 1 | 2 | 3 | 3 | 3 | 4 | 4 | 5 |
| Score of sample $a_{2}\left(\mathbf{s}_{i}(2)\right)$ | 2 | 2 | 3 | 3 | 3 | 5 | 5 | 5 | 5 |

Table 7.7: The scores assigned by the panellists in Example 7.5

From Table 7.7, it can be clearly seen that the median for sample $a_{1}$ is 3 and the median for sample $a_{2}$ is 3 (i.e., the score in the middle, in this case for $i=5$ ). Analogously, we consider the problem defined by Eq. 7.11), and we compute the sum of distances between the scores provided by the panellists and every possible vector of scores ${ }^{8}$. The results are illustrated in Table 7.8.

We see that the minimizer of the values in the first column - thus the mode - is the vector of scores $(3,5)$, the minimizer of the values in the second column - thus the median - is the vector of scores $(3,3)$, and the minimizer of the values in the third column - thus the mean - is the vector of scores $(3,4)$. Note that, as expected, these vectors coincide with the result of performing the mode, median and mean in each of the different samples separately.

In the setting where it is known that the first sample is fresher than the second sample (they are samples from different time spans of the shelf life of the same food), it is expected that the score of the first sample should be greater than or equal to the score of the second sample. Finding a solution by simply looking at the table is not an easy task. Thus, the problem defined by Eq. 7.12) is considered. The set of constraints $\mathcal{S}$ is formed by the vectors of scores in which the first sample is assigned a score greater than or equal to the score of the second sample. Such vectors are highlighted in gray in Table 7.8. It can be seen that the minimizer of the values in each column - thus the constrained mode, median and mean - is the vector of scores $(3,3)$. A conclusion is reached that the score assigned to sample $a_{1}$ should be greater than that originally assigned.

[^28]| $\mathbf{s}$ | $\sum_{j=1}^{2} \sum_{i=1}^{9} \partial_{0}\left(\mathbf{s}(j), \mathbf{s}_{i}(j)\right)$ | $\sum_{j=1}^{2} \sum_{i=1}^{9} \partial_{1}\left(\mathbf{s}(j), \mathbf{s}_{i}(j)\right)$ | $\sum_{j=1}^{2} \sum_{i=1}^{9} \partial_{2}\left(\mathbf{s}(j), \mathbf{s}_{i}(j)\right)$ |
| :---: | :---: | :---: | :---: |
| $(1,1)$ | 16 | 41 | 125 |
| $(1,2)$ | 14 | 32 | 86 |
| $(1,3)$ | 13 | 27 | 65 |
| $(1,4)$ | 16 | 28 | 62 |
| $(1,5)$ | 12 | 29 | 77 |
| $(2,1)$ | 17 | 36 | 100 |
| $(2,2)$ | 15 | 27 | 61 |
| $(2,3)$ | 14 | 23 | 40 |
| $(2,4)$ | 17 | 24 | 37 |
| $(2,5)$ | 13 | 33 | 52 |
| $(3,1)$ | 15 | 24 | 93 |
| $(3,2)$ | 13 | $\mathbf{1 9}$ | 54 |
| $(3,3)$ | 12 | 20 | $\mathbf{3 3}$ |
| $(3,4)$ | 15 | 36 | $\mathbf{3 0}$ |
| $(3,5)$ | 11 | 27 | 45 |
| $(4,1)$ | 16 | 22 | 104 |
| $(4,2)$ | 14 | 23 | 65 |
| $(4,3)$ | 13 | 43 | 44 |
| $(4,4)$ | 16 | 34 | 41 |
| $(4,5)$ | 12 | 29 | 56 |
| $(5,1)$ | 17 | 30 | 133 |
| $(5,2)$ | 15 | 31 | 94 |
| $(5,3)$ | 14 |  | 73 |
| $(5,4)$ | 17 | 13 |  |
| $(5,5)$ | 13 |  | 80 |

Table 7.8: Sum of zero-one distance $\partial_{0}$, absolute distance $\partial_{1}$ and squared difference $\partial_{2}$ between the scores provided by the panellists and all possible vectors of scores. The minimizers are shown in bold and the vectors of scores in which the first sample is assigned a score greater than or equal to the score of the second sample are highlighted in blue.

Remark Note that the difference between the number of panellists for each experiment can be extremely large in some instances. For instance, a very small number of panellists are gathered for one experiment, and a larger number of panellists are gathered for another experiment. Such a scenario can be approached from two different points of view: (a) each panellist is represented by one evaluation or (b) each sample is represented by one evaluation. The former approach is analogous to the problem defined by Eq. 7.12 , whereas the latter approach can be defined as follows:

$$
\begin{equation*}
\mathbf{s}^{*}=\underset{\mathbf{s} \in \mathcal{\delta}}{\arg \min } \sum_{j=1}^{n} \sum_{i=1}^{m_{j}} \frac{\partial\left(\mathbf{s}(j), \mathbf{s}_{i}(j)\right)}{m_{j}} \tag{7.13}
\end{equation*}
$$

where the evaluations are averaged based on the number of panellists $m_{j}$ that provide scores to the $j$-the sample. It must be noted that both approaches are
equivalent if all $m_{j}$ 's coincide. Moreover, both problems are also equivalent if there are no constraints (i.e., solving the problem defined by Eq. 7.8.

Example 7.6. Consider a simple setting where one panellist assigns a score to a given food sample and nine panellists each assign a score to a second given food sample on the 5-point scale fixed in Figure 4.2, and only one panellist assigns a score to a second given food sample. The scores are represented in increasing value in Table 7.9 .

| The $i$-th lowest score | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Score sample $1\left(s_{i}(1)\right)$ | 2 |  |  |  |  |  |  |  |  |
| Score sample $2\left(s_{i}(2)\right)$ | 2 | 2 | 3 | 3 | 3 | 5 | 5 | 5 | 5 |

Table 7.9: The scores assigned by the panellists in Example 7.6

To determine the vector of consensus scores, the problem defined by Eq. 7.11) is considered, and, for each of the 25 possible vectors of scores, the sum of distances to the scores provided by the panellists is computed. The minimizer of the sum of zero-one distances - thus the mode - is the vector of scores $(2,5)$, the minimizer of the sum of absolute distances - thus the median - is the vector of scores (2,3), and the minimizer of the sum of $\ell_{2}$-distances - thus the mean - is the vector of scores $(2,4)$.

Consider now that we know that the first sample is fresher than the second sample. Based on the approach where each panellist is represented by one evaluation, the problem defined by Eq. 7.12) is solved, resulting in the vectors of scores $(5,5)$ as the constrained mode, $(3,3)$ as the constrained median, and $(3,3)$ and $(4,4)$ as the constrained means. Based on the approach where each sample is represented by one evaluation, the problem defined by Eq. 7.13) is solved, resulting in the vectors of scores $(2,2)$ as the constrained mode, $(2,2)$ as the constrained median, and $(3,3)$ as the constrained mean. If we consider that each panellist is represented by one evaluation (i.e., the score of each sample depends on the number of panellists), then it seems logical to change the mode and median score of $a_{1}$ to 3. However, if we consider that each sample is represented by one evaluation (i.e., the score of each sample depends on the proportion of panellists), then it seems logical to change the mode or median score of $a_{2}$ to 2 .

### 7.6. Application to sensory data

This section illustrates the proposed methods by presenting an experiment on raw Atlantic salmon (Salmo salar), where trained panellists were asked to evaluate each salmon sample and assign a score. As the number of trained panellists was limited,
untrained panellists were asked to rank the salmon samples, and the influence of integrating the rankings on the quality of the assessment of the overall freshness of salmon samples was studied. This is illustrated by determining the consensus scores of each salmon sample at different storage days. In addition, using the same data, we illustrate the influence of integrating the scores on the quality of the assessment of the consensus ranking of salmon samples. Finally, we apply the method of incorporating additional information, such as knowledge of storage days, consensus rankings and results of a clustering analysis, of the salmon samples with the scores provided by the trained panellists to jointly find the consensus score of each of these samples.

### 7.7. Integrating rankings for assigning consensus scores

To determine the consensus score that can be assigned to each of the four samples on each day, while integrating the rankings in Table A. 12 with the scores in Table A.11, we considered the problem defined by Eq. 7.6) and computed $C_{\alpha}(\mathbf{s})$ for each of the 625 possible vectors of scores. Figure 7.13 illustrates all the minimizers $\mathbf{s}_{\alpha}^{*}$ that minimized $C_{\alpha}(\mathbf{s})$ for at least one value of $\left.\left.\alpha \in\right] 0,1\right]$ on each of the five days (Monday-Friday).

Figure 7.13 illustrates (in black) all the vectors of scores $\mathbf{s}_{\alpha}^{*}$ that were not contradicted by the minimizer(s) $\precsim^{*}$ as a fan-shaped pattern starting at $\alpha=0$ and the (highlighted) vectors of scores that were minimizers for $\alpha \in] 0,1]$. It can be seen in Figures 7.13 (b) and 7.13 (c) that there was a single minimizer $\mathbf{s}_{\alpha}^{*}$ on Thursday and Monday for all values of $\alpha \in] 0,1]$ that was also not contradicted by $\precsim^{*}$. Since we wanted to study the influence of the rankings on the quality of the assessment of the overall freshness of salmon samples, we focused on the results of Tuesday, Wednesday and Friday, where there were multiple minimizers $\mathbf{s}_{\alpha}^{*}$ for different values of $\alpha \in] 0,1]$.

The obtained minimizers $\mathbf{s}_{\alpha}^{*}$ for Tuesday, Wednesday and Friday are summarized as follows:

$$
\begin{aligned}
& \mathbf{s}_{\alpha, \text { Tue }}^{*}=\left\{\begin{array}{ccr}
\{(4,4,3,4)\} & , \text { if } & 0<\alpha<\frac{32}{41}, \\
\{(4,4,3,4),(5,4,3,4)\} & , \text { if } & \alpha=\frac{32}{41}, \\
\{(5,4,3,4)\} & , \text { if } & \alpha>\frac{32}{41} .
\end{array}\right. \\
& \mathbf{s}_{\alpha}^{*} \text {, Wed }=\left\{\begin{array}{ccr}
\{(4,2,2,2)\} & , \text { if } & 0<\alpha<\frac{50}{71}, \\
\{(4,2,2,2),(4,3,2,2)\} & , \text { if } & \alpha=\frac{50}{71}, \\
\{(4,3,2,2)\} & , \text { if } & \alpha>\frac{50}{71} .
\end{array}\right. \\
& \mathbf{s}_{\alpha, \text { Fri }}^{*}=\left\{\begin{array}{ccr}
\{(4,4,2,2)\} & , \text { if } & 0<\alpha<\frac{280}{349}, \\
\{(4,4,2,2),(4,3,2,2)\} & , \text { if } & \alpha=\frac{280}{349}, \\
\{(4,3,2,2)\} & , \text { if } & \alpha>\frac{280}{349} .
\end{array}\right.
\end{aligned}
$$

The influence of integrating the rankings on the consensus vectors of scores for $\alpha \in] 0,1]$ is visualized in Figure 7.14 . We deduce on the basis of the vectors of scores provided by the trained panellists that $\mathrm{A}^{1}$ was fresher than $\mathrm{B}^{2}$ and $\mathrm{D}^{4}$. However, integrating the rankings on that day indicates that these samples were similar in terms of freshness. Similarly, we deduce on the basis of the vectors of scores provided by the trained panellists that $\mathrm{B}^{5}$ was fresher than $\mathrm{C}^{6}$ and $\mathrm{D}^{7}$. However, integrating the rankings on that day indicates that these samples were similar. Finally, we deduce on the basis of the vectors of scores provided by the trained panellists that $\mathrm{B}^{6}$ was less fresh than $\mathrm{A}^{5}$. However, integrating the rankings on that day indicates that these samples were similar.

It must be noted that the different salmon fillets were not identical, and thus, the storage days could not be used as the only tool to compare samples from different fillets. For instance, one cannot simply presume that samples from different fillets that have been stored for the same number of days will be similar. Similarly, one cannot presume that a sample from one fillet is always preferred over a sample from another fillet that has been stored for longer. Therefore, interpreting the resulting consensus scores cannot be done based solely on the storage day of the salmon samples.

As odours (and flavours) are often produced by volatile organic compounds (VOCs), the composition of the VOC profiles has been successfully used to evaluate the quality of food, such as seafood [65] and meats 66. SIFT-MS has attracted the attention of many researchers for rapid and accurate characterization of VOCs and has been validated for fish metabolite research [65, 61].

To establish a relation between the VOC profiles and the resulting scores, hierarchical agglomerative clustering [67] was performed on the VOC profiles in Table B.14. To measure how well a generated dendrogram reflects data accuracy, one can compute the cophenetic correlation coefficient [72, 175]. Generally, the results of different distance metrics and clustering algorithms should be compared and the


Figure 7.13: The vectors of scores that minimize $C_{\alpha}(\mathbf{s})$ for at least one value of $\left.\left.\alpha \in\right] 0,1\right]$. combination resulting in the largest cophenetic correlation coefficient ${ }^{9}$ is normally

[^29]

Figure 7.14: A visualization of the change in scores assigned to each sample on (a) Tuesday for $0<\alpha<\frac{32}{41}$ (b) Wednesday for $0<\alpha<\frac{50}{71}$ and (c) Friday for $0<\alpha<\frac{280}{349}$.
selected. To determine the optimal number of clusters, the gap value, which is commonly used in hierarchical clustering [73], is computed. To obtain an ideal clustering, the number of clusters that maximizes the gap value must be selected. However, in many real-world datasets, the clusters are not so well-defined. As a result, a balance between maximizing the gap value and the parsimony of the model is usually determined.

The clusters were represented on the basis of the Euclidean distance between their centroids resulted in the largest cophenetic correlation coefficient of 0.8053 . The corresponding dendrogram is shown in Figure 7.15(a), where it naturally divided the samples into distinct clusters (i.e., the groups of samples were densely packed in certain areas and not in others). From Figure 7.15(b), it can be seen that the minimum number of clusters that maximized the gap value (i.e., the point where the rate of increase of the gap value starts to decrease) was four. In Figure 7.15(a), these four clusters correspond to a horizontal slice across the dendrogram at a height of 1.8.

First, to obtain an ideal clustering, the number of clusters that maximizes the gap value must be selected. However, in many real-world datasets, the clusters are


Figure 7.15: (a) Hierarchical clustering of the samples of fillets A, B, C and D on the basis of the log-transformed VOC profile data, with the optimal number of clusters corresponding to the horizontal slice (dashed line). (b) Estimation of the optimal number of clusters in hierarchical clustering by maximizing the gap value.
not as well-defined. As a result, a balance between maximizing the gap value and the parsimony of the model is usually determined. Then, based on the clustered samples in Figure 7.15 (a), we try to explain the changes in the scores assigned by the trained panellists when integrating the rankings provided by the untrained panellists. We deduce, based on the dendrogram in Figure 7.15 (a), that samples $A^{1}$ and $A^{2}$ form the first cluster, sample $B^{2}$ forms the second cluster, samples $\mathrm{A}^{3}$, $\mathrm{A}^{4}, \mathrm{~A}^{5}, \mathrm{~A}^{6}, \mathrm{~B}^{3}, \mathrm{~B}^{4}, \mathrm{~B}^{5}, \mathrm{~B}^{6}, \mathrm{C}^{3}, \mathrm{C}^{4}, \mathrm{C}^{6}, \mathrm{D}^{4}, \mathrm{D}^{5}$ and $\mathrm{D}^{6}$ form the third cluster, and
samples $\mathrm{C}^{5}, \mathrm{C}^{7}, \mathrm{D}^{7}$ and $\mathrm{D}^{8}$ form the fourth cluster.
We conclude that samples $B^{6}$ and $A^{5}$ are similar in terms of their VOC profiles, thus, the change in the score assigned to sample $\mathrm{B}^{6}$ on Friday for $0<\alpha<\frac{280}{349}$ agrees with this conclusion. Similarly, we conclude that samples $\mathrm{B}^{5}$ and $\mathrm{C}^{6}$ are similar in terms of their VOC profiles, thus, the change in the score assigned to sample $\mathrm{B}^{5}$ on Wednesday for $0<\alpha<\frac{50}{71}$ also agrees with this conclusion. Finally, since we cannot conclude that samples $\mathrm{A}^{1}, \mathrm{~B}^{2}$ and $\mathrm{D}^{4}$ are similar, the change in the score assigned to sample $\mathrm{A}^{1}$ on Tuesday for $0<\alpha<\frac{50}{71}$ is not validated by the conclusion we obtained from the dendrogram.

In light of these findings, it appears that integrating rankings showed an increase in the quality of the assigned consensus score. As relations between the resulting consensus scores for any of the salmon samples of the different fillets and the clustering of the samples were significant, these results support the idea of combining scoring and ranking methods.

### 7.7.1. Integrating scores for determining consensus rankings

We now illustrate the method of improving the quality of the assessment of a consensus ranking using the scores and rankings in Table A.11 and Table A.12. respectively. Although these sensory data were gathered for the aim of improving the quality of the assessment of a consensus vector of scores, we apply our method to these data as an illustration of integrating scores to improve the quality of the assessment of a consensus ranking.

To determine the consensus ranking of the four samples on each day, while integrating the gathered scores with the rankings, we considered the problem defined by Eq. 7.10) and computed $D_{\alpha}(\precsim)$ for each of the 75 possible rankings. Figure 7.16 illustrates all the minimizers $\precsim_{\alpha}^{*}$ that minimized $D_{\alpha}(\precsim)$ for at least one value of $\alpha \in[0,1[$ on each of the five days (Monday-Friday).

Figure 7.16 illustrates (in black) all the rankings $\precsim_{\alpha}^{*}$ that were not contradicted by the minimizer(s) $\mathbf{s}^{*}$ as a fan-shaped pattern starting at $\alpha=1$ and the (highlighted) rankings that were minimizers for $\alpha \in[0,1[$. It can be seen in Figures 7.16(b) and 7.16(c) that there were two overlapping minimizers $\precsim_{\alpha}^{*}$ on Thursday and Monday for all values of $\alpha \in\left[0,1\left[\right.\right.$ that was also not contradicted by $\mathbf{s}^{*}$. Since we wanted to study the influence of the vectors of scores on the quality of the assessment of the overall ranking of salmon samples, we focused on the results of Tuesday, Wednesday and Friday, where there were multiple minimizers $\swarrow_{\alpha}^{*}$ for different values of $\alpha \in[0,1[$.

The obtained minimizers $\mathbf{s}_{\alpha}^{*}$ for Tuesday, Wednesday and Friday are summarized as follows:

$$
\begin{aligned}
& \precsim_{\alpha, \text { Tue }}^{*}=\left\{\begin{array}{ccc}
\left\{\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~B}^{2} \sim \mathrm{~A}^{1}\right\} & , \text { if } & 0<\alpha<\frac{32}{41}, \\
\left\{\begin{array}{c}
\mathrm{C}^{3} \\
\prec \mathrm{D}^{4} \prec \mathrm{~B}^{2} \prec \mathrm{~A}^{1} \\
\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~B}^{2} \sim \mathrm{~A}^{1}
\end{array}\right\}, & \text { if } & \alpha=\frac{32}{41}, \\
\left\{\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~B}^{2} \prec \mathrm{~A}^{1}\right\} \quad, & \text { if } & \alpha>\frac{32}{41} .
\end{array}\right. \\
& \precsim_{\alpha, \text { Wed }}^{*}=\left\{\begin{array}{ccc}
\left\{\mathrm{D}^{7} \prec \mathrm{C}^{6} \sim \mathrm{~B}^{5} \prec \mathrm{~A}^{4}\right\} & , \text { if } & \alpha<\frac{50}{71}, \\
\left\{\begin{array}{c}
\mathrm{D}^{7}
\end{array} \mathrm{C}^{6} \sim \mathrm{~B}^{5} \prec \mathrm{~A}^{4}\right. \\
\left.\mathrm{D}^{7} \prec \mathrm{C}^{6} \prec \mathrm{~B}^{5} \prec \mathrm{~A}^{4}\right\}, & , \text { if } & \alpha=\frac{50}{71}, \\
\left\{\mathrm{D}^{7} \prec \mathrm{C}^{6} \prec \mathrm{~B}^{5} \prec \mathrm{~A}^{4}\right\} & , \text { if } & \frac{50}{71}<\alpha<1 .
\end{array}\right.
\end{aligned}
$$

We deduce on the basis of the rankings provided by the untrained panellists that $\mathrm{A}^{1}$ and $\mathrm{B}^{2}$ are tied. However, integrating the scores on that day indicates that $A^{1}$ is preferred over $B^{2}$. Similarly, we deduce that $B^{5}$ and $C^{6}$ are tied, however, integrating the scores on that day indicates that $\mathrm{B}^{5}$ is preferred over $\mathrm{C}^{6}$. Finally, we deduce that $B^{6}$ is preferred over $A^{5}$, however, integrating the scores on that day indicates that $\mathrm{A}^{5}$ is preferred over $\mathrm{B}^{6}$.

### 7.7.2. Incorporating knowledge of storage days for assigning joint consensus scores

To determine the consensus vectors of scores that should be assigned to the five samples of each fillet, we considered the problem defined by Eq. 7.11), and, for each of the 3125 vectors of scores, we computed the sum of distances to the vectors of scores provided by the panellists in Table A.11 Similarly, to determine the consensus vectors of scores that should be assigned to the five samples of the same fillet, while integrating the knowledge of storage days of the samples, we considered the problem defined by Eq. (7.12), where the constraints of a set $\mathcal{S}$ are summarized in Table 7.10 .

The median and the constrained median for each sample of every fillet are summarized in Table 7.11

The constrained medians for the samples are illustrated in Figure 7.17. After incorporating the knowledge of storage days of the samples from each of fillets C and D, improved results are seen in Figure 7.17, where there are no increasing values of evaluations for any fillet.


Figure 7.16: The rankings that minimize $D_{\alpha}(\precsim)$ for at least one value of $\alpha \in[0,1[$.

Note that including the knowledge of storage days should be considered when several factors of the initial conditions of the samples are similar, namely, their

| Fillet | Constraints |
| :---: | :---: |
| A | $\mathrm{A}^{5} \prec \mathrm{~A}^{4} \prec \mathrm{~A}^{3} \prec \mathrm{~A}^{2} \prec \mathrm{~A}^{1}$ |
| B | $\mathrm{~B}^{6} \prec \mathrm{~B}^{5} \prec \mathrm{~B}^{4} \prec \mathrm{~B}^{3} \prec \mathrm{~B}^{2}$ |
| C | $\mathrm{C}^{7} \prec \mathrm{C}^{6} \prec \mathrm{C}^{5} \prec \mathrm{C}^{4} \prec \mathrm{C}^{3}$ |
| D | $\mathrm{D}^{8} \prec \mathrm{D}^{7} \prec \mathrm{D}^{6} \prec \mathrm{D}^{5} \prec \mathrm{D}^{4}$ |

Table 7.10: The constraints based on the storage days of the samples.

| Method | $\mathrm{A}^{1}$ | $\mathrm{~A}^{2}$ | $\mathrm{~A}^{3}$ | $\mathrm{~A}^{4}$ | $\mathrm{~A}^{5}$ | $\mathrm{~B}^{2}$ | $\mathrm{~B}^{3}$ | $\mathrm{~B}^{4}$ | $\mathrm{~B}^{5}$ | $\mathrm{~B}^{6}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| The median | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 3 |
| The constrained median | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 3 |
|  |  |  |  |  |  |  |  |  |  |  |
| Method | $\mathrm{C}^{3}$ | $\mathrm{C}^{4}$ | $\mathrm{C}^{5}$ | $\mathrm{C}^{6}$ | $\mathrm{C}^{7}$ | $\mathrm{D}^{4}$ | $\mathrm{D}^{5}$ | $\mathrm{D}^{6}$ | $\mathrm{D}^{7}$ | $\mathrm{D}^{8}$ |
| The median | 3 | 4 | 3 | 2 | 2 | 4 | 2 | 4 | 2 | 2 |
| The constrained median | 3 | 3 | 3 | 2 | 2 | 4 | 3 | 3 | 2 | 2 |

Table 7.11: The median and the constrained median for each sample after incorporating knowledge of storage days. The differences are highlighted.
contamination, dimensions, composition and packaging and storage conditions. Verifying the similarity of the samples (equivalently, their rate of spoilage) requires measurements. From Table 8, it is deduced that, in an ideal situation where the samples from the same fillet were initially similar, incorporating the knowledge of storage days of the samples from fillet $C$ indicates that sample $C^{4}$ should be assigned a lower score and that samples $\mathrm{C}^{3}, \mathrm{C}^{4}$, and $\mathrm{C}^{5}$ should be equally scored in terms of freshness. Moreover, it is deduced that incorporating the knowledge of storage days of the samples from fillet D indicates that sample $\mathrm{D}^{5}$ should be assigned a higher score and sample $\mathrm{D}^{6}$ should be assigned a lower score and these samples should be scored equally in terms of freshness.

There exist several potential risks when incorporating knowledge of storage days of samples that are not initially similar. The main concern is that an assumption is made that the samples have similar spoilage rates, and, thus, their assigned scores might be incorrectly related to their storage days. Since microbiological analysis of each salmon sample was not performed, the storage days will not be used as the only tool to compare these samples.


Figure 7.17: Illustrating the median and constrained median of fillets A, B, C and D over time after incorporating knowledge of storage days.

### 7.7.3. Incorporating results of a clustering analysis for assigning joint consensus scores

Based on the aforedescribed methods of using SIFT-MS to quantify the VOC profiles, the results of the clustering are summarized in Table 7.12. To determine the consensus vectors of scores that should be assigned to the samples in each group, while incorporating the results of the clustering analysis of the samples, we considered the problem defined by Eq. 7.12), where the constraints are the clusters summarized in Table 7.12 .

To determine the consensus vector of scores that should be assigned to all the samples, while incorporating the results of the clustering analysis of the samples, the problem defined by Eq. 7.12 with a threshold $\epsilon=1$ on the absolute difference of the scores of the samples is considered. The medians and the constrained medians

| Cluster | Samples |
| :---: | :---: |
| Cluster 1 | $\left\{\mathrm{A}^{1}, \mathrm{~A}^{2}\right\}$ |
| Cluster 2 | $\left\{\mathrm{B}^{2}\right\}$ |
|  | $\left\{\mathrm{A}^{3}, \mathrm{~A}^{4}, \mathrm{~A}^{5}\right.$, |
| Cluster 3 | $\mathrm{B}^{3}, \mathrm{~B}^{4}, \mathrm{~B}^{5}, \mathrm{~B}^{6}$, |
|  | $\mathrm{C}^{3}, \mathrm{C}^{4}, \mathrm{C}^{6}$, |
|  | $\left.\mathrm{D}^{4}, \mathrm{D}^{5}, \mathrm{D}^{6}\right\}$ |
| Cluster 4 | $\left\{\mathrm{C}^{5}, \mathrm{C}^{7}, \mathrm{D}^{7}, \mathrm{D}^{8}\right\}$ |

Table 7.12: Clustered samples based on the similarity of their VOC profile.
for the samples of each fillet are gathered in Table 7.13 .

| Method | Cluster 1 |  | $\frac{\text { Cluster } 2}{\text { B }^{2}}$ | Cluster 4 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{A}^{1}$ | $\mathrm{A}^{2}$ |  | $\mathrm{C}^{5}$ | $\mathrm{C}^{7}$ | $\mathrm{D}^{7}$ | $\mathrm{D}^{8}$ |
| The median | 5 | 5 | 4 | 3 | 2 | 2 | 2 |
| The constrained median | 5 | 5 | 4 | 3 | 2 | 2 | 2 |


| Method | Cluster 3 |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{A}^{3}$ | $\mathrm{A}^{4}$ | $\mathrm{A}^{5}$ | $\mathrm{B}^{3}$ | $B^{4}$ | $\mathrm{B}^{5}$ | $\mathrm{B}^{6}$ | $\mathrm{C}^{3}$ | $\mathrm{C}^{4}$ | $\mathrm{C}^{6}$ | $\mathrm{D}^{4}$ | $\mathrm{D}^{5}$ | $\mathrm{D}^{6}$ |
| The median | 4 | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 4 | 2 | 4 | 2 | 4 |
| The constrained median | 4 | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 4 | 3 | 4 | 3 | 4 |

Table 7.13: The median and the constrained median for each sample after incorporating the results of clustering analysis. The differences are highlighted.

From Table 7.13 , it is deduced on the basis of the scores provided by the panellists to all the panellists that in cluster 3 , samples $\mathrm{C}^{6}$ and $\mathrm{D}^{5}$ are less fresh than all the other samples in the cluster. However, incorporating the results of the clustering analysis of all the samples indicates that samples $\mathrm{C}^{6}$ and $\mathrm{D}^{5}$ should each be assigned a higher score.

Note that including the results of a clustering analysis should only be considered when the measurements are accurate and the clusters are well defined. Otherwise, there exists the potential risk of inaccurately clustering samples, and, thus, assigning incorrect scores. In many real-world datasets, there is no absolute optimal number of clusters. As a result, a balance between a clustering that reflects the data best and a parsimonious model should be determined. For instance, a very small number of clusters may result in a large number of samples in a single cluster, and, thus, their assigned scores

### 7.7.4. Incorporating consensus rankings for assigning joint consensus scores

To determine the consensus scores and the consensus rankings of the samples on each day, the problems defined by Eq. 7.1), using the absolute distance function $\partial_{1}$, and Eq. 7.2 were considered, and the minimizers s* and $\precsim^{*}$ of the scores in Table A.11 and the rankings in Table A.12, respectively, were computed. The results are gathered in Table 7.14 .

|  | Group 1 | Group 2 |  |
| :---: | :---: | :---: | :---: |
| Tuesday | Thursday | Group 3 <br> Monday |  |
| $\mathbf{s}^{*}$ | $(5,4,3,4)$ | $(5,4,2,2)$ | $(4,4,3,4)$ |
| $\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~B}^{2} \sim \mathrm{~A}^{1}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \sim \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \sim \mathrm{~A}^{3} \prec \mathrm{~B}^{4}$ |  |
|  |  | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \sim \mathrm{~B}^{3} \sim \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~A}^{3} \sim \mathrm{~B}^{4}$ |
|  |  | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \sim \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ |  |
|  |  | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ |  |


|  | Group 4 | Group 5 |
| :---: | :---: | :---: |
|  | Wednesday | Friday |
| $\mathbf{s}^{*}$ | $(4,3,2,2)$ | $(4,3,2,2)$ |
| $\precsim^{*}$ | $\mathrm{D}^{7} \prec \mathrm{C}^{6} \sim \mathrm{~B}^{5} \prec \mathrm{~A}^{4}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |

Table 7.14: The minimizers $\mathrm{s}^{*}$ and $\precsim^{*}$ of salmon samples of fillets (A, B, C, D) on each day of the week in the order shown in Table 4.14 .

To determine the consensus vectors of scores that should be assigned to the samples in each group, while incorporating the consensus rankings of the samples, we considered the problem defined by Eq. 7.12, where the constraints of a defined set $\mathcal{S}$ are the consensus rankings summarized in Table 7.14 .

To determine the consensus vectors of scores that should be assigned to the four samples in each group, while incorporating the consensus rankings of the samples, the problem defined by Eq. 7.12 is considered. The medians and the constrained medians for the samples in each group are summarized in Table 7.15

From Table 7.15, it is deduced on the basis of the vectors of scores provided by the panellists to the samples in group 4 that sample $\mathrm{B}^{5}$ is fresher than sample $\mathrm{C}^{6}$ and sample $\mathrm{D}^{7}$. However, incorporating the consensus ranking of the samples in group 4 indicates that sample $B^{5}$ is similar to sample $C^{6}$. Therefore, sample $B^{5}$ should be assigned a lower score, resulting in equal scores assigned to samples $\mathrm{B}^{5}$, $C^{6}$ and $D^{7}$. Similar conclusions can be drawn for groups ( $1,2,3$ and 5 ).

| Method | Group 1 |  |  |  | Group 2 |  |  |  | Group 3 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{A}^{1}$ | $\mathrm{B}^{2}$ | $\mathrm{C}^{3}$ | $\mathrm{D}^{4}$ | $\mathrm{A}^{2}$ | $\mathrm{B}^{3}$ | $\mathrm{C}^{4}$ | $\mathrm{D}^{5}$ | $\mathrm{A}^{3}$ | $B^{4}$ | $\mathrm{C}^{5}$ | $\mathrm{D}^{6}$ |
| The median | 5 | 4 | 3 | 4 | 5 | 4 | 4 | 2 | 4 | 4 | 3 | 4 |
| The constrained median | 4 | 4 | 3 | 4 | 5 | 4 | 4 | 2 | 4 | 4 | 3 | 4 |
|  |  |  |  |  | 4 | 4 | 4 | 2 |  |  |  |  |


|  | Group 4 |  |  |  | Group 5 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Method | $\mathrm{A}^{4}$ | $B^{5}$ | $\mathrm{C}^{6}$ | $\mathrm{D}^{7}$ | $\mathrm{A}^{5}$ | $\mathrm{B}^{6}$ | $\mathrm{C}^{7}$ | $\mathrm{D}^{8}$ |
| The median | 4 | 3 | 2 | 2 | 4 | 3 | 2 | 2 |
| The constrained median | 4 | 2 | 2 | 2 | 4 | 4 | 2 | 2 |

Table 7.15: The median and constrained median for each sample in every group after incorporating knowledge of consensus rankings in Table 7.14. The differences are highlighted.

Note that incorporating a consensus ranking should only be considered when the number of panellists providing a ranking is large enough. Otherwise, there exists the potential risk of inaccurately ordering samples, and, thus, assigning incorrect scores. In this experiment, the number of panellists providing rankings on the salmon samples was between 23 and 28 depending on the group. This number may be considered to be large enough for obtaining consensus rankings, however, having more panellists would result in more reliable consensus rankings. Therefore, in this study we will not use the consensus rankings as the only tool to compare these samples.

### 7.7.5. Comparing the consensus vectors of scores

To study the influence of incorporating the previously discussed information that invoked different constraints on the median of each sample, we summarize all the medians and the constrained medians for each setting in Table 7.16

It can be seen that the medians and the constrained medians for some of the samples are equal. A conclusion is reached that the scores assigned to these samples agree with each of the additional information of the samples. However, the medians and the constrained medians for the other samples differ. The additional constraints might provide a better understanding of the score that should be assigned to each sample.

It is important for the reader to note that choosing an optimal source of additional information depends on the quality of that information. For instance, to use the previously described sources of information, it is recommended that the initial conditions of the samples should be similar, the clustering methods should be

| Method | Group 1 |  |  |  | Group 2 |  |  |  | Group 3 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{A}^{1}$ | $\mathrm{B}^{2}$ | $\mathrm{C}^{3}$ | $\mathrm{D}^{4}$ | $\mathrm{A}^{2}$ | $\mathrm{B}^{3}$ | $\mathrm{C}^{4}$ | $\mathrm{D}^{5}$ | $\mathrm{A}^{3}$ | $\mathrm{B}^{4}$ | $\mathrm{C}^{5}$ | $\mathrm{D}^{6}$ |
| The median | 5 | 4 | 3 | 4 | 5 | 4 | 4 | 2 | 4 | 4 | 3 | 4 |
| $\mathrm{s}_{\alpha}^{*}$ | 4 | 4 | 3 | 4 | 5 | 4 | 4 | 2 | 4 | 4 | 3 | 4 |
| The constrained median |  |  |  |  |  |  |  |  |  |  |  |  |
| Inc. knowledge of storage days | 5 | 4 | 3 | 4 | 5 | 4 | 3 | 3 | 4 | 4 | 3 | 3 |
| Inc. results of clustering analysis | 5 | 4 | 3 | 4 | 5 | 4 | 4 | 3 | 4 | 4 | 3 | 4 |
| Inc. consensus ranking | 4 | 4 | 3 | 4 | 5 | 4 | 4 | 2 | 4 | 4 | 3 | 4 |
|  |  |  |  |  | 4 | 4 | 4 | 2 |  |  |  |  |


|  | Group 4 |  |  |  | Group 5 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Method | $\mathrm{A}^{4}$ | $\mathrm{B}^{5}$ | $\mathrm{C}^{6}$ | $D^{7}$ | $\mathrm{A}^{5}$ | $\mathrm{B}^{6}$ | $\mathrm{C}^{7}$ | $\mathrm{D}^{8}$ |
| The median | 4 | 3 | 2 | 2 | 4 | 3 | 2 | 2 |
| $\mathbf{s}_{\alpha}^{*}$ | 4 | 2 | 2 | 2 | 4 | 4 | 2 | 2 |


| The constrained median |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Inc. knowledge of storage days | 4 | 3 | 2 | 2 | 4 | 3 | 2 | 2 |
| Inc. results of clustering analysis | 4 | 3 | 3 | 2 | 4 | 3 | 2 | 2 |
| Inc. consensus ranking | 4 | 2 | 2 | 2 | 4 | 4 | 2 | 2 |

Table 7.16: The median, the consensus vector of scores $\mathbf{s}_{\alpha}^{*}$ (for small values of $\alpha$ close to but not equal to zero) and the constrained medians for each sample in every group after incorporating knowledge of storage days, results of clustering analysis and consensus rankings.
reliable and consistent, and the number of panellists performing ranking tests should be large.

Note that simultaneously incorporating knowledge of storage days, results of a clustering analysis and consensus ranking(s) of the samples would result in many constraints. As a result, the constrained median of samples $\mathrm{C}^{7}, \mathrm{D}^{7}$ and $\mathrm{D}^{8}$ stays the same while the constrained median of all the other samples is either 3 or 4 . The reader should bear in mind that adding too many constraints may result in forcing the scores of all the samples to be similar.

Interestingly, in groups $1,3,4$, and 5 , the consensus vector of scores $\mathbf{s}_{\alpha}^{*}$ (for small values of $\alpha$ close to but not equal to zero) is identical to the constrained median after incorporating the consensus rankings. However, in group 2 they are identical only for the consensus rankings where sample $\mathrm{A}^{2}$ is ranked higher than sample $B^{3}$.

### 7.8. Conclusions

In this chapter, we have proposed three methods for combining scores and rankings. We have presented a first method for assigning consensus scores to food samples, while integrating rankings. This is done by combining the median and the Kemeny median that allows to compute the 'distance' between each possible vector of scores and both the vectors of scores and the rankings provided by the trained and untrained panellists, respectively. This method is especially useful for sensory evaluation problems where the number of trained panellists providing scores is very small, and where it is easier to obtain additional information, in the form of rankings, by recruiting untrained panellists. This method is, however, not limited by the number of untrained panellists or samples.

We have presented a second method for determining a consensus ranking of food samples, while integrating scores. This is done by combining the median and the Kemeny median that allows to compute the 'distance' between each possible ranking and both the vectors of scores and the rankings provided by the trained and untrained panellists, respectively. This method is especially useful for sensory evaluation problems where the number of untrained panellists providing rankings is very small, and where it is helpful to obtain additional information, in the form of scores, by recruiting trained panellists. Similarly, this method is not limited by the number of trained panellists or samples.

We have presented a third method for assigning a constrained mode, median or mean by incorporating other types of information of the samples. The additional information is incorporated in the form of constraints, and the constrained mode, median or mean is the minimizer of sum of different distances. This method is thus not limited to a certain distance. This method is especially useful for sensory evaluation problems where the number of trained panellists providing scores is very small, and where additional information on the samples is known.

We have illustrated the use of these three methods on an experiment for determining the freshness of raw Atlantic salmon, as described in Chapter 4 . We have shown the influence of integrating rankings provided by untrained panellists on the consensus vector of scores, the influence of integrating vectors of scores provided by trained panellists on the consensus ranking, and the influence of incorporating additional information of the samples, namely storage days, clustering analyses and consensus rankings, on the consensus vector of scores. In addition, other potential applications for the aforementioned methods include, but are not limited to, decision making problems [130, 176], online valuation [177] and recommender systems (e.g., social matching systems and gift, music and movie recommenders) 120 . We end by noting that, in the field of food science, researchers are not only interested in determining the quality of food samples, but also in understanding and identifying the reasons for the scores assigned to the samples. Thus, the resulting consensus scores can be
of use in relating characteristics of samples to their assigned scores.
One could note that these methods are a starting point to apply the novel methods presented in Chapters 5 and 6. Thus, it would be interesting to improve the search for the closest monotone matrix of labels (respectively, the closest monotone profile of rankings), while integrating rankings with ordinal labels. One approach would be to consider a combination of the search for a closest monotone matrix of labels and the search for the closest monotone profile of rankings. Another approach would be to consider the optimization problem in Section 5.5 that leads to the computation of a closest monotone matrix of labels while further satisfying the constraints of $\mathcal{S}$. Expressing the monotonicity property when dealing with ordinal labels and rankings, at the same time, still remains an open problem that will be addressed in the near future.

# PREDICTION OF SENSORY EVALUATIONS 

## 8 Learning to predict ordinal labels

## Table of Contents

8.1 Introduction<br>8.2 Predictive modelling of ordinal labels<br>8.2.1 The ordinal regression problem<br>8.2.2 Performance measures for sensory evaluations<br>8.2.3 Regularization in ordinal regression<br>8.3 Experimental analysis<br>8.3.1 General experimental setup<br>8.3.2 Experimental settings using synthetic data<br>8.4 Application to sensory data<br>8.4.1 Chicken breast<br>8.4.2 Atlantic cod<br>8.4.3 Brown shrimp<br>8.4.4 Atlantic salmon<br>8.5 Conclusions

### 8.1. Introduction

The problem of correlating certain data of food with sensory evaluation is not new and has been tackled in the past [66, 178, 179, 180, 181, 182, 183]. Those studies focus on statistical modelling of sensory evaluations and on testing statistical hypotheses. Recently, studies have shown various ways of predicting sensory evaluations [60, [75, 83, 84, 184]. However, those studies focus on unsupervised statistical techniques.

In this chapter, we consider the problem of predicting the appreciation of a food sample. A traditional approach for solving this supervised prediction problem starts with a data collection step, where a number of samples is gathered and a feature representation of these samples is obtained. Subsequently, several people are asked to express their appreciation of these samples. From these data, we can, using statistical approaches, learn a model that can map a sample's feature representation to an appreciation and thus solve the prediction problem. The feature representation consists of a number of chemical and/or physical characteristics of these samples. To obtain an appreciation, the samples are evaluated by one or more panellists in terms of perceived freshness. Recently, the prediction of the appreciaiton of food samples has appeared in several studies on food quality [84, 185, 186, 187, where
the appreciation is in the form of a score. However, these scores are often associated with ordinal labels, where the distance between them is latent. Therefore, in this chapter, we answer the following question:

## Question IV.1: How can we predict an ordinal label?

To predict ordinal labels, we use traditional ordinal regression models, but improve their predictive performance by introducing regularization. Notably, numerous models that describe absolute evaluations (in case of ordinal regression) use a latent variable. These latent variable models assume that every sample can be mapped to a real value using a so-called appreciation function. The variable obtained in this way is called the latent variable, and the appreciation function is sometimes called the latent variable function. Generally, the higher the value of the latent variable of a sample, the more it will be appreciated. This latent variable thus provides a way to rank samples according to their appreciation. Most preference learning strategies try to approximate this (unknown) appreciation function by means of a mapping $g$ from the feature space $\mathscr{X}$ to $\mathbb{R}$, and simply call $g$ the latent variable function. The use of the mapping $g$ to predict the result of a relative evaluation is rather straightforward. Consider a first sample with feature vector $\mathbf{x}_{1} \in \mathscr{X}$ and a second sample with feature vector $\mathbf{x}_{2} \in \mathscr{X}$. Whenever $g\left(\mathbf{x}_{2}\right)<g\left(\mathbf{x}_{1}\right)$, it is predicted that the first sample is ranked higher than the second. Analogously, whenever $g\left(\mathbf{x}_{1}\right)<g\left(\mathbf{x}_{2}\right)$, it is predicted that the second sample is ranked higher than the first, and, whenever $g\left(\mathbf{x}_{1}\right)=g\left(\mathbf{x}_{2}\right)$, it is predicted that the first sample is ranked equal to the second. To predict the result of an absolute evaluation, i.e., to which class of a set of linearly ordered classes a sample belongs, an additional step is needed. The sampling space of the latent variable (typically $\mathbb{R}$ ) is partitioned in as many intervals as there are classes. Each class is then linked with an interval of the partition such that the ordering of the classes matches the ordering of the intervals. As a result, the partition provides a natural way to map the predicted value of the latent variable to a class label and can thus be used to predict the result of an ordinal label.

Over the last few decades, several ordinal regression methods have been proposed to analyse ordinal data and learn a mapping from a feature space to a set of linearly ordered labels. Notable examples include the proportional-odds model [188], support vector ordinal regression [189] and nearest neighbours for ordinal classification [190]. Unfortunately, training and (subsequently) collecting information from trained panellists usually carries big expenses. For this reason, there usually is a limited amount of data available to learn a good predictive model, which does not suffer from overfitting or low predictive power. Statistical methods for predicting an ordinal label using sensory data and high-dimensional features are lacking. Therefore, in the following sections, we propose a strategy for reducing overfitting in ordinal regression models and improving their predictive power.

### 8.2. Predictive modelling of ordinal labels

In this section, we present the ordinal regression problem and an approach that can be adopted to include $\ell_{1}$-norm regularization.

### 8.2.1. The ordinal regression problem

As a starting point, we introduce some notations. Assume that our training data are identically and independently drawn observations of an unknown distribution over $\mathscr{X} \times \mathscr{Y}$, where $\mathscr{X}=\mathbb{R}^{p}$ is the input space and $p$ is the number of features. $\mathscr{Y}=\left\{L_{1}, \ldots, L_{q}\right\}$ is the label set containing $q$ linearly ordered labels:

$$
L_{1} \prec \ldots \prec L_{k} \prec \ldots \prec L_{q}
$$

where $k \in\{2, \ldots, q-1\}$. We represent each sample by a $p$-dimensional vector $\mathrm{x} \in \mathscr{X}$ and an ordinal label $y \in \mathscr{Y}$. We describe our dataset $\mathscr{D}$ as a set of $n$ couples $\left(\mathbf{x}_{i}, y_{i}\right)$ with $\mathbf{x}_{i}=\left(x_{i 1}, \ldots, x_{i p}\right)$, where the couples ( $\left.\mathbf{x}_{i}, y_{i}\right)$ are realizations of the random vector $(\mathcal{X}, \mathcal{Y})$.

Ordinal regression has been studied quite extensively in statistics [110, 189, 191, 192, 193, 194, 195, 196, 197. There are several types of ordinal regression models, however, the latent variable models are generally considered the most important ones [198, 199, 200, 201, 202. These models assume that a linearly ordered label set is the result of the discretization of an unobserved latent variable. Therefore, these ordinal regression models $f: \mathscr{X} \rightarrow \mathbb{R}$ are of the following form:

$$
f(\mathbf{x})= \begin{cases}L_{1} & , \text { if } g(\mathbf{x}) \leq \theta_{1}  \tag{8.1}\\ L_{k} & , \text { if } \theta_{k-1}<g(\mathbf{x}) \leq \theta_{k}, \quad \text { for } k \in\{2, \ldots, q-1\} \\ L_{q} & , \text { if } \theta_{q-1}<g(\mathbf{x})\end{cases}
$$

with $\theta_{1}<\ldots<\theta_{k-1}<\theta_{k}<\ldots<\theta_{q-1}$ the set of threshold parameters and $g: \mathscr{X} \rightarrow \mathbb{R}$ the function that models the latent variable.

Therefore, fitting an ordinal regression model consists of estimating a function $g$ and a set of thresholds, where $g$ assigns a real value to a sample, and the thresholds are used to partition the real line to obtain an ordinal response.

We now define the relationship between the output random variable $\mathcal{Y}$ and the input random vector $\mathcal{X}$ as follows:

$$
\mathcal{Y}= \begin{cases}L_{1} & , \text { if } g(\mathcal{X})+\mathcal{E} \leq \theta_{1}  \tag{8.2}\\ L_{k} & , \text { if } \theta_{k-1}<g(\mathcal{X})+\mathcal{E} \leq \theta_{k}, \quad \text { for } k \in\{2, \ldots, q-1\} \\ L_{q} & , \text { if } \theta_{q-1}<g(\mathcal{X})+\mathcal{E}\end{cases}
$$

where $\mathcal{E}$ is an error term following a random distribution that is uncorrelated with $\mathcal{X}$.

To specify the model fully, it is necessary to select a probability distribution for $\mathcal{Y}$ and, equivalently, for $\mathcal{E}$. We say that if the probability of observing a larger outcome $\mathcal{Y}$ increases slowly for small values of $\mathcal{X}$, more rapidly for intermediate values of $\mathcal{X}$, and more slowly for large values of $\mathcal{X}$, then either the normal distribution or logistic distribution is appropriate for $\mathcal{E}$. The former distribution yields the ordered probit model [203] and the latter distribution yeilds the proportional odds model [188. One reason for their popularity is their connection to the motivation of a continuous latent variable.

The difference in the overall results of both models is usually small, however, the interpretation of the proportional odds model is known to be more intuitive 204]. Thus, we assume that $\mathcal{E}$ follows a logistic distribution with 0 mean and that the latent variable function $g$ can be written as a linear function of the type $g(\mathcal{X})=\mathbf{w} \cdot \mathcal{X}$. The parameter vector $\mathbf{w}$ describes the effect of a unit change in vector $\mathbf{x}$ on the unobserved function $g$. This can be seen in Figure 8.1(a) that $\mathcal{Y}$ falls in the category $L_{k}$ when the latent variable falls in the $k$-th interval of values. Here, we assume $\mathcal{G}$ to be a latent random variable that follows a logistic distribution with a mean $g(\mathcal{X})$ for a given vector $\mathcal{X}=\mathbf{x}$.

The cumulative probability of observing an outcome $\mathcal{Y}$ smaller than or equal to $L_{k}$, given a variable vector $\mathcal{X}=\mathbf{x}$, is $\operatorname{Pr}\left(\mathcal{Y} \leq L_{k} \mid \mathcal{X}=\mathbf{x}\right)$. Given the relationship in Eq. 8.2, it follows

$$
\begin{equation*}
\operatorname{Pr}\left(\mathcal{Y} \leq L_{k} \mid \mathcal{X}=\mathbf{x}\right)=\operatorname{Pr}\left(g(\mathbf{x})+\mathcal{E} \leq \theta_{k}\right)=\operatorname{Pr}\left(\mathcal{E} \leq \theta_{k}-g(\mathbf{x})\right) \tag{8.3}
\end{equation*}
$$

To provide an estimate of the model parameters $\mathbf{w}$ and $\boldsymbol{\theta}=\left\{\theta_{1}, \ldots, \theta_{q-1}\right\}$, instead of fitting a decision rule $f: \mathscr{X} \rightarrow \mathscr{Y}$ directly, this model defines a probability density function over the ordinal labels for a given variable vector $\mathbf{x}$.

Using the cumulative distribution function of the logistic distribution, it can be seen that the probability of observing an outcome $\mathcal{Y}$ smaller than or equal to $L_{k}$, given $\mathcal{X}=\mathbf{x}$, is:


Figure 8.1: (a) A model underlying ordinal data for the case of 3 ordinal labels, the horizontal axis indicates the value of a (one-dimensional) feature vector $\mathbf{x}$, the vertical axes contain the random variable $\mathcal{Y}$ that falls in category $L_{k}$ when the latent variable $g$ falls in the $k$-th interval of values. The intervals are determined by the thresholds $\theta_{1}$ and $\theta_{2}$ on the latent variable. The latent random variable $\mathcal{G}$ is assumed to follow a logistic distribution with mean $g(\mathcal{X}=\mathbf{x})$. The highlighted areas indicate the probabilities for the case of $k=2$, such that $\operatorname{Pr}\left(\mathcal{Y}=L_{2} \mid \mathcal{X}=\mathbf{x}_{1}\right)$ and $\operatorname{Pr}\left(\mathcal{Y}=L_{2} \mid \mathcal{X}=\mathbf{x}_{2}\right)$. (b) A visualization of the proportional odds model with the latent variable on the horizontal axis and the cumulative probabilities on the vertical axis.

$$
\operatorname{Pr}\left(\mathcal{Y} \leq L_{k} \mid \mathcal{X}=\mathbf{x}\right)=\left\{\begin{array}{cl}
\frac{\exp \left(-\mathbf{w} \cdot \mathbf{x}+\theta_{k}\right)}{1+\exp \left(-\mathbf{w} \cdot \mathbf{x}+\theta_{k}\right)} & , \text { if } k=1, \ldots, q-1  \tag{8.4}\\
1 & , \text { if } k=q
\end{array}\right.
$$

This is visualised in Figure 8.1(b), where the probability of observing an outcome $\mathcal{Y}$ equal to $L_{k}$, given $\mathcal{X}=\mathbf{x}$, is simply the difference between the cumulative probability of that ordinal label and the one below it. This probability is formulated as follows:

$$
\operatorname{Pr}\left(\mathcal{Y}=L_{k} \mid \mathcal{X}=\mathbf{x}\right)=\left\{\begin{array}{cl}
\bar{f}_{k}(\mathbf{x}) & , \text { if } k=1  \tag{8.5}\\
\bar{f}_{k}(\mathbf{x})-\bar{f}_{k-1}(\mathbf{x}) & , \text { if } k=2, \ldots, q
\end{array}\right.
$$

where we used the short-hand notation $\bar{f}_{k}(\mathbf{x})=\operatorname{Pr}\left(\mathcal{Y} \leq L_{k} \mid \mathcal{X}=\mathbf{x}\right)$.
The parameter vectors wand $\boldsymbol{\theta}=\left(\theta_{1}, \ldots, \theta_{q-1}\right)$ are estimated by maximizing the likelihood function:

$$
\begin{equation*}
\ell(\mathbf{w}, \boldsymbol{\theta})=\prod_{i=1}^{n} \operatorname{Pr}\left(\mathcal{Y}=y_{i} \mid \mathcal{X}=\mathbf{x}_{i}\right) \tag{8.6}
\end{equation*}
$$

Because of tradition and computational convenience, the negative log-likelihood is minimized instead of maximizing the likelihood in Eq. 8.6) directly. There exist various minimization algorithms that can be used to perform the computation 94 , 205, 206. Minimizing the log-likelihood is also the approach we adopt in this chapter. Using the new variable $\bar{y}_{i, k}$, where $\bar{y}_{i, k}=1$ if $y_{i}=L_{k}$ and $\bar{y}_{i, k}=0$, otherwise, we compute the negative log-likelihood function as follows:

$$
\begin{align*}
-\log \ell(\mathbf{w}, \boldsymbol{\theta})=-\sum_{i=1}^{n} & \log \left(\bar{f}_{1}\left(\mathbf{x}_{i}\right)^{\bar{y}_{i, 1}}\right)  \tag{8.7}\\
& -\sum_{i=1}^{n} \sum_{k=2}^{q} \log \left(\left(\bar{f}_{k}\left(\mathbf{x}_{i}\right)-\bar{f}_{k-1}\left(\mathbf{x}_{i}\right)\right)^{\bar{y}_{i, k}}\right) .
\end{align*}
$$

As a final step, we minimize Eq. 8.7 w.r.t. w and $\boldsymbol{\theta}$ as follows:

$$
\begin{equation*}
\underset{\mathbf{w}, \boldsymbol{\theta}}{\operatorname{minimize}}-\log \ell(\mathbf{w}, \boldsymbol{\theta}) \tag{8.8}
\end{equation*}
$$

### 8.2.2. Performance measures for sensory evaluations

To evaluate the performance of an ordinal regression model, performance measures such as the concordance index (C-index) [207], or the volume under the ROC surface [208, 209] can be used. These measures use the predicted ordinal labels to compute the performance. However, as the proportional odds model is a probabilistic model, we can, given the estimates for $\mathbf{w}$ and $\boldsymbol{\theta}$, also estimate the probability distribution over the ordinal labels. Moreover, in the setting of sensory evaluation of food, it is common that a single food sample is evaluated by multiple panellists. As a result, these replicated evaluations can be used to compute the empirical distribution function of the panellists' evaluations. Thus, for each sample two distributions arise, one that is observed and one that is predicted (or modelled).

To compare probability distributions, one of several (dis)similarity measures or distance metrics described in literature (see for instance [210]) can be used. We distinguish two types of measures: measures that are invariant to a change of the ordering of the classes and measures that are not invariant to such a change. Examples of the former type include the Hellinger distance metric, the KullbackLeibler divergence (KL divergence) measure, the total variation distance metric and the $\chi^{2}$-distance, whereas examples of the latter type include the Kolmogorov distance metric, the Levy distance metric and the Earth mover's distance metric. Note that, for the aforementioned measures and distance metrics, the type trivially follows from their definition. For simplicity, we focus on the KL divergence measure in this chapter. Given two probability distribution functions $P$ and $\hat{P}$ over a
common sample space (in our case $\mathscr{Y}$ ), the KL divergence $d_{\mathrm{KL}}(P, \hat{P})$ is computed as:

$$
\begin{equation*}
d_{\mathrm{KL}}(P, \hat{P})=\sum_{L \in \mathscr{Y}} P(L) \log \frac{P(L)}{\hat{P}(L)} \tag{8.9}
\end{equation*}
$$

We consider the distribution $P$ to be the "test distribution" (usually considered to be the true distribution) and the distribution $\hat{P}$ to be the estimated distribution. A high value of $d_{\mathrm{KL}}(P, \hat{P})$ corresponds to a large separation of the two distributions, i.e., $\hat{P}$ is very dissimilar to $P$. Analogously, a low value of $d_{\mathrm{KL}}(P, \hat{P})$ corresponds to a better estimation. It must be noted that KL divergence is considered as a measure of entropy increase due to the use of an estimated distribution $\hat{P}$ to the true distribution $P$ rather than $P$ itself. From Eq 8.9), it is clear that the order on the classes does not influence the obtained KL divergence.

### 8.2.3. Regularization in ordinal regression

When estimating model parameters by maximizing the likelihood (minimizing the negative log-likelihood in our case), the resulting model might overfit the data and have a low predictive power. To counter that, $\ell_{1}$ or $\ell_{2}$ regularization is usually performed. Typically, $\ell_{1}$ regularization is more useful since it tends to produce sparse parameters (i.e., only few parameters are non-zero) by encouraging the sum of the absolute values of the parameters $\mathbf{w}$ to be small. Although $\ell_{2}$ regularization produces non-sparse parameters, it is computationally more efficient than $\ell_{1}$ regularization. However, in the setting where there is a limited amount of data available to learn an interpretable model, the advantages of $\ell_{1}$ regularization outweigh the computational efficiency of $\ell_{2}$ regularization. The former is done by adding a penalty term to the negative log-likelihood function, such that

$$
\begin{equation*}
-(1-\alpha) \log \ell(\mathbf{w}, \boldsymbol{\theta})+\alpha \sum_{i=1}^{p}\left|w_{i}\right| \tag{8.10}
\end{equation*}
$$

Note that the regularization parameter $\lambda>0$ is transformed into a parameter $\alpha$ to write the expression as a convex combination. The parameter $\alpha \in[0,1[$ controls the complexity of the model via a trade-off between fitting the data well, and having a well-regularized model with a reduced number of model parameters 1 .

The $\ell_{1}$-regularized likelihood function in Eq. 8.10 is a convex function. Therefore, it can potentially be minimized efficiently using convex-programming solvers. However, the sparsity-inducing regularization term is non-smooth. There exist

[^30]several strategies to solve optimization problems with sparsity-inducing penalties [105, 106, 107, 108. One of the advantages of the log-likelihood function is the possibility of easily including constraints on the model parameters. We choose to rewrite $\|\mathbf{w}\|_{1}$ as $\|\boldsymbol{\beta}\|_{1}$, such that $-\boldsymbol{\beta} \leq \mathbf{w} \leq \boldsymbol{\beta}$ for $\boldsymbol{\beta} \geq \mathbf{0}_{p}$, and, thus, transform Eq. 8.10) into the equivalent linearly constrained convex optimization problem:
\[

$$
\begin{array}{lll}
\underset{\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\beta}}{\operatorname{minimize}} & -(1-\alpha) \log \ell(\mathbf{w}, \boldsymbol{\theta}) & +\alpha \sum_{i=1}^{p} \beta_{i}  \tag{8.11}\\
\text { subject to } & 0 \leq \beta_{i}-w_{i} & , \text { for } i=1, \ldots, p \\
& 0 \leq \beta_{i}+w_{i} & , \text { for } i=1, \ldots, p
\end{array}
$$
\]

Note that the model parameters now consist of $\mathbf{w}, \boldsymbol{\theta}$ and $\boldsymbol{\beta}$.

### 8.3. Experimental analysis

This section is devoted to illustrate the proposed method using synthetic examples that are representative of typical problem settings. The goal is to investigate how the addition of an $\ell_{1}$-regularization term to the constrained optimization problem 8.11) influences the predictive performance of an ordinal regression model.

### 8.3.1. General experimental setup

It is to be expected that the performance of our ordinal regression model depends on several characteristics of the data and the underlying phenomenon, including:

- The number of samples $(n)$ provided to trained panellists.
- The number of labelling tasks $\left(n_{l}\right)$ performed by trained panellists, with one sample in each labelling task.
- The number of features $(p)$.

The combined influence of these characteristics on the overall predictive performance of the ordinal regression model is investigated using synthetically generated data. Note that, in this chapter, we do not focus on the number of trained panellists, but on the number of tasks. In other words, we do not distinguish between a trained panellist providing labels for all $n$ samples or $n$ trained panellists providing a label for one sample.

We demonstrate the influence of adding an $\ell_{1}$-regularization term to ordinal regression models. We do this by considering the setting where five labelling tests
are conducted, and, in each test, trained panellists are provided with $n$ samples to perform labelling tasks. These five tests are used as folds for applying 5 -fold cross-validation 100 and extracting the cross-validated performance of the model.

The general data simulation process for the experiment is as follows:
Generate the feature vectors (x): Feature vectors (for all $n$ samples evaluated by trained panellists in each of the five labelling tests) are independent and identically distributed observations of a Gaussian distribution with mean 0 and standard deviation 1. This can be seen as drawing observations from a $p$-variate Gaussian distribution with a covariance matrix equal to the identity matrix.

Defining the latent variable and thresholds: We consider different settings where the number of features $p$ is increased. The parameters in the vector $\mathbf{w}$ of the latent variable function $g$ are independent and identically distributed observations of a Gaussian distribution with mean 0 and standard deviation 1. This can be seen as drawing observations from a $p$-variate Gaussian distribution with a covariance matrix equal to the identity matrix. We consider a setting with five classes $(q=5)$. To obtain the threshold parameters $\left(\theta_{1}, \ldots, \theta_{4}\right)$, a large sample of the latent variable is simulated, i.e., a large number of feature vectors is generated for each vector $\mathbf{w}$ and only used to compute the latent variable ( $\mathbf{w} \cdot \mathbf{x}$ ). The thresholds are chosen such that the empirical distribution of the labels is uniform over the label set.

Sampling absolute evaluations (ordinal labels): In each of the five labelling tests, trained panellists are provided with $n$ samples to perform a total of $n_{l}$ labelling tasks. To simulate this, based on the proportional odds model, we compute, for each of the $n$ generated feature vectors, the probability mass function over the labels. To obtain labels from $n_{l}$ labelling tasks, we sample a feature vector from the set of $n$ feature vectors, $n_{l}$ times. For each sampled feature vector, we sample a label from the label set $\left(\mathscr{Y}=\left\{L_{1}, \ldots, L_{5}\right\}\right)$ according to the corresponding probability mass function. After sampling all the labels, we compute, for each feature vector, the empirical distribution function over the different labels.

Tuning and evaluating the model: By considering the optimization problem defined by Eq. 8.11), we learn a linearly constrained $\ell_{1}$-regularized ordinal regression model. To determine the optimal values of the tuning parameter $\alpha$ for solving Eq. (8.11) and to evaluate the performance of the model learned, we perform nested 5 -fold cross-validation. Nested cross-validation is used to avoid biased evaluations of performance that that result from using the same data to tune the model and evaluate its performance of the model [211]. First, we use inner cross-validation on $4 n$ feature vectors to determine the optimal values of the tuning parameter in the set $\left\{2^{-10}, 2^{-9}, \ldots, 2^{-1}, 2^{-0.9}, \ldots, 2^{-0.1}\right\}$ by selecting the tuning parameter that provides the best model. Second, we perform outer cross-validation to evaluate the model selected by the inner cross-validation on the remaining $n$ feature vectors in the labelling tests. We evaluate the performance in the five outer folds by
calculating the KL divergence defined by Eq. 8.9 between the predicted probability distribution using the learned model and the originally computed empirical probability distribution over the label set for each sample. We then compute the mean KL divergence to extract the cross-validated performance of the model. Note that the standard ordinal regression model does not require tuning, therefore, 5 -fold cross-validation is performed to assess its predictive power [100].

### 8.3.2. Experimental settings using synthetic data

We consider three settings in this experiment: in the first, trained panellists are given four samples $(n=4)$ to perform 40 labelling tasks $\left(n_{l}=40\right)$ in each of the five labelling tests; in the second, trained panellists are given five samples $(n=5)$ to perform 50 labelling tasks $\left(n_{l}=50\right)$ in each of the five labelling tests; and, in the third, trained panellists are given six samples $(n=6)$ to perform 60 labelling tasks $\left(n_{l}=60\right)$ in each of the five labelling tests. In each setting, 1000 vectors $\mathbf{w}$ are first randomly generated to compute a latent variable ( $\mathbf{w} \cdot \mathbf{x}$ ) for each $\mathbf{w}$ and the threshold parameters $\left(\theta_{1}, \ldots, \theta_{4}\right)$ are obtained. Then, $n$ feature vectors are generated (for all evaluated samples) and $n_{l}$ labels are sampled. Finally, we compute the empirical probability mass function over the label set $\left(\mathscr{Y}=\left\{L_{1}, \ldots, L_{5}\right\}\right)$ for each sample.

By considering the optimization problem defined by Eq. 8.8), we learn an ordinal regression model on $4 n$ feature vectors and $4 n_{l}$ labels gathered from trained panellists for each number of features $p \in\{10,20, \ldots, 100,150, \ldots, 300\}$. We then consider the optimization problem defined by Eq. 8.11 to learn a linearly constrained $\ell_{1}$-regularized ordinal regression model on the $4 n$ feature vectors and the $4 n_{l}$ labels gathered from trained panellists for each number of features in $p$. We then compute the KL divergence between the predicted probability distribution using the model learned on the sampled $4 n_{l}$ labels and the originally computed empirical probability distribution over the label set for each sample. We expect the overall performance of the model to improve after including an $\ell_{1}$-regularization term. Note that, for each setting, we expect the performance to diminish after increasing the number of features since the number of feature vectors is small in comparison.

Figure 8.2 shows the mean KL divergence (cross-validated performance) of the standard ordinal regression models and is considered as a reference. It is clear that $\ell_{1}$-regularized ordinal regression models outperform the standard ordinal regression models. In Figures 8.2 (a)-(c), we see that, as the number of features increases, the overall performance of the standard ordinal regression models diminishes, and, consequently, increasing the KL divergence. In comparison, the overall performance of the $\ell_{1}$-regularized ordinal regression models diminished less and, consequently, resulting in lower and somewhat steady KL divergence values. Note that the
variation in KL divergence values for the reference is much larger than that for the $\ell_{1}$-regularized ordinal regression models.

We now refer to the works of Murphy [212] and Gelman et al. [213] for detailed information on a Bayesian interpretation of regularization. In general, $\ell_{1}$-regularization naturally arises when the model coefficients have a Laplace distribution. It can be argued that fixing the model coefficients at a predefined value in the synthetic experiment, such that they have the same importance, deviates from this theoretically ideal situation in which they are sampled from a Laplacian distribution. However, we have two reasons why we still prefer to use $\ell_{1}$-regularization:

1. Using some form of regularization will improve the predictive accuracy of the predictive model as we have considered a setting with a large number of features ( 50 in the synthetic experiment) and a limited number of labelling tasks (no more than 60 for the synthetic experiment). This means that without using regularization the model will highly over-fit the data (there might not even be a unique solution to the optimization problem, or we would observe a perfect fit on the training data), which would lead to poor predictive power. The bias-variance trade-off, which is often used to motivate the need for regularization, does not assume that the importance of the elements in $\mathbf{w}$ is unequal. Other forms of regularization could be more applicable to the scenario that we study in the synthetic experiments where all inputs are equally important (such as for instance $\ell_{2}$ ). However, we do not want to focus on finding an optimal regularizer for this synthetic experiment (mainly because the effect of the type of regularization ( $\ell_{1}$ versus $\left.\ell_{2}\right)$ is often not that important in practice). Moreover, the translation of the $\ell_{1}$ regularizer into a set of linear inequality constraints in the optimisation problem closely resembles the way in which the ranking information is encoded in the optimisation problem (also by means of linear constraints).
2. In a real-life setting, the feature-selecting behaviour of $\ell_{1}$ regularization is a highly appreciated property. For instance, if a good model can be found that only uses a subset of the original number of VOCs this would reduce the efforts needed in the chemical lab. Therefore, using $\ell_{1}$ regularization is particularly interesting here.

### 8.4. Application to sensory data

In this section, we apply the method introduced in this chapter to the datasets gathered from labelling tests in Chapter 4 that measure the degree of freshness of chicken breasts, cod, brown shrimp and salmon samples. Using the sensory data, we illustrate the effect of adding $\ell_{1}$ regularization to ordinal regression models.

(c)

Figure 8.2: Experiment illustrating the mean KL divergence values of standard ordinal regression models as reference (blue line) with possible values (shaded blue) in comparison to the mean KL divergence values of $\ell_{1}$-regularized ordinal regression models (orange line) with possible values (shaded orange) for an increasing number of features $p$.

### 8.4.1. Chicken breast

First, the feature representations of chicken samples are obtained through the use of selected-ion flow-tube mass spectrometry (SIFT-MS) to measure concentrations of 23 VOCs. Table B. 2 summarizes the VOC profile (i.e., the feature vector) of each of the eight chicken samples. Subsequently, 33 panellists were each asked to give their appreciation of eight chicken breast samples (i.e., $n=8$ and $n_{l}=264$ ) by evaluating them, on an ordinal scale consisting of three labels: "Spoiled" (SP), "Marginal" (M), and "Fresh" (F) such that SP $\prec \mathrm{M} \prec \mathrm{F}$. The labels provided by the panellists are gathered in Table A.1, and, for each feature vector, we compute the empirical distribution function over the different labels.

In this case, the model is learned on only eight samples, which is insufficient to create large partitions for training and testing without losing significant training or testing capabilities. Here, nested leave-one-out cross-validation is performed using $n$ folds to determine the optimal value of the tuning parameter $\alpha$ for solving Eq. 8.11) and to estimate the general performance of the model. The resulting mean KL divergence of the standard ordinal regression models is 25.8102 , and the resulting mean KL divergence of the $\ell_{1}$-regularized ordinal regression models is 0.5602 . To interpret these KL divergence values, we compare them with the entropy of distribution $P=1.8684$. Thus, the standard ordinal regression model result in a $1381 \%$ entropy increase, while the $\ell_{1}$-regularized ordinal regression model result in a $30 \%$ entropy increase. The large increase in entropy is due to the very small dataset used in training the models.

It is important to consider a higher number of samples in the training dataset. Therefore, to ensure good performance of the models, it is recommended to gather more data, preferably from the same or similar storage experiments.

### 8.4.2. Atlantic cod

First, the feature representations of cod samples are obtained through the use of selected-ion flow-tube mass spectrometry (SIFT-MS) to measure concentrations of 20 VOCs. Table B. 5 summarizes the VOC profile (i.e., the feature vector) of each of the 32 cod samples. Subsequently, panellists (between 8 and 12) were each asked to give their appreciation of the cod samples (i.e., $n=32$ and $n_{l}=320$ ), as described in Table 4.5 by evaluating them, on an ordinal scale consisting of five labels: "Spoiled" (SP), "Marginal" (M), "Satisfactory" (S), "Fresh" (F) and Very Fresh (VF), such that $\mathrm{SP} \prec \mathrm{M} \prec \mathrm{S} \prec \mathrm{F} \prec \mathrm{VF}$. The labels provided by the panellists are used to compute the empirical distribution function over the different labels, for each feature vector. The results are shown in Table A.5.

In this case, the model is learned on only 32 samples, which is insufficient to create large partitions for training and testing without losing significant training
or testing capabilities. Here, nested leave-one-out cross-validation is performed using $n$ folds to determine the optimal value of the tuning parameter $\alpha$ for solving Eq. 8.11) and to estimate the general performance of the model. The resulting mean KL divergence of the standard ordinal regression models is 18.4988 , and the resulting mean KL divergence of the $\ell_{1}$-regularized ordinal regression models is 0.8129 . To interpret these KL divergence values, we compare them with the entropy of distribution $P=2.6834$. Thus, the standard ordinal regression model result in a $689 \%$ entropy increase, while the $\ell_{1}$-regularized ordinal regression model result in a $30.3 \%$ entropy increase. The large increase in entropy is due to the very small dataset used in training the models. However, it is clear that $\ell_{1}$ regularization reduced the amount of increased entropy.

### 8.4.3. Brown shrimp

First, the feature representations of shrimp samples are obtained through the use of selected-ion flow-tube mass spectrometry (SIFT-MS) to measure concentrations of 20 VOCs. Table B. 8 summarizes the VOC profile (i.e., the feature vector) of each of the 16 shrimp samples. Subsequently, nine or ten panellists were each asked to give their appreciation of the shrimp samples (i.e., $n=16$ and $n_{l}=147$ ), as described in Table 4.8 by evaluating them on an ordinal scale consisting of five labels: "Spoiled" (SP), "Marginal" (M), "Satisfactory" (S), "Fresh" (F) and Very Fresh (VF), such that SP $\prec \mathrm{M} \prec \mathrm{S} \prec \mathrm{F} \prec \mathrm{VF}$. The labels provided by the panellists are gathered in Table A.7, and, for each feature vector, we compute the empirical distribution function over the different labels.

In this case, the model is learned on only 16 samples, which is insufficient to create large partitions for training and testing without losing significant training or testing capabilities. Here, nested leave-one-out cross-validation is performed using $n$ folds to determine the optimal value of the tuning parameter $\alpha$ for solving Eq. 8.11) and to estimate the general performance of the model. The resulting mean KL divergence of the standard ordinal regression models is 26.3282 , and the resulting mean KL divergence of the $\ell_{1}$-regularized ordinal regression models is 2.5755. To interpret these KL divergence values, we compare them with the entropy of distribution $P=3.4648$. Thus, the standard ordinal regression model result in a $759 \%$ entropy increase, while the $\ell_{1}$-regularized ordinal regression model result in a $74 \%$ entropy increase. The large increase in entropy is due to the very small dataset used in training the models. However, it is clear that $\ell_{1}$ regularization reduced the amount of increased entropy.

### 8.4.4. Atlantic salmon

First, the feature representations of salmon samples are obtained through the use of selected-ion flow-tube mass spectrometry (SIFT-MS) to measure concentrations of 25 VOCs. Table B.13 summarizes the VOC profile (i.e., the feature vector) of each of the 72 salmon samples. Subsequently, panellists (between 5 and 12) were each asked to give their appreciation of the shrimp samples (i.e., $n=72$ and $n_{l}=748$ ), as described in Table 4.11, by evaluating them, on an ordinal scale consisting of five labels: "Spoiled" (SP), "Marginal" (M), "Satisfactory" (S), "Fresh" (F) and Very Fresh (VF), such that $\mathrm{SP} \prec \mathrm{M} \prec \mathrm{S} \prec \mathrm{F} \prec \mathrm{VF}$. The labels provided by the panellists are gathered in Table A.9 and, for each feature vector, we compute the empirical distribution function over the different labels.

In this case, the model is learned on 72 samples, which are quite sufficient to create large partitions for training and testing without losing significant training or testing capabilities. Here, nested 10 -fold cross-validation is performed using 10 folds to determine the optimal value of the tuning parameter $\alpha$ for solving Eq. 8.11) and to estimate the general performance of the model. The resulting mean KL divergence of the standard ordinal regression models is 16.6718 , and the resulting mean KL divergence of the $\ell_{1}$-regularized ordinal regression models is 0.9551 . To interpret these KL divergence values, we compare them with the entropy of distribution $P=2.982$. Thus, the standard ordinal regression model result in a $559 \%$ entropy increase, while the $\ell_{1}$-regularized ordinal regression model result in a $32 \%$ entropy increase. The large increase in entropy is due to the very small dataset used in training the models. However, it is clear that $\ell_{1}$ regularization reduced the amount of increased entropy.

It is important to note that not only would gathering larger amounts of data help the predictive performance of the models, but also the quality of the gathered data. We notice that the MAP conditions of the storage experiments vary a lot for each food type, particularly in the levels of CO2. Therefore, in light of these findings, in order to ensure good performance of the models, it is not only recommended to gather more data, but also from the same or similar storage experiments.

### 8.5. Conclusions

In this chapter, we have presented an ordinal regression strategy that allows to include $\ell_{1}$ regularization. This is done by rewriting the ordinal regression model as a constrained non-linear optimization problem that includes $\ell_{1}$ regularization in the form of constraints. The validity and application of the presented strategy has been shown through simulation and real-world experimentation on the sensory and instrumental data gathered in Chapter 4. Simulation studies show that the performance of standard ordinal regression models decreases when the models
are learned on a larger number of features. Moreover, these studies showed that $\ell_{1}$-regularized ordinal regression models are more robust to the increase in the number of features, and, thus, their performance did not decrease as much as that of the standard ordinal regression models.

Studies on the real-world data showed that the performance of the standard ordinal regression models learned on the sensory data was very poor. This is due to the fact that in each setting, the number of samples is small in comparison to the total number of features. The performance of the $\ell_{1}$-regularized ordinal regression models learned on the same data was improved. We conclude that this strategy is useful, especially for problems where the number of samples is very small in comparison to the total number of features. The demonstrated strategy ensures the sparsity of the feature vectors, and, thus, reducing overfitting and improving the model's predictive power. Moreover, results of the studies on the sensory data were consistent with the results of the synthetic studies.

It should be noted that our method is developed for settings where ordered labels are provided. For the setting where each untrained panellist expresses a ranking on the set of samples, it can be seen that most preference learning and learning to rank strategies also use a latent variable. We shed light on the use of a latent variable function in the next chapter (Chapter 9 ).

However, in other settings, different kinds of information can be provided [129, 130, 131. For instance, trained panellists and untrained panellists may provide different kinds of information, such as, absolute and relative information, respectively. These types of information can be combined in order to exploit the information expressed by both trained and untrained panellists. In Chapter 10 we introduce a strategy for combining ordinal labels and rankings to augment ordinal regression models.

## 9 Learning to predict rankings

## Table of Contents

9.1 Introduction<br>9.2 Approaches to predictive modelling of rankings 9.2.1 The pointwise approach 9.2.2 The pairwise approach<br>9.2.3 From preferences to rankings<br>9.2.4 Performance measures for preferences and rankings<br>9.3 Application to sensory data<br>9.3.1 Chicken breast<br>9.3.2 Atlantic cod<br>9.3.3 Brown shrimp<br>9.3.4 Atlantic salmon<br>9.4 Conclusions

### 9.1. Introduction

Over the last few decades, learning from preference data (data resulting from relative evaluations) has received a lot of attention, especially in the machine learning subfields: learning-to-rank [214, 215] and preference learning [216, 217, 218].

It can be seen that the learning-to-rank problem shares properties with the ordinal regression problem. The former focuses on predicting a relative order of samples while the latter focuses on predicting an ordinal label for each sample. Interestingly, it is possible for a ranking model (but not for an ordinal regression model) to predict wrong labels and incur no loss, as long as the relative order of the predicted labels is correct. Although learning-to-rank and ordinal regression problems are different problems, there has been some confusion resulting from the unclear distinction between both problems. For instance, some of the ordinal regression methods featured the word "ranking" [219, 220, and, similarly, some methods presented with the word "ordinal regression" in the title would be considered today ranking methods 221.

Learning-to-rank methods have been commonly used to solve information retrieval problems [215]. In this chapter, we introduce the most prominent approaches, namely, the pointwise and the pairwise approach, and we answer the following question:

### 9.2. Approaches to predictive modelling of rankings

As a starting point, we define the problem of constructing a predictive model to predict the ranking of packaged food samples in terms of freshness given the concentrations of the VOCs in the headspace of the package. To tackle the problem of predicting a ranking of samples, machine learning methods have been prominent in building effective ranking models [222. In particular, we focus on the pointwise and the pairwise approach. Note that these approaches model the process of learning-to-rank in different ways. They define different input and output spaces, employ different loss functions, and use different hypothesis spaces.

### 9.2.1. The pointwise approach

In general, the pointwise approach is similar to the ordinal regression approach. The input space $\mathscr{X} \subseteq \mathbb{R}^{p}$ of the pointwise approach contains vectors of $p$ features for samples and the output space $\mathscr{Y}=\left\{L_{1}, \ldots, L_{q}\right\}$ contains the ranking scores of the samples in a ranking. We assume that our training data are identically and independently drawn observations of an unnown distribution over $\mathscr{X} \times \mathscr{Y}$. The hypothesis space contains functions that take the feature vector of a sample as input and predict the ranking score of the sample in a ranking. We usually call such a function $f$ the appreciation function. Note that, based on the appreciation function, one can order all the samples and produce the final ranked list.

In literature, several algorithms have been proposed for predicting rankings 219, 223, 224. However, we focus on the method that has been previously discussed in Chapter 8, where the relationship between the output random variable $\mathcal{Y}$ and the input random variable $\mathcal{X}$ is defined in Eq. (8.2). We represent each sample by a $p$-dimensional vector $\mathbf{x} \in \mathscr{X}$ and a ranking score $y \in \mathscr{Y}$. We describe our dataset $\mathscr{D}$ as a set of $n$ couples $\left(\mathbf{x}_{i}, y_{i}\right)$ with $\mathbf{x}_{i}=\left(x_{i 1}, \ldots, x_{i p}\right)$, where the couples $\left(\mathbf{x}_{i}, y_{i}\right)$ are realizations of the random vector $(\mathcal{X}, \mathcal{Y})$.

By considering the optimization problem defined by Eq. (8.8), an ordinal regression model is learned on the couples $\left(\mathbf{x}_{i}, y_{i}\right)$. Similarly, by considering the optimization problem defined by Eq. 8.11) a linearly constrained $\ell_{1}$-regularized ordinal regression model is learned on the couples ( $\mathbf{x}_{i}, y_{i}$ ). Since the loss function is defined on the basis of single samples, the ranking score of each sample is predicted, and, consequently, a final ranked list is produced. Thus, the interdependency (i.e. the relative order) among these samples is naturally not considered in the learning process.

Furthermore, this approach ignores the fact that some samples are associated with the same experiment while some others are not. As a result, the ranking score of each sample in the ranked list is invisible to the loss function. This may cause the loss function to unconsciously overemphasize some of the samples that have the same ranking score but are relatively different, and, thereby, injecting bias into the loss function. Considering that sensory evaluation is mainly performed through multiple experiments, intuitively speaking, the pointwise approach has its limitations. So tackle the problem, an attempt has been made by considering pairs of samples associated with the same experiment as input. This is called the pairwise approach, where the relative order among samples can be better modelled.

### 9.2.2. The pairwise approach

Consider the set of samples $\mathscr{A}=\left\{a_{1}, \ldots, a_{n}\right\}$. In general, a ranking of $n$ samples corresponds to information for $\frac{n(n-1)}{2}$ pairs of samples, or equivalently, $n(n-1)$ couples of samples. For instance, the pair $\left\{\mathbf{x}_{1}, \mathbf{x}_{2}\right\}$ represents both couples $\left(\mathbf{x}_{1}, \mathbf{x}_{2}\right)$ and $\left(\mathbf{x}_{2}, \mathbf{x}_{1}\right)$. Since both couples provide the same information, we restrict to the couples where the index of the first sample is smaller than the index of the second sample, resulting in $n_{\mathscr{P}}=\frac{n(n-1)}{2}$ couples. We then order all these $n_{\mathscr{P}}$ couples lexicographically and identify each of them with an index $j \in\left\{1, \ldots, n_{\mathscr{P}}\right\}$. We denote by $\left(\mathbf{x}_{1}^{j}, \mathbf{x}_{2}^{j}\right)$ the feature vectors of the couples of samples, where $\mathbf{x}_{1}^{j} \in \mathscr{X}$ is the feature vector of the first sample of the $j$-th couple and $\mathbf{x}_{2}^{j} \in \mathscr{X}$ is the feature vector of the second sample of the $j$-th couple, for any $j \in\left\{1, \ldots, n_{\mathscr{P}}\right\}$.

In general, the input space $\mathscr{X}^{2}=\mathbb{R}^{p} \times \mathbb{R}^{p}$ of the pairwise approach contains couples of samples with feature vectors $\left(\mathrm{x}_{1}^{j}, \mathrm{x}_{2}^{j}\right) \in \mathscr{X}^{2}$. We describe our dataset $\mathscr{D} \subset X^{2} \times \mathscr{Y}$ as a set of $\frac{n(n-1)}{2}$ couples $\left(\mathrm{x}_{1}^{j}, \mathbf{x}_{2}^{j}\right)$, where the couples $\left(\mathrm{x}_{1}^{j}, \mathbf{x}_{2}^{j}\right)$ are realizations of the random vector $\left(\mathcal{X}_{1}, \mathcal{X}_{2}\right)$. Note that for any $j \neq k$, it holds that $\left\{\mathbf{x}_{1}^{j}, \mathbf{x}_{2}^{j}\right\} \neq\left\{\mathbf{x}_{1}^{k}, \mathbf{x}_{2}^{k}\right\}$.

Interestingly, in the pairwise approach, models that describe relative evaluations use a latent variable function $g$, as explained in the Chapter 8. Consider a first sample with feature vector $\mathbf{x}_{1} \in \mathscr{X}$ and a second sample with feature vector $\mathbf{x}_{2} \in \mathscr{X}$. Whenever $g\left(\mathbf{x}_{2}\right)<g\left(\mathbf{x}_{1}\right)$, it is predicted that the first sample is ranked higher than the second. Analogously, whenever $g\left(\mathbf{x}_{1}\right)<g\left(\mathbf{x}_{2}\right)$, it is predicted that the second sample is ranked higher than the first, and, whenever $g\left(\mathbf{x}_{1}\right)=g\left(\mathbf{x}_{2}\right)$, it is predicted that the first sample is ranked equal to the second.

Note that, in real life situations, the case $g\left(\mathbf{x}_{1}\right)=g\left(\mathbf{x}_{2}\right)$ is mathematically difficult to occur, unless the feature vectors $\mathbf{x}_{1}$ and $\mathbf{x}_{2}$ are identical. Therefore, in this chapter, we only consider strict rankings. As a result, the ordinal regression problem is reduced to a logistic regression problem, where the output is binary (i.e., either $\mathbf{x}_{1} \prec \mathbf{x}_{2}$ or $\mathbf{x}_{2} \prec \mathbf{x}_{1}$ ). Therefore, using the logistic distribution function, it can be seen that the probability of observing an outcome $\mathcal{X}_{1}$ preferred over $\mathcal{X}_{2}$, given
$\mathcal{X}_{1}=\mathbf{x}_{1}$ and $\mathcal{X}_{2}=\mathbf{x}_{2}$, is defined as follows:

$$
\begin{equation*}
\operatorname{Pr}\left(\mathcal{X}_{2} \prec \mathcal{X}_{1} \mid \mathcal{X}_{1}=\mathbf{x}_{1}, \mathcal{X}_{2}=\mathbf{x}_{2}\right)=\frac{\exp \left(\mathbf{w} \cdot\left(\mathbf{x}_{1}-\mathbf{x}_{2}\right)\right)}{1+\exp \left(\mathbf{w} \cdot\left(\mathbf{x}_{1}-\mathbf{x}_{2}\right)\right)} \tag{9.1}
\end{equation*}
$$

We use the notation $y_{j} \in\{0,1\}$ and $f\left(\mathbf{x}_{1}-\mathbf{x}_{2}\right)=\operatorname{Pr}\left(\mathcal{X}_{2} \prec \mathcal{X}_{1} \mid \mathcal{X}_{1}=\mathbf{x}_{1}, \mathcal{X}_{2}=\mathbf{x}_{2}\right)$, and we generalize Eq. 9.1) by computing the following likelihood function:

$$
\begin{equation*}
\ell\left(\mathbf{w}, \boldsymbol{\theta}, \mathcal{X}_{2} \prec \mathcal{X}_{1}\right)=\prod_{j=1}^{n_{\mathscr{F}}}\left(f\left(\mathbf{x}_{1}^{j}-\mathbf{x}_{2}^{j}\right)\right)^{y_{j}}\left(1-f\left(\mathbf{x}_{1}^{j}-\mathbf{x}_{2}^{j}\right)\right)^{1-y_{j}} \tag{9.2}
\end{equation*}
$$

For computational convenience, the negative log-likelihood is considered in Eq. 9.2 as the loss function. To follow the tradition of minimization in optimization, the loss function is minimized as follows:

$$
\begin{equation*}
\underset{\mathbf{w}}{\operatorname{minimize}}-\sum_{j=1}^{n \mathscr{A}}\left(y_{j} \log f\left(\mathbf{x}_{1}^{j}-\mathbf{x}_{2}^{j}\right)+\left(1-y_{j}\right) \log \left(1-f\left(\mathbf{x}_{1}^{j}-\mathbf{x}_{2}^{j}\right)\right)\right)+\lambda\|\mathbf{w}\|_{1} . \tag{9.3}
\end{equation*}
$$

### 9.2.3. From preferences to rankings

After predicting pairwise preferences for a set of samples, the next step is to derive an associated final ranking. This is non trivial, since the predicted preferences do not straighforwardly result in a unique ranking. In fact, this problem has received a lot of attention in the field of fuzzy preference modelling and (multiple criteria) decision making [225]. In the case of preference learning, several studies have compared different ways of combining the predictions of preferences into a ranking [226, 227].

The most common approach to deriving a ranking from preferences is a simple scoring approach. This approach, similar to the Borda count (Eq. 6.1)), makes use of a scoring function $S\left(\mathbf{x}_{i}\right)$ that assigns a score to the $i$-th sample $a_{i}$ with feature vector $\mathbf{x}_{i}$, and a ranking is derived as follows: $\left(S\left(\mathbf{x}_{i}\right) \leq S\left(\mathbf{x}_{j}\right)\right) \Rightarrow\left(a_{i} \precsim a_{j}\right)$, for $i \neq j$ and $i, j \in\{1, \ldots, n\}$.

The simplest scoring function is defined by the sum of weighted preferences, and is commonly used in pairwise classification and ranking [217. However, since we are dealing with probabilities, the scoring function $S\left(\mathbf{x}_{i}\right)$ or maximum likelihood function becomes difficult to compute. This is particularly due to the unknown conditional probability of a pairwise preference in a ranking given the other relevant pairwise preferences in the ranking. One way of solving this is by assuming that there are no conditional probabilities, thus, simplifying the scoring function $S\left(\mathbf{x}_{i}\right)$
into the product of the probabilities of preferences that contain the sample with feature vector $\mathbf{x}_{i}$.

Note that this is just a preliminary approach to derive a ranking from pairwise preferences.

### 9.2.4. Performance measures for preferences and rankings

To evaluate the performance of a ranking model, the Kendall (Kemeny for rankings with ties) distance between the predicted ranking and the consensus ranking is computed. Moreover, as the logistic regression model is a probabilistic model, we can, given the estimates for $\mathbf{w}$, also estimate the probability of a pairwise preference. Moreover, in the setting of sensory evaluation of food, it is common that a set of samples is ranked by multiple panellists. These rankings result in replicated preferences that can be used to compute the empirical probability distribution of the panellists' preferences. Thus, for each couple of samples, two probabilities arise, one that is observed and one that is predicted (modelled). These probabilities can be compared by computing their absolute differences, and the overall performance of the models is computed as the sum of these absolute differences.

In sensory evaluation of food, it is common that experiments include a different number of samples. Therefore, for each experiment, the Kendall distance and the sum of absolute differences are averaged based on the corresponding number of couples.

### 9.3. Application to sensory data

In this section, we apply the method introduced in this chapter to the datasets gathered from the ranking tests in Chapter 4 that order chicken breasts, cod, brown shrimp and salmon samples in terms of freshness. Using the sensory data, we illustrate the effect of adding $\ell_{1}$ regularization to ordinal regression models.

### 9.3.1. Chicken breast

First, the feature representations of chicken samples are obtained through the use of SIFT-MS to measure concentrations of 23 VOCs. Table B. 3 summarizes the VOC profile (i.e., the feature vector) of each of the chicken samples. Subsequently, a number of panellists (between 2 and 14) was asked to rank different sets of chicken breast samples as described in Tables A. 2 and A.3. As can be seen, there are eight sets of three samples in each set, eight sets of four samples in each set and two sets of five samples in each set. For each set of samples, we compute the distribution of the panellists' preferences on the couples of samples.

The model is learned on the computed distribution of the panellists' preferences on the couples of chicken samples. Here, nested $k$-fold cross validation is performed using $k$ folds, where each fold contains the corresponding number of couples for each experiment, to determine the optimal value of the tuning parameter $\alpha$ for solving Eq. (9.3). In addition, this procedure is performed to estimate the performance of the model in every fold by computing the averaged absolute difference of the predicted and observed preferences and the averaged Kendall distance between the predicted and observed rankings. The results are gathered in Table 9.1.

From Table 9.1, it is difficult to relate the sum of absolute differences of the probabilities of preferences to the Kendall distances between the predicted and consensus rankings. One of the main reasons is the fact that, in the case of experiments L4 and H4, a very small number (between two and four) of panellists provided rankings on the different sets of samples. However, note that the model was learned on the computed distribution of the panellists' preferences on the couples of samples in experiments L8 and H8, where 14 panellists provided rankings on the sets of samples. Therefore, we can extract some meaningful conclusions concerning the experiments where there were multiple consensus rankings. Interestingly, the results hint at an overall ranking of $a^{15} \prec a^{9} \prec a^{0}$ for group 2 in session 4 of experiment L4, an overall ranking of $a^{11} \prec a^{7} \prec a^{5} \prec a^{0}$ for group 1 in session 2 of experiment H4 and an overall ranking of $a^{15} \prec a^{9} \prec a^{7} \prec a^{0}$ for group 1 in session 3 of experiment H4.

In the case of experiments L8 and H8, the sum of absolute differences of the probabilities of preferences was mostly lower than that in the other experiments. Furthermore, since the consensus rankings of the samples in these experiments were based on a larger number of panellists, then we can conclude that the probability distributions of the preferences better resemble the overall preferences.

It is important to consider the number of the panellists who provided preferences on the couples of samples; the larger this number is, the higher quality the consensus rankings are for computing the Kendall distance. Furthermore, it is important to consider a higher number of couples of samples in the training dataset. Therefore, to ensure good performance of the models, it is recommended to gather more data, preferably from the same or similar storage experiments.

### 9.3.2. Atlantic cod

The feature representations of cod samples are obtained through the use of SIFTMS to measure concentrations of 20 VOCs. Table B. 6 summarizes the VOC profile (i.e., the feature vector) of each of the 32 cod samples. Subsequently, a number of panellists (between eight and ten) are asked to rank different sets of cod samples as described in Table A.6. As can be seen, there are eight sets of four samples in each set. For each set of samples, we compute the distribution of the panellists'

| Ranking test | Session | Group | Absolute difference | Predicted ranking | Consensus ranking | Kendall distance |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| L4 | 1 | 1 | 0.0879 | $a^{9} \prec a^{7} \prec a^{0} \prec a^{5}$ | $a^{9} \prec a^{5} \prec a^{7} \prec a^{0}$ | 0.333 |
|  |  | 2 | 0.0459 | $a^{5} \prec a^{0} \prec a^{7} \prec a^{11}$ | $a^{11} \prec a^{5} \prec a^{7} \prec a^{0}$ | 0.833 |
|  |  | 3 | 0.0157 | $a^{9} \prec a^{15} \prec a^{7} \prec a^{0}$ | $a^{9} \prec a^{15} \prec a^{0} \prec a^{7}$ | 0.167 |
|  |  | 4 | 0.0319 | $a^{13} \prec a^{11} \prec a^{7} \prec a^{5}$ | $a^{13} \prec a^{11} \prec a^{5} \prec a^{7}$ | 0.167 |
|  | 2 | 1 | 0.234 | $a^{13} \prec a^{11} \prec a^{15}$ | $a^{15} \prec a^{13} \prec a^{11}$ | 0.667 |
|  |  | 2 | 0.235 | $a^{13} \prec a^{15} \prec a^{9}$ | $\begin{aligned} & a^{15} \prec a^{13} \prec a^{9} \\ & a^{15} \prec a^{13} \prec a^{9} \end{aligned}$ | $\begin{aligned} & 0.667 \\ & 0.667 \end{aligned}$ |
|  |  | 3 | 0.0076 | $a^{13} \prec a^{11} \prec a^{5}$ | $a^{11} \prec a^{13} \prec a^{5}$ | 0.333 |
|  |  | 4 | 0.0230 | $a^{15} \prec a^{9} \prec a^{0}$ | $\begin{aligned} & a^{15} \prec a^{9} \prec a^{0} \\ & a^{9} \prec a^{15} \prec a^{0} \end{aligned}$ | $\begin{gathered} 0 \\ 0.333 \end{gathered}$ |
| H4 | 1 | 1 | 0.0351 | $a^{9} \prec a^{7} \prec a^{5} \prec a^{0}$ | $\begin{aligned} & a^{9} \prec a^{7} \prec a^{0} \prec a^{5} \\ & a^{9} \prec a^{0} \prec a^{7} \prec a^{5} \end{aligned}$ | $\begin{aligned} & 0.167 \\ & 0.333 \end{aligned}$ |
|  |  | 2 | 0.0215 | $a^{11} \prec a^{7} \prec a^{5} \prec a^{0}$ | $\begin{aligned} & a^{11} \prec a^{7} \prec a^{5} \prec a^{0} \\ & a^{11} \prec a^{7} \prec a^{0} \prec a^{5} \\ & a^{11} \prec a^{5} \prec a^{0} \prec a^{7} \\ & a^{11} \prec a^{5} \prec a^{7} \prec a^{0} \\ & a^{11} \prec a^{0} \prec a^{7} \prec a^{5} \end{aligned}$ | 0 0.167 0.333 0.167 0.333 |
|  |  | 3 | 0.0505 | $a^{15} \prec a^{9} \prec a^{7} \prec a^{0}$ | $\begin{aligned} & a^{15} \prec a^{9} \prec a^{7} \prec a^{0} \\ & a^{15} \prec a^{7} \prec a^{9} \prec a^{0} \\ & a^{15} \prec a^{7} \prec a^{0} \prec a^{9} \\ & a^{9} \prec a^{15} \prec a^{7} \prec a^{0} \\ & a^{9} \prec a^{7} \prec a^{15} \prec a^{0} \\ & a^{7} \prec a^{15} \prec a^{9} \prec a^{0} \\ & a^{7} \prec a^{15} \prec a^{0} \prec a^{9} \\ & a^{7} \prec a^{9} \prec a^{15} \prec a^{0} \end{aligned}$ | $\begin{gathered} 0 \\ 0.167 \\ 0.333 \\ 0.167 \\ 0.333 \\ 0.333 \\ 0.5 \\ 0.5 \end{gathered}$ |
|  |  | 4 | 0.0299 | $a^{13} \prec a^{11} \prec a^{7} \prec a^{5}$ | $a^{11} \prec a^{13} \prec a^{7} \prec a^{5}$ | 0.167 |
|  | 2 | 1 | 0.0567 | $a^{13} \prec a^{15} \prec a^{11}$ | $a^{11} \prec a^{13} \prec a^{15}$ | 0.667 |
|  |  | 2 | 0.0343 | $a^{13} \prec a^{15} \prec a^{9}$ | $a^{13} \prec a^{15} \prec a^{9}$ | 0 |
|  |  | 3 | 0.0533 | $a^{13} \prec a^{11} \prec a^{5}$ | $a^{11} \prec a^{13} \prec a^{5}$ | 0.333 |
|  |  | 4 | 0.0728 | $a^{15} \prec a^{9} \prec a^{0}$ | $a^{9} \prec a^{15} \prec a^{0}$ | 0.333 |
| L8 | - | - | 0.0196 | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ | $\begin{aligned} & a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0} \\ & a^{5} \prec a^{6} \prec a^{4} \prec a^{2} \prec a^{0} \end{aligned}$ | $\begin{gathered} 0 \\ 0.1 \end{gathered}$ |
| H8 | - | - | 0.0106 | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ | $\begin{aligned} & a^{6} \prec a^{5} \prec a^{4} \prec a^{0} \prec a^{2} \\ & a^{6} \prec a^{5} \prec a^{0} \prec a^{4} \prec a^{2} \\ & a^{6} \prec a^{5} \prec a^{0} \prec a^{2} \prec a^{4} \end{aligned}$ | $\begin{aligned} & 0.1 \\ & 0.3 \\ & 0.2 \end{aligned}$ |

Table 9.1: The averaged sum of absolute differences of the predicted and observed pairwise preferences and the averaged Kendall distances between the predicted and observed rankings of chicken samples.
preferences on the couples of samples.

The model is learned on the computed distribution of the panellists' preferences on the couples of cod samples. Here, nested $k$-fold cross validation is performed using $k$ folds, where each fold contains the corresponding number of couples for each
experiment, to determine the optimal value of the tuning parameter $\alpha$ for solving Eq. 9.3). In addition, this procedure is performed to estimate the performance of the model in every fold by computing the averaged absolute difference of the predicted and observed preferences and the averaged Kendall distance between the predicted and observed rankings. The results are gathered in Table 9.2 .

| Ranking test | Session | Absolute difference | Predicted ranking | Consensus ranking | Kendall distance |
| :---: | :---: | :---: | :---: | :---: | :---: |
| L4 | 1 | 0.0070 | $a^{13} \prec a^{8} \prec a^{4} \prec a^{0}$ | $a^{13} \prec a^{8} \prec a^{4} \prec a^{0}$ | 0 |
|  | 2 | 0.0273 | $a^{8} \prec a^{7} \prec a^{6} \prec a^{5}$ | $a^{8} \prec a^{7} \prec a^{6} \prec a^{5}$ | 0 |
| H4 | 1 | 0.0476 | $a^{13} \prec a^{8} \prec a^{6} \prec a^{4}$ | $a^{8} \prec a^{13} \prec a^{6} \prec a^{4}$ | 0.167 |
|  | 2 | 0.0384 | $a^{11} \prec a^{8} \prec a^{7} \prec a^{6}$ | $a^{11} \prec a^{8} \prec a^{7} \prec a^{6}$ | 0 |
| H8 | 1 | 0.0313 | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ | $\begin{aligned} & a^{7} \prec a^{5} \prec a^{3} \prec a^{0} \\ & a^{5} \prec a^{7} \prec a^{3} \prec a^{0} \end{aligned}$ | $\begin{gathered} 0 \\ 0.167 \end{gathered}$ |
|  | 2 | 0.0521 | $a^{5} \prec a^{3} \prec a^{6} \prec a^{4}$ | $\begin{aligned} & a^{5} \prec a^{6} \prec a^{4} \prec a^{3} \\ & a^{5} \prec a^{6} \prec a^{3} \prec a^{4} \end{aligned}$ | $\begin{aligned} & 0.333 \\ & 0.167 \end{aligned}$ |
| L8 | - | 0.0311 | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ | 0 |
| A4 | - | 0.0570 | $a^{0} \prec a^{1} \prec a^{2} \prec a^{3}$ | $a^{3} \prec a^{2} \prec a^{0} \prec a^{1}$ | 0.833 |

Table 9.2: The averaged sum of absolute differences of the predicted and observed pairwise preferences and the averaged Kendall distances between the predicted and observed rankings of cod samples.

It can be seen from Table 9.2 that the sum of absolute differences of the probabilities of the preferences is positively correlated to the Kendall distances between the predicted and consensus rankings. Therefore, we can conclude that the model was able to predict most of the actual consensus rankings, particularly for storage experiments L4, H4, L8 and H8 (session 1). Interestingly, we notice that for experiment A4, the predicted and consensus ranking were different from each other. This can be explained by the fact that the model was learned on mostly samples stored for at least three days, while the predicted ranking in experiment A4 was of samples stored for at most 3 days.

### 9.3.3. Brown shrimp

The feature representations of shrimp samples are obtained through the use of SIFT-MS to measure concentrations of 20 VOCs. Table B. 9 summarizes the VOC profile (i.e., the feature vector) of each of the 16 shrimp samples. Subsequently, a number of panellists (nine or ten) are asked to rank different sets of shrimp samples as described in Table A.8. As can be seen, there are four sets of four samples in each set. for each set of samples, we compute the distribution of the panellists' preferences on the couples of samples.

The model is learned on the computed distribution of the panellists' preferences on the couples of shrimp samples. Here, nested $k$-fold cross validation is performed using $k$ folds, where each fold contains the corresponding number of couples for each experiment, to determine the optimal value of the tuning parameter $\alpha$ for solving Eq. 9.3). In addition, this procedure is performed to estimate the performance of the model in every fold by computing the averaged absolute difference of the predicted and observed preferences and the averaged Kendall distance between the predicted and observed rankings. The results are gathered in Table 9.3 .

| Ranking <br> test | Session | Absolute <br> difference | Predicted <br> ranking | Consensus <br> ranking | Kendall <br> distance |
| :---: | :---: | :---: | :---: | :---: | :---: |
| L4 | 1 | 0.0284 | $a^{10} \prec a^{5} \prec a^{0} \prec a^{3}$ | $a^{10} \prec a^{5} \prec a^{0} \prec a^{3}$ | 0 |
|  | 2 | 0.0187 | $a^{10} \prec a^{0} \prec a^{5} \prec a^{3}$ | $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ | 0.333 |
| H 4 | 1 | 0.0276 | $a^{12} \prec a^{7} \prec a^{3} \prec a^{0}$ | $a^{12} \prec a^{7} \prec a^{3} \prec a^{0}$ | 0 |
|  | 2 | 0.0191 | $a^{7} \prec a^{5} \prec a^{0} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ | 0.167 |

Table 9.3: The averaged sum of absolute differences of the predicted and observed pairwise preferences and the averaged Kendall distances between the predicted and observed rankings of shrimp samples.

From Table 9.3, it is difficult to relate the sum of absolute differences of the probabilities of the preferences to the Kendall distances between the predicted and consensus rankings. One reason could be due to the small dataset of 24 couples. Another reason could be the fact that, in the second session of the experiments, some of the observed probabilities of preferences were close to 0.5 (i.e., close to half of the panellists had the opposite preference than that of the other panellists). Thus, even though most of the predicted probabilities were close to the observed probabilities of preferences, a small difference could still result in an opposite preference. Therefore, the Kendall distance between the predicted and the consensus rankings was larger than that of the other experiments.

### 9.3.4. Atlantic salmon

The feature representations of salmon samples are obtained through the use of SIFT-MS to measure concentrations of 25 VOCs. Table B. 11 summarizes the VOC profile (i.e., the feature vector) of each of the 72 salmon samples. Subsequently, a number of panellists (between eight and ten) are asked to rank different sets of shrimp samples as described in Table A.10. As can be seen, there are 18 sets of four samples in each set. for each set of samples, we compute the distribution of the panellists' preferences on the couples of samples.

The model is learned on the computed distribution of the panellists' preferences on the couples of salmon samples. Here, nested $k$-fold cross validation is performed
using $k$ folds, where each fold contains the corresponding number of couples for each experiment, to determine the optimal value of the tuning parameter $\alpha$ for solving Eq. (9.3). In addition, this procedure is performed to estimate the performance of the model in every fold by computing the averaged absolute difference of the predicted and observed preferences and the averaged Kendall distance between the predicted and observed rankings. The results are gathered in Table 9.4

| Ranking test | Session | Absolute difference | Predicted ranking | Consensus ranking | Kendall distance |
| :---: | :---: | :---: | :---: | :---: | :---: |
| H4 | 1 | 0.1261 | $a^{5} \prec a^{1} \prec a^{9} \prec a^{11}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | 0.833 |
|  | 2 | 0.0535 | $a^{7} \prec a^{1} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | 0.167 |
|  | 3 | 0.0282 | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | 0 |
|  | 4 | 0.0566 | $a^{7} \prec a^{5} \prec a^{9} \prec a^{3}$ | $a^{9} \prec a^{7} \prec a^{5} \prec a^{3}$ | 0.333 |
| AN4 | 1 | 0.0254 | $a^{13} \prec a^{9} \prec a^{5} \prec a^{1}$ | $\begin{aligned} & a^{13} \prec a^{9} \prec a^{5} \prec a^{1} \\ & a^{9} \prec a^{13} \prec a^{5} \prec a^{1} \end{aligned}$ | $\begin{gathered} \hline 0 \\ 0.167 \end{gathered}$ |
|  | 2 | 0.0185 | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ | 0 |
|  | 3 | 0.0413 | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | 0 |
|  | 4 | 0.0117 | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | 0.167 |
| ANH4 | 1 | 0.0160 | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | 0 |
|  | 2 | 0.0300 | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ | $a^{7} \prec a^{3} \prec a^{1} \prec a^{5}$ | 0.333 |
|  | 3 | 0.0367 | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $\begin{aligned} & a^{11} \prec a^{9} \prec a^{5} \prec a^{1} \\ & a^{11} \prec a^{5} \prec a^{9} \prec a^{1} \\ & a^{11} \prec a^{5} \prec a^{1} \prec a^{9} \end{aligned}$ | $\begin{gathered} 0 \\ 0.167 \\ 0.333 \end{gathered}$ |
|  | 4 | 0.0302 | $a^{7} \prec a^{3} \prec a^{5} \prec a^{1}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | 0.333 |
| A4 | 1 | 0.0220 | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | 0.167 |
|  | 2 | 0.0095 | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ | 0 |
|  | 3 | 0.0291 | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ | 0.167 |
|  | 4 | 0.0248 | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ | 0.167 |
| L4 | - | 0.0273 | $a^{11} \prec a^{7} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{7} \prec a^{5} \prec a^{1}$ | 0 |
| M4 | - | 0.0271 | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | 0 |

Table 9.4: The averaged sum of absolute differences of the predicted and observed pairwise preferences and the averaged Kendall distances between the predicted and observed rankings of salmon samples.

It can be seen from Table 9.4 that the sum of absolute differences of the probabilities of the preferences is positively correlated to the Kendall distance between the predicted and consensus rankings. Therefore, we can conclude that the model was able to predict most of the actual consensus rankings, particularly for storage experiments AN4, A4, L4 and M4. However, we notice some predictive errors (i.e.,
large Kendall distance and absolute difference), particularly for experiments H 4 and ANH4.

As was previously mentioned, it is important to gather a large number of experiments, particularly a large number of couples of samples. Even though the salmon dataset is large ( 108 couples of samples), it seems that the performance of the models was affected by the quality of the data. We notice that the MAP conditions of the storage experiments vary a lot, particularly in the levels of CO2. Therefore, in light of these findings, in order to ensure good performance of the models, it is recommended to gather more data, preferably from the same or similar storage experiments.

### 9.4. Conclusions

In this chapter, we have briefly described two approaches for the prediction of rankings. Learning-to-rank problems based on the pointwise approach share properties with ordinal regression problems. However, the relative order among the samples is not well-modelled. The pairwise approach tackles this limitation by considering information on pairs of samples associated with the same experiment as inputs. Based on the pairwise approach, we have illustrated an $\ell_{1}$-regularized logistic regression strategy that allows to include as input the difference of the feature vectors of the samples in each pair. The application of the presented strategy has been shown through real-world experimentation on the sensory and instrumental data gathered in Chapter 4. This strategy is useful, especially for problems where the number of samples is very small in comparison to the total number of features. The demonstrated strategy exploits sparsity-inducing penalties, and, thus, reducing overfitting and improving the model's predictive power.

It should be noted that the pairwise approach is used to predict probabilities of pairwise preferences. We have presented a preliminary approach to derive rankings from the predicted pairwise preferences. However, it would be interesting to expand other approaches that could potentially result in improved prediction of rankings.

In the next chapter (Chapter 10), we borrow ideas from the pairwise approach to integrate rankings in ordinal regression models. We introduce a strategy for combining ordinal labels and rankings to augment ordinal regression models.

# 10 Integrating rankings in ordinal regression models 

## Table of Contents

10.1 Introduction<br>10.2 The ordinal regression problem<br>10.3 Incorporating relative evaluations<br>10.3.1 Rankings<br>10.3.2 Constraints<br>10.4 Experimental analysis<br>10.4.1 General experimental setup<br>10.4.2 Experimental settings using synthetic data<br>10.5 Application to sensory data<br>10.5.1 Chicken breast<br>10.5.2 Atlantic cod<br>10.5.3 Atlantic brown shrimp<br>10.5.4 Atlantic salmon<br>10.6 Conclusions

### 10.1. Introduction

In this chapter, we consider the problem defined in Chapter 8 of predicting the appreciation of a sample when there is a limited amount of data available. As a result, learning a good statistical model becomes a difficult task. Therefore, any additional source of information could potentially lead to learning an improved model. One way of obtaining additional data is by using more panellists to evaluate the samples. As the use of additional trained panellists is too expensive, untrained panellists could provide a cost-efficient source of additional information. Although untrained panellists may not be well equipped to make accurate absolute evaluations, they can instead provide reliable information by comparing multiple samples and ranking them according to their perceived freshness [228]. The evaluations obtained in such a way are relative (as opposed to absolute); therefore, we call them relative evaluations. A simple example of relative evaluation arises when untrained panellists are asked to compare only two samples at a time.

As we have seen in the previous chapters, learning from absolute evaluation data and learning from relative evaluation data are usually considered as two separate
problems. It might be worthwhile to develop strategies to exploit these different types of data simultaneously into a single learning problem.

The purpose of this chapter is to develop a learning strategy that can be used to predict absolute evaluations. To achieve this goal, we use traditional ordinal regression models, but augment them by exploiting information obtained from the relative evaluations provided by untrained panellists. Notably, numerous models that describe absolute evaluations (in case of ordinal regression) or relative evaluations (in case of preference learning) use a latent variable, as explained in the previous chapters. Therefore, in this chapter, we answer the following question:

Question IV.3: How can we combine ordinal labels and rankings to improve the prediction of an ordinal label?

The path we follow is similar to the one proposed by Ye and Doermann 229], who developed an approach to integrate data resulting from absolute and relative evaluations using a probabilistic model. Similar to our work, Ye and Doermann assumed that both types of data stem from the same ordinal principle. However, their goal was not to construct a predictive model, and they did not use a feature representation of the objects. It should be noted that there exist other, nonprobabilistic, approaches to integrate these different types of data [129, 130, 131, 230. However, we will not follow this line of research in this paper as it mainly focuses on reaching a consensus in decision-making problems rather than inferring an absolute evaluation.

We present a framework that is capable of learning an ordinal regression model while combining two types of data: firstly, data that represent absolute evaluations, namely ordinal data, and secondly, data that represent relative evaluations, namely rankings. In what follows, we propose and solve a linearly constrained convex optimization problem that guarantees that the learned mapping takes both types of data into account, and is capable of attributing an ordinal label to a new sample based on its features. We do this by relying on principles from machine learning and optimization theory, combined with ideas from information fusion.

### 10.2. The ordinal regression problem

In this section, we introduce some notations and briefly review the methods that constitute the main building blocks of our approach.

As a starting point of this section, recall the problem setting presented in the previous chapter, where a predictive model needs to be constructed to predict the freshness of packaged foods given the concentrations of the VOCs in the headspace of the package. On one hand, the output space $\mathscr{Y}=\left\{L_{1}, \ldots, L_{q}\right\}$ is the label set
containing $q$ linearly ordered labels. We represent each sample by a $p$-dimensional vector $\mathbf{x} \in \mathscr{X}$ and an ordinal label $y \in \mathscr{Y}$. We describe our dataset $\mathscr{D}$ as a set of $n$ couples ( $\mathbf{x}_{i}, y_{i}$ ) with $\mathbf{x}_{i}=\left(x_{i 1}, \ldots, x_{i p}\right)$, where the couples ( $\mathbf{x}_{i}, y_{i}$ ) are realizations of the random vector $(\mathcal{X}, \mathcal{Y})$.

### 10.3. Incorporating relative evaluations

So far, we have described a traditional ordinal regression model with $\ell_{1}$ regularization. This model, as seen in Eq. 8.11, can be used to predict absolute evaluations. Due to limitations in gathering large amounts of data to learn a good model, any additional source of information could potentially lead to an improved model. Therefore, we develop hereafter a learning strategy to augment ordinal regression models by exploiting information obtained from relative evaluations provided by untrained panellists.

### 10.3.1. Rankings

As a starting point, recall the problem of setting presented in the previous chapter, where a number of untrained panellists compare $n$ samples, resulting in a ranking (with ties) for each untrained panellist. We denote by $\left(\mathbf{x}_{1}^{j}, \mathbf{x}_{2}^{j}\right)$ the feature vectors of the couples of samples, where $\mathbf{x}_{1}^{j} \in \mathscr{X}$ is the feature vector of the first sample of the $j$-th couple and $\mathbf{x}_{2}^{j} \in \mathscr{X}$ is the feature vector of the second sample of the $j$-th couple, for any $j \in\left\{1, \ldots, n_{\mathscr{P}}\right\}$. Subsequently, we gather all $n_{\mathscr{P}}$ couples in the following set:

$$
\begin{equation*}
\mathscr{P}=\left\{\left(\left(\mathbf{x}_{1}^{j}, \mathbf{x}_{2}^{j}\right),\left(\pi_{j}, \pi_{j}^{\prime}, \mu_{j}\right)\right) \mid j \in\left\{1, \ldots, n_{\mathscr{P}}\right\}\right\}, \tag{10.1}
\end{equation*}
$$

where $\pi_{j}$ (respectively $\pi_{j}^{\prime}$ ) is the number of untrained panellists ranking $\mathbf{x}_{1}^{j}$ higher than $\mathbf{x}_{2}^{j}$ (respectively $\mathbf{x}_{2}^{j}$ higher than $\mathbf{x}_{1}^{j}$ ) and $\mu_{j}$ is the number of untrained panellists expressing a tie between $\mathbf{x}_{1}^{j}$ and $\mathbf{x}_{2}^{j}$. Note that for any $j \neq k$, it holds that $\left\{\mathbf{x}_{1}^{j}, \mathbf{x}_{2}^{j}\right\} \neq\left\{\mathbf{x}_{1}^{k}, \mathbf{x}_{2}^{k}\right\}$.
Example 10.1. Consider four samples with feature vectors $\mathbf{x}_{1}, \mathbf{x}_{2}, \mathbf{x}_{3}, \mathbf{x}_{4} \in \mathscr{X}$. Hence, the lexicographic order $\left(\prec_{\text {lex }}\right)$ on the set of $n_{\mathscr{P}}=6$ couples is

$$
\left(\mathbf{x}_{1}, \mathbf{x}_{2}\right) \prec_{l e x}\left(\mathbf{x}_{1}, \mathbf{x}_{3}\right) \prec_{l e x}\left(\mathbf{x}_{1}, \mathbf{x}_{4}\right) \prec_{l e x}\left(\mathbf{x}_{2}, \mathbf{x}_{3}\right) \prec_{l e x}\left(\mathbf{x}_{2}, \mathbf{x}_{4}\right) \prec_{l e x}\left(\mathbf{x}_{3}, \mathbf{x}_{4}\right) .
$$

These four samples are compared by an untrained panellist, resulting in a ranking (with ties), say $\mathbf{x}_{4}<\mathbf{x}_{1} \sim \mathbf{x}_{3}<\mathbf{x}_{2}$, expressing that this untrained panellist prefers $\mathbf{x}_{2}$ over $\mathbf{x}_{3}$, is indifferent between $\mathbf{x}_{3}$ and $\mathbf{x}_{1}$, and prefers $\mathbf{x}_{1}$ over $\mathbf{x}_{4}$. Due to the inherent transitivity of a ranking, this ranking also expresses that this untrained
panellist prefers $\mathbf{x}_{2}$ over $\mathbf{x}_{1}, \mathbf{x}_{2}$ over $\mathbf{x}_{4}$, and $\mathbf{x}_{3}$ over $\mathbf{x}_{4}$. This information is gathered as follows:

$$
\begin{aligned}
\mathscr{P}=\{ & \left(\left(\mathbf{x}_{1}, \mathbf{x}_{2}\right),(0,1,0)\right),\left(\left(\mathbf{x}_{1}, \mathbf{x}_{3}\right),(0,0,1)\right),\left(\left(\mathbf{x}_{1}, \mathbf{x}_{4}\right),(1,0,0)\right) \\
& \left.\left(\left(\mathbf{x}_{2}, \mathbf{x}_{3}\right),(1,0,0)\right),\left(\left(\mathbf{x}_{2}, \mathbf{x}_{4}\right),(1,0,0)\right),\left(\left(\mathbf{x}_{3}, \mathbf{x}_{4}\right),(1,0,0)\right)\right\} .
\end{aligned}
$$

Typically, not all untrained panellists rank all the samples. Actually, untrained panellists may rank different subsets of the set of samples. As a result, for every couple of samples $\left(\mathbf{x}_{1}^{j}, \mathbf{x}_{2}^{j}\right)$, we have $\pi_{j}+\pi_{j}^{\prime}+\mu_{j}=m_{j}$, where $m_{j}$ is the number of untrained panellists that ranked samples $\mathbf{x}_{1}^{j}$ and $\mathbf{x}_{2}^{j}$. Note that $m_{j}$ might be equal to zero.

Example 10.2. Consider four samples with feature vectors $\mathbf{x}_{1}, \mathbf{x}_{2}, \mathbf{x}_{3}, \mathbf{x}_{4} \in \mathscr{X}$. Three untrained panellists are asked to rank three of the four samples, resulting in three rankings (with ties), say $\mathbf{x}_{1}>\mathbf{x}_{3} \sim \mathbf{x}_{4}, \mathbf{x}_{3}>\mathbf{x}_{1}>\mathbf{x}_{2}$, and $\mathbf{x}_{1} \sim \mathbf{x}_{4}>\mathbf{x}_{3}$ This information is gathered as follows:

$$
\begin{aligned}
\mathscr{P}=\{ & \left(\left(\mathbf{x}_{1}, \mathbf{x}_{2}\right),(1,0,0)\right),\left(\left(\mathbf{x}_{1}, \mathbf{x}_{3}\right),(2,1,0)\right),\left(\left(\mathbf{x}_{1}, \mathbf{x}_{4}\right),(1,0,1)\right) \\
& \left.\left(\left(\mathbf{x}_{2}, \mathbf{x}_{3}\right),(0,1,0)\right),\left(\left(\mathbf{x}_{2}, \mathbf{x}_{4}\right),(0,0,0)\right),\left(\left(\mathbf{x}_{3}, \mathbf{x}_{4}\right),(0,1,1)\right)\right\} .
\end{aligned}
$$

### 10.3.2. Constraints

We now consider a proportional odds model with a latent variable function $g: X \rightarrow \mathbb{R}$. We can use this model to predict the preference of one sample, with feature vector $\mathbf{x}_{1}^{j} \in \mathscr{X}$, over another, with feature vector $\mathbf{x}_{2}^{j} \in \mathscr{X}$, for any $j \in\left\{1, \ldots, n_{\mathscr{P}}\right\}$. If $g\left(\mathbf{x}_{1}^{j}\right)>g\left(\mathbf{x}_{2}^{j}\right)$, then we predict that $\mathbf{x}_{1}^{j}$ is preferred over $\mathbf{x}_{2}^{j}$; if $g\left(\mathbf{x}_{1}^{j}\right)<g\left(\mathbf{x}_{2}^{j}\right)$, then we predict that $\mathbf{x}_{2}^{j}$ is preferred over $\mathbf{x}_{1}^{j}$; finally, if $g\left(\mathbf{x}_{1}^{j}\right)=g\left(\mathbf{x}_{2}^{j}\right)$, then we predict that $\mathbf{x}_{1}^{j}$ and $\mathbf{x}_{2}^{j}$ are indifferent. Clearly, a useful model should (to some extent) respect the information in Eq. 10.1). This imposes a number of constraints on the potential latent variable functions.

If $\pi_{j}>0$, then we consider the constraint $g\left(\mathbf{x}_{1}^{j}\right)>g\left(\mathbf{x}_{2}^{j}\right)$. Since $g$ is linear, i.e., $g(\mathbf{x})=\mathbf{w}^{\top} \mathbf{x}$, we obtain $\mathbf{w}^{\top} \mathbf{x}_{1}^{j}>\mathbf{w}^{\top} \mathbf{x}_{2}^{j}$, or, equivalently,

$$
\begin{equation*}
\mathbf{w}^{\top}\left(\mathbf{x}_{2}^{j}-\mathbf{x}_{1}^{j}\right)<0 . \tag{10.2}
\end{equation*}
$$

However, due to the complexity of the problem or noise in the data, such inequality might be too strict for a linear model. For this reason, we relax the constraint by
introducing a slack variable $\xi_{j} \in \mathbb{R}^{+}$and require

$$
\begin{equation*}
\mathbf{w}^{\top}\left(\mathbf{x}_{2}^{j}-\mathbf{x}_{1}^{j}\right) \leq \xi_{j} . \tag{10.3}
\end{equation*}
$$

Similarly, if $\pi_{j}^{\prime}>0$, then we consider the constraint $g\left(\mathbf{x}_{2}^{j}\right)>g\left(\mathbf{x}_{1}^{j}\right)$ and relax it by introducing a slack variable $\xi_{j}^{\prime} \in \mathbb{R}^{+}$and require

$$
\begin{equation*}
\mathbf{w}^{\top}\left(\mathbf{x}_{1}^{j}-\mathbf{x}_{2}^{j}\right) \leq \xi_{j}^{\prime} . \tag{10.4}
\end{equation*}
$$

Finally, if $\mu_{j}>0$, then we consider the constraint $g\left(\mathbf{x}_{1}^{j}\right)=g\left(\mathbf{x}_{2}^{j}\right)$ and relax it by introducing a slack variable $\zeta_{j} \in \mathbb{R}^{+}$and require

$$
\begin{equation*}
-\zeta_{j} \leq \mathbf{w}^{\top}\left(\mathbf{x}_{1}^{j}-\mathbf{x}_{2}^{j}\right) \leq \zeta_{j} \tag{10.5}
\end{equation*}
$$

The constrained optimization problem defined by Eq. 8.11 can now be extended to include constraints defined in Eq. 10.3), Eq. 10.4, and Eq. 10.5) as follows

$$
\begin{array}{lll}
\underset{\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\xi}^{\prime} \boldsymbol{\zeta}}{\operatorname{minimize}} & -\left(1-\alpha_{1}\right) \log \mathcal{L}(\mathbf{w}, \boldsymbol{\theta}) & \\
& +\alpha_{1}\left(\left(1-\alpha_{2}\right) \sum_{i=1}^{p} \beta_{i}\right. & \left.+\alpha_{2}\left(\sum_{j=1}^{n_{\mathscr{P}}} \pi_{j} \xi_{j}+\pi_{j}^{\prime} \xi_{j}^{\prime}+\mu_{j} \zeta_{j}\right)\right), \\
\text { subject to } & 0 \leq \beta_{i}-w_{i} & , \text { for } i=1, \ldots, p, \\
0 \leq \beta_{i}+w_{i} & , \text { for } i=1, \ldots, p, \\
0 \leq \xi_{j}-\mathbf{w}^{\top}\left(\mathbf{x}_{2}^{j}-\mathbf{x}_{1}^{j}\right) & , \text { for } j=1, \ldots, n_{\mathscr{P}},  \tag{10.6}\\
0 \leq \xi_{j}^{\prime}-\mathbf{w}^{\top}\left(\mathbf{x}_{1}^{j}-\mathbf{x}_{2}^{j}\right) & , \text { for } j=1, \ldots, n_{\mathscr{P}}, \\
0 \leq \zeta_{j}-\mathbf{w}^{\top}\left(\mathbf{x}_{2}^{j}-\mathbf{x}_{1}^{j}\right) & , \text { for } j=1, \ldots, n_{\mathscr{P}}, \\
0 \leq \zeta_{j}-\mathbf{w}^{\top}\left(\mathbf{x}_{1}^{j}-\mathbf{x}_{2}^{j}\right) & , \text { for } j=1, \ldots, n_{\mathscr{P}}, \\
0 \leq \xi_{j} & , \text { for } j=1, \ldots, n_{\mathscr{P}}, \\
0 \leq \xi_{j}^{\prime} & , \text { for } j=1, \ldots, n_{\mathscr{P}},
\end{array}
$$

The parameters $\left.\alpha_{1}, \alpha_{2} \in\right] 0,1[$ control the complexity of the model via a trade-off between fitting the data well, and having a well-regularized model with a reduced number of model parameters and additional ranking information.

Note that the objective function takes into account the multiplicity of the constraints by weighing the slack variables $\xi_{j}, \xi_{j}^{\prime}$ and $\zeta_{j}$ with $\pi_{j}, \pi_{j}^{\prime}$ and $\mu_{j}$, respectively. Note also that if $\pi_{j}=0, \pi_{j}^{\prime}=0$, or $\mu_{j}=0$, then the corresponding constraint becomes inactive (in accordance with the fact that we have actually not introduced corresponding slack variables in these cases).

The fact that approaches for integrating absolute and relative evaluations for predicting ordinal labels, as far as we know, do not yet exist is one of the reasons why our method is useful. Since the proportional odds model is a traditional model that is considered to be the simplest to understand, it is natural to first develop an augmented version of this model based on our proposed approach. As a following step, it would be interesting to apply this approach to other models that are designed for ordinal regression.

As an example, we show that relative information can be integrated to augment support vector ordinal regression models [189, 231] based on our proposed approach. First, we denote by $\mathbf{x}_{i}^{k} \in X$ the $i$-th feature vector in the $k$-th class label and by $n^{k}$ the number of feature vectors in the $k$-th class label, where $k \in\{1, \ldots, q\}$ and $i \in\left\{1, \ldots, n^{k}\right\}$, such that $n=\sum_{k=1}^{q} n^{k}$ is the total number of feature vectors. Denote by $\phi(\mathbf{x})$ the feature vector in a high-dimensional reproducing kernel Hilbert space related to $\mathbf{x}$ by transformation. Thus, the problem is tackled by optimizing multiple thresholds to define parallel discriminant hyperplanes for the ordinal scale as follows:

$$
\begin{array}{ccl}
\underset{\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \gamma^{\prime}}{\operatorname{minimize}} & \frac{1}{2}\|\mathbf{w}\|_{2}^{2}+\alpha_{1}\left(\sum_{k=1}^{q} \sum_{i=1}^{n^{k}}\left(\gamma_{i}^{k}+\gamma_{i}^{\prime k}\right)\right), & \\
\text { subject to } & 0 \leq-1+\gamma_{i}^{k}-\mathbf{w}^{\top} \phi\left(\mathbf{x}_{i}^{k}\right)+\theta_{k} & , \text { for } i=1, \ldots, n^{k} \\
& & \text { and } k=1, \ldots, q,  \tag{10.7}\\
0 \leq-1+\gamma_{i}^{\prime k}+\mathbf{w}^{\top} \phi\left(\mathbf{x}_{i}^{k}\right)-\theta_{k-1} & , \text { for } i=1, \ldots, n^{k} \\
\theta_{k-1} \leq \theta_{k} & \text { and } k=1, \ldots, q, \\
0 \leq \gamma_{i}^{k} & & \text { for } k=2, \ldots, q-1, \\
& \text { for } i=1, \ldots, n^{k} \\
0 \leq \gamma_{i}^{\prime k} & \text { and } k=1, \ldots, q, \\
& \text {, for } i=1, \ldots, n^{k} \\
& \text { and } k=1, \ldots, q,
\end{array}
$$

where $\gamma^{k}$ and $\boldsymbol{\gamma}^{\prime k}$ are slack variables for the $k$-th parallel discriminant hyperplane and $\alpha_{1}>0$ controls the complexity of the model.

The proposed approach for integrating relative information can be similarly applied to extend support vector ordinal regression models as follows:

$$
\begin{align*}
& \underset{\mathbf{w}, \boldsymbol{\theta}, \gamma, \gamma^{\prime}, \boldsymbol{\xi}, \boldsymbol{\xi}^{\prime} \boldsymbol{\zeta}}{\operatorname{minimize}} \quad \frac{1}{2}\|\mathbf{w}\|_{2}^{2}+\alpha_{1}\left(\sum_{k=1}^{q} \sum_{i=1}^{n^{k}}\left(\gamma_{i}^{k}+\gamma_{i}^{\prime k}\right)\right) \\
& +\alpha_{2}\left(\sum_{j=1}^{n_{\text {g }}}\left(\pi_{j} \xi_{j}+\pi_{j}^{\prime} \xi_{j}^{\prime}+\epsilon_{j} \zeta_{j}\right)\right), \\
& \text { subject to } \\
& 0 \leq-1+\gamma_{i}^{k}-\mathbf{w}^{\top} \phi\left(\mathbf{x}_{i}^{k}\right)+\theta_{k} \quad, \text { for } \quad i=1, \ldots, n^{k} \\
& \text { and } k=1, \ldots, q \text {, } \\
& 0 \leq-1+\gamma_{i}^{\prime k}+\mathbf{w}^{\top} \phi\left(\mathbf{x}_{i}^{k}\right)-\theta_{k-1} \quad, \text { for } \quad i=1, \ldots, n^{k} \\
& \text { and } k=1, \ldots, q \text {, } \\
& \theta_{k-1} \leq \theta_{k} \quad \text {, for } \quad k=2, \ldots, q-1 \text {, } \\
& 0 \leq \xi_{j}-\mathbf{w}^{\top}\left(\mathbf{x}_{2}^{j}-\mathbf{x}_{1}^{j}\right) \quad \text {, for } \quad j=1, \ldots, n_{\mathscr{P}},  \tag{10.8}\\
& 0 \leq \xi_{j}^{\prime}-\mathbf{w}^{\top}\left(\mathbf{x}_{1}^{j}-\mathbf{x}_{2}^{j}\right) \quad \text {, for } \quad j=1, \ldots, n_{\mathscr{P}}, \\
& 0 \leq \zeta_{j}-\mathbf{w}^{\top}\left(\mathbf{x}_{2}^{j}-\mathbf{x}_{1}^{j}\right) \quad \text {, for } \quad j=1, \ldots, n_{\mathscr{P}}, \\
& 0 \leq \zeta_{j}-\mathbf{w}^{\top}\left(\mathbf{x}_{1}^{j}-\mathbf{x}_{2}^{j}\right) \quad \text {, for } \quad j=1, \ldots, n_{\mathscr{P}}, \\
& 0 \leq \gamma_{i}^{k} \quad \text {, for } \quad i=1, \ldots, n^{k} \\
& \text { and } k=1, \ldots, q \text {, } \\
& 0 \leq \gamma_{i}^{\prime k} \quad \text {, for } \quad i=1, \ldots, n^{k} \\
& \text { and } k=1, \ldots, q \text {, } \\
& 0 \leq \xi_{j} \quad \text {, for } \quad j=1, \ldots, n_{\mathscr{P}}, \\
& 0 \leq \xi_{j}^{\prime} \quad \text {, for } \quad j=1, \ldots, n_{\mathscr{P}},
\end{align*}
$$

where $\xi_{j}, \xi_{j}^{\prime}$ and $\zeta_{j}$ are slack variables with respective weights $\pi_{j}, \pi_{j}^{\prime}$, and $\epsilon_{j}$, $n_{\mathscr{P}}$ is the number of couples in the set $\mathscr{P}$ of couples of objects, and $\alpha_{1}, \alpha_{2}>1$ control the complexity of the model. Note that the $\ell_{2}$ regularization term can be replaced with an $\ell_{1}$ regularization term in an approach similar to the one shown in Eq. 8.11.

The nearest neighbours model [190 for ordinal classification is another model that can be augmented based on our approach. We hope that this approach inspires readers to further develop other models.

### 10.4. Experimental analysis

This section is devoted to illustrate the proposed method using synthetic examples and a real-life example that are representative of typical problem settings. The goal is to investigate how the incorporation of relative evaluations (ranking information) through constrained optimization problem defined by Eq. (10.6) influences the predictive performance of an ordinal regression model.

### 10.4.1. General experimental setup

It is to be expected that the performance of our data integration strategy depends on several characteristics of the data and the underlying phenomenon, including:

- The number of samples ( $n$ ) provided to trained panellists and untrained panellists.
- The number of labelling tasks $\left(n_{l}\right)$ performed by trained panellists, with one sample in each labelling task.
- The number of ranking tasks $\left(n_{r}\right)$ performed by untrained panellists, with $n_{o}$ samples in each ranking task.

The combined influence of these characteristics on the overall predictive performance of the ordinal regression model is investigated using synthetically generated data. Note that we do not focus on the number of trained panellists or untrained panellists, but on the number of tasks. In other words, we do not distinguish between an trained panellist providing labels for all $n$ samples or $n$ trained panellists providing a label for one sample, or equivalently, an untrained panellist providing $n_{r}$ rankings of $n_{o}$ samples or $n_{r}$ untrained panellists providing one ranking of $n_{o}$ samples.

We demonstrate the influence of removing a fraction of all labels gathered from trained panellists and instead integrating ranking information from untrained panellists. We do this by considering the setting where five labelling and ranking experiments are conducted, and, in each experiment, trained panellists and untrained panellists are provided with $n$ samples to perform labelling and ranking tasks. These five experiments are used as folds for applying 5 -fold cross validation 100 and extracting the cross-validated performance of the model.

Let us consider a case study of predicting the appreciation of meat products. In the field of sensory evaluation of food, volatile organic compounds (VOCs) are seen as good indicators of food spoilage [228, 232]. We try to mimic this setting and choose a lower bound on the number of VOCs present in various types of meat products [233, 234, 235] as the total number of features $p$ equal to 50 .

The general data simulation process for the experiments is as follows:
Generate the feature vectors (x): Feature vectors (for all $n$ samples evaluated by both trained panellists and untrained panellists in each of the five labelling and ranking experiments) are independent and identically distributed observations of a Gaussian distribution. This can be seen as drawing observations from a $p$-variate Gaussian distribution with a covariance matrix equal to the identity matrix.

Defining the latent variable and thresholds: For simplicity, we assume that all considered features are equally relevant (features that have an influence on the response). Therefore, the parameters in the vector $\mathbf{w}$ of the latent variable function $g$ are set to one. We consider a setting with five classes $(q=5)$. To
obtain the threshold parameters $\left(\theta_{1}, \ldots, \theta_{4}\right)$, a large sample of the latent variable is simulated, i.e., a large number of feature vectors is generated and only used to compute the latent variable $\left(\mathbf{w}^{\top} \mathbf{x}\right)$. The thresholds are chosen such that the empirical distribution of the labels is uniform over the label set. This results in the vector of thresholds given by $\boldsymbol{\theta}=[-7.6948,-2.6006,2.4936,7.5878]$.
Sampling absolute (trained) evaluations (labels): In each of the five labelling experiments, trained panellists are provided with $n$ samples to perform a total of $n_{l}$ labelling tasks. To simulate this, based on the proportional odds model, we compute, for each of the $n$ generated feature vectors, the probability mass function over the labels. To obtain labels from $n_{l}$ labelling tasks, we sample a feature vector from the set of $n$ feature vectors, $n_{l}$ times. For each sampled feature vector, we sample a label from the label set $(\mathscr{y}=\{1, \ldots, 5\})$ according to the corresponding probability mass function. After sampling all the labels, we compute, for each feature vector, the empirical distribution function over the different labels.

Sampling relative (untrained) evaluations (rankings): In each of the five ranking experiments, untrained panellists are also given the same $n$ samples as the trained panellists and are asked to perform $n_{r}$ ranking tasks on $n_{o}$ samples. In sensory analysis of food, due to olfactory fatigue of panellists, it is recommended to rank no more than six samples 24. In our experimental analysis, we consider two to four samples in each ranking task. To simulate this setting, we start by sampling without replacement $n_{o}$ feature vectors from the set of $n$ feature vectors. For each of the $n_{o}$ feature vectors, we sample a label from the label set according to the corresponding probability mass function. We gather $n_{\mathscr{P}}=\frac{n_{o}\left(n_{o}-1\right)}{2}$ couples and identify each of them with the index $j$ corresponding to the position of the couple in the lexicographic order. In each couple, if the first label is larger than the second, then the value of $\pi_{j}$ is increased by one. If the second label is larger than the first, then the value of $\pi_{j}^{\prime}$ is increased by one. Otherwise, if the labels are equal, then the value of $\mu_{j}$ is increased by one. We repeat this procedure $n_{r}$ times and gather the ranking information in a set $\mathscr{P}$, as in Eq. 10.1).

Tuning and evaluating the model: By considering the optimization problem defined by Eq. 10.6), we learn a linearly constrained ordinal regression model while combining the labels and rankings. To determine the optimal values of the tuning parameters $\alpha_{1}$ and $\alpha_{2}$ for solving Eq. 10.6 ) and to evaluate the performance of the model learned, we perform nested 5 -fold cross validation. We compute the mean KL divergence and the Earth mover's distance to extract the cross-validated performance of the model, as described in Chapter 8. It must be noted that the Earth mover's distance $d_{\text {EMD }}(P, \hat{P})$ between $P$ and $\hat{P}$ is influenced by the ordering of the class labels. For a complete description of this distance metric, we refer to [236]. Intuitively, however, it can be described as the minimal amount of work that needs to be performed to transform one distribution into another by moving "probability mass" between classes. Transporting one unit of mass to an adjacent class requires one unit of work. Therefore, transporting mass to classes that are
farther apart will require more work. Note, however, that this choice implies, in addition to the ordering, an equidistance assumption on the classes. Finally, to obtain the overall performance of a predictive model over an entire dataset, the mean dissimilarity or distance is computed over all objects in the dataset.

### 10.4.2. Experimental settings using synthetic data

We consider three settings: in the first, trained panellists are given four samples $(n=4)$ to perform 40 labelling tasks $\left(n_{l}=40\right)$ in each of the five labelling experiments; in the second, trained panellists are given five samples $(n=5)$ to perform 50 labelling tasks $\left(n_{l}=50\right)$ in each of the five labelling experiments; and, in the third, trained panellists are given six samples $(n=6)$ to perform 60 labelling tasks $\left(n_{l}=60\right)$ in each of the five labelling experiments. Note that, after $n$ feature vectors are generated (for all evaluated samples) and $n_{l}$ labels are sampled, we compute the empirical probability mass function over the label set $(\mathscr{y}=\{1, \ldots, 5\})$ for each sample.

By considering the optimization problem defined by Eq. 8.11), we learn a linearly constrained $\ell_{1}$-regularized ordinal regression model on $4 n$ feature vectors and $4 n_{l}$ labels gathered from trained panellists. We now simulate the influence of removing half of the $4 n_{l}$ labels that are available and instead integrate ranking information from untrained panellists. We consider the optimization problem defined by Eq. 10.6, however, we learn a model on the $4 n$ feature vectors and just $4 n_{l} / 2$ labels. We gather these labels by sampling $n_{l} / 2$ labels in each labelling experiment. We then compute the KL divergence between the predicted probability distribution using the model learned on the sampled $4 n_{l} / 2$ labels and the originally computed empirical probability distribution over the label set for each sample. We expect the performance of the model to improve after integrating ranking information gathered from untrained panellists and, after large amounts of ranking information, the performance to equal or maybe exceed the performance of the model learned on all the labels gathered from trained panellists. Note that, for each setting, the number of integrated ranking tasks $\left(n_{r}\right)$ is increased by multiples of the number of labelling tasks $\left(n_{l}\right)$.

Figure 10.1 shows the mean KL divergence (cross-validated performance) of the $\ell_{1}$-regularized ordinal regression model learned on all labels and is considered as reference. It is clear that including ranking information provided by untrained panellists, in the form of an increasing number of ranking tasks $\left(n_{r}\right)$, improves the overall performance of the models and, consequently, decreases the KL divergence. In Figure 10.1(a), Figure 10.1(c) and Figure 10.1(e), we see that, for a certain number of ranking tasks, the models learned on half the labels gathered from trained panellists outperform the models learned on all the labels gathered from trained panellists. For a very large number of ranking tasks, the performances of
all the augmented models seem to converge.


Figure 10.1: Experiments comparing $\ell_{1}$-regularized ordinal regression models learned on all labels as reference (dashed black line) with models learned on half the labels and integrating an increasing number of ranking tasks of two samples (blue line), three samples (green line), and four samples (orange line).

Figures 10.2 (a) and 10.2 (b) show the performance of the models using the Earth mover's distance metric. When comparing Figures 10.1 (a) and 10.2 (a), it can be seen that the overall patterns are highly similar. The most notable difference can be found for the case $n_{o}=2$. When the KL divergence measure is used as a performance measure, only a limited number of ranking tasks are needed to outperform the baseline model. However, when the Earth mover's distance metric is used, this number increases significantly. It can be understood that the equidistance assumption on the classes implied by the Earth mover's distance metric may have resulted in some errors in the calculation of $d_{\text {EMD }}(P, \hat{P})$. For larger values of $n_{o}$, this increase is less dramatical. The same conclusion can be drawn when comparing Figures 10.1(c) and 10.2(b).


Figure 10.2: Re-evaluation of the results shown in Figures 10.1 a) and 10.1 c) using the Earth mover's distance metric.

It is interesting to study the influence of additional information resulting from the inherent transitivity of a ranking on the performance of the models. As we have previously mentioned, a ranking of $n_{o}$ samples corresponds to information for $n_{\mathscr{P}}=\frac{n_{o}\left(n_{o}-1\right)}{2}$ couples of samples. In particular, a ranking of two samples corresponds to information for one couple, a ranking of three samples corresponds to information for three couples, and a ranking of four samples corresponds to information for six couples. Hence, we rescale Figures 10.1(a), 10.1(c) and 10.1(e) into Figures 10.1 (b), 10.1(d) and 10.1(f), respectively, by multiplying the number of ranking tasks by the respective number of couples $\left(n_{\mathscr{P}} \cdot n_{r}\right)$. These figures suggest that additional information resulting from the inherent transitivity of a ranking improves the performance of the models.

It must be noted that the value of $\alpha_{2}$ cannot be interpreted easily because the $\log$-likelihood, the parameters $\beta_{i}, \xi_{j}, \xi_{j}^{\prime}$ and $\zeta_{j}$ are not measured on the same scale. Parameters $\beta_{i}$ have the same scale as the parameter vectors, whereas parameters $\xi_{j}, \xi_{j}^{\prime}$ and $\zeta_{j}$ have the scale of the latent variable. Both scales might have different orders of magnitude. Therefore, even if, e.g., $\alpha=0.95$ might suggest that the
constraints have a strong influence as compared to the $\ell_{1}$ regularization at first sight, the opposite might be true if the scale of the latent variable is an order of magnitude larger than the scale of the parameters. Moreover, the fact that the KL divergence is decreasing as the number of ranking experiments increases shows that the optimisation problem gives a considerable amount of 'importance' to the novice rankings (otherwise the solution, and thus the resulting model, and its performance would be the same). The change in KL divergence is thus already an indirect quantification of this importance.

### 10.5. Application to sensory data

In this section, we apply the method introduced in this chapter to the datasets gathered from labelling tests in Chapter 4 that measure the degree of freshness of chicken breasts, cod, brown shrimp and salmon samples. Using the sensory data, we illustrate the effect of integrating ranking information on $\ell_{1}$-regularized ordinal regression models.

### 10.5.1. Chicken breast

First, the feature representations of eight chicken breast samples are obtained through the use of selected-ion flow-tube mass spectrometry (SIFT-MS) to measure concentrations of 23 VOCs and are gathered in Table B.2. Subsequently, 33 panellists were each asked to give their appreciation of eight chicken breast samples (i.e., $n=8$ and $n_{l}=264$ ) by evaluating them, on an ordinal scale consisting of three labels: "Spoiled" (SP), "Marginal" (M), and "Fresh" (F) such that SP $\prec \mathrm{M} \prec \mathrm{F}$. These evaluations are gathered in Table A.1. In addition, the feature representations of 25 chicken breast samples are gathered in Table B. 3 . Subsequently, different groups of panellists were asked to rank different sets of these chicken breast samples. These evaluations are gathered in Table A. 2 and Table A.3. We summarize these ranking tests in Table 10.1. In each ranking test, $n_{r}$ panellists perform one ranking task of $n_{o}$ samples. Note that the labelled chicken breast samples are different from the ranked chicken breast samples, hence, resulting in different sets of feature vectors.

The ranking data in Table 10.1 are integrated with the labels provided by trained panellists to fit an augmented ordinal regression model. In this case, the model is learned on only eight samples, which are insufficient to create large partitions for training and testing without losing significant training or testing capabilities. Here, nested leave-one-out cross validation is performed using $n$ folds to determine the optimal values of the tuning parameters $\alpha_{1}$ and $\alpha_{2}$ for solving Eq. 10.6 and to estimate the general performance of the model.

| Ranking test | $R_{1}$ | $R_{2}$ | $R_{3}$ | $R_{4}$ | $R_{5}$ | $R_{6}$ | $R_{7}$ | $R_{8}$ | $R_{9}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n_{r}$ | 3 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 3 |
| $n_{o}$ | 4 | 4 | 3 | 3 | 4 | 4 | 3 | 3 | 4 |
| Ranking test | $R_{10}$ | $R_{11}$ | $R_{12}$ | $R_{13}$ | $R_{14}$ | $R_{15}$ | $R_{16}$ | $R_{17}$ | $R_{18}$ |
| $n_{r}$ | 2 | 3 | 2 | 4 | 4 | 4 | 4 | 14 | 14 |
| $n_{o}$ | 4 | 3 | 3 | 4 | 4 | 3 | 3 | 5 | 5 |

Table 10.1: Ranking tests on chicken breast samples with different numbers of ranking tasks $\left(n_{r}\right)$ and samples ( $n_{o}$ ) per task.

To illustrate the general impact of exploiting the ranking information, we monitor the KL divergence while gradually increasing the number of ranking tests in Table 10.1. For each $m \in\{1, \ldots, 18\}$, we randomly sample 10 different sets of $m$ ranking tests, and compute the mean KL divergence. In Figure 10.3, we show the mean KL divergence after integrating ranking information from 10 different sets of $m$ ranking tests, and we see a decrease in the mean KL divergence value. In other words, as the amount of additional ranking information increases, the performance of the model improves. Moreover, to illustrate the variation in the KL divergence at every number of ranking tests, we plot the possible values between the maximum and the minimum KL divergence values as shaded. Interestingly, between three and ten ranking tests, the maximum KL divergence values slightly exceed the "reference" KL divergence value. This could be due to incorrect ranking of samples by panellists. However, it is clear that as we increase the number of ranking tests, the variation in the KL divergence values decreases.

On the basis of the parameter vector $\mathbf{w}$, it was observed that acetic acid, 2,3butanediol, acetoin, ethanol, ammonia and sulfur compounds were considered relevant. Several studies on meat spoilage, particularly on chicken, have identified that 3-methylbutanol, 2-propanol, acetic acid, 2,3-butanediol, acetoin, dimethyl sulfide, ethanol and ethyl acetate have an impact on the spoilage of chicken [37, 237. Therefore, the additional VOCs may also be considered as strong contributors for detecting food spoilage, regardless of the different MAP conditions and temperatures.

### 10.5.2. Atlantic cod

First, the feature representations of 24 cod samples are obtained through the use of selected-ion flow-tube mass spectrometry (SIFT-MS) to measure concentrations of 20 VOCs and are summarized in Table B.5. Subsequently, different groups of panellists were asked to give their appreciation of these cod samples


Figure 10.3: Experiment illustrating the KL divergence value of an $\ell_{1}$-regularized ordinal regression model as reference (dashed black line) in comparison to the mean (blue line) and range (shaded) of the KL divergence values of $\ell_{1}$-regularized ordinal regression models with integrated ranking information of chicken breasts.
by evaluating them, on an ordinal scale consisting of five labels: "Spoiled" (SP), "Marginal" (M), "Satisfactory" (S), "Fresh" (F) and "Very Fresh" (VF) such that $\mathrm{SP} \prec \mathrm{M} \prec \mathrm{S} \prec \mathrm{F} \prec \mathrm{VF}$. These evaluations are gathered in Table 4.5. In addition, the feature representations of another 24 cod samples are gathered in Table B.6. Subsequently, different groups of panellists were asked to rank different sets of these cod samples. These evaluations are gathered in Table A.6. We summarize these ranking tests in Table 10.2. In each ranking test, $n_{r}$ panellists perform one ranking task of $n_{o}$ samples. Note that the labelled cod samples are different from the ranked cod samples, hence, resulting in different sets of feature vectors.

| Ranking test | $R_{1}$ | $R_{2}$ | $R_{3}$ | $R_{4}$ | $R_{5}$ | $R_{6}$ | $R_{7}$ | $R_{8}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n_{r}$ | 10 | 8 | 8 | 8 | 9 | 8 | 9 | 8 |
| $n_{o}$ | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |

Table 10.2: Ranking tests on cod samples with different numbers of ranking tasks $\left(n_{r}\right)$ and samples ( $n_{o}$ ) per task.

The ranking data in Table 10.2 are integrated with the labels provided by trained panellists to fit an augmented ordinal regression model. In this case, the model is learned on only 24 samples, which are insufficient to create large partitions for training and testing without losing significant training or testing capabilities. Here, nested leave-one-out cross validation is performed using $n$ folds to determine the optimal values of the tuning parameters $\alpha_{1}$ and $\alpha_{2}$ for solving Eq. 10.6) and to
estimate the general performance of the model.
To illustrate the general impact of exploiting the ranking information, we monitor the KL divergence while gradually increasing the number of ranking tests in Table 10.2 . For each $m \in\{1, \ldots, 8\}$, we randomly sample 10 different sets of $m$ ranking tests, and compute the mean KL divergence. In Figure 10.4 we show the mean KL divergence after integrating ranking information from 10 different sets of $m$ ranking tests, and we see a decrease in the mean KL divergence value. In other words, as the amount of additional ranking information increases, the performance of the model improves. Moreover, to illustrate the variation in the KL divergence at every number of ranking tests, we plot the possible values between the maximum and the minimum KL divergence values as shaded. Interestingly, between one and three ranking tests, the maximum KL divergence values slightly exceed the "reference" KL divergence value. This could be due to incorrect ranking of samples by panellists. However, it is clear that as we increase the number of ranking tests, the variation in the KL divergence values decreases.


Figure 10.4: Experiment illustrating the KL divergence value of an $\ell_{1}$-regularized ordinal regression model as reference (dashed black line) in comparison to the mean (blue line) and range (shaded) of the KL divergence values of $\ell_{1}$-regularized ordinal regression models with integrated ranking information of Atlantic cod.

It was observed that the following VOCs were relevant: ethanol, 2,3-butanediol, 2-methylpropanol, 3-methylbutanol, dimethylamine and trimethylamine. Respectively, ethanol, 2,3-butanediol and 3-methylbutanol, have frequently been identified as potential spoilage indicators of cod under different MAP conditions and temperatures [238, 239, 240]. Therefore, the additional VOCs, 2-methylpropanol, dimethylamine and trimethylamine, can also be recognized as strong contributors for the predicting spoilage of cod fillets, regardless of the applied atmosphere.

### 10.5.3. Atlantic brown shrimp

First, the feature representations of 16 shrimp samples are obtained through the use of selected-ion flow-tube mass spectrometry (SIFT-MS) to measure concentrations of 20 VOCs and are summarized in Table B. 8 . Subsequently, different groups of panellists were asked to give their appreciation of these shrimp samples by evaluating them, on the ordinal scale used to evaluate cod samples. These evaluations are gathered in Table A.7. In addition, the feature representations of anther 16 shrimp samples are gathered in Table B.9. Subsequently, different groups of panellists were asked to rank different sets of these shrimp samples. These evaluations are gathered in Table A.8. These ranking tests are suammrized in Table 10.3. In each ranking test, $n_{r}$ panellists perform one ranking task of $n_{o}$ samples. Note that the labelled shrimp samples are different from the ranked shrimp samples, hence, resulting in different sets of feature vectors.

| Ranking test | $R_{1}$ | $R_{2}$ | $R_{3}$ | $R_{4}$ |
| :---: | :---: | :---: | :---: | :---: |
| $n_{r}$ | 9 | 9 | 10 | 10 |
| $n_{o}$ | 4 | 4 | 4 | 4 |

Table 10.3: Ranking tests on shrimp samples with different numbers of ranking tasks $\left(n_{r}\right)$ and samples ( $n_{o}$ ) per task.

The ranking data in Table 10.3 are integrated with the labels provided by trained panellists to fit an augmented ordinal regression model. Similar to the case study on cod samples, the model is learned on only 24 samples. Thus, nested leave-one-out cross validation is performed using $n$ folds to determine the optimal values of the tuning parameters $\alpha_{1}$ and $\alpha_{2}$ for solving Eq. 10.6) and to estimate the general performance of the model.

To illustrate the general impact of exploiting the ranking information, we monitor the KL divergence while gradually increasing the number of ranking tests in Table 10.3 For each $m \in\{1, \ldots, 4\}$, we randomly sample 10 different sets of $m$ ranking tests, and compute the mean KL divergence. In Figure 10.5, we show the mean KL divergence after integrating ranking information from 10 different sets of $m$ ranking tests, and we see a decrease in the mean KL divergence value. In other words, as the amount of additional ranking information increases, the performance of the model improves. Moreover, to illustrate the variation in the KL divergence at every number of ranking tests, we plot the possible values between the maximum and the minimum KL divergence values as shaded. It is clear that as we increase the number of ranking tests, the variation in the KL divergence values decreases.

In a recent study on brown shrimp during aerobic storage, it was observed that the relevant compounds are 1,2-butanediol, 2-propanol, 2-pentanone, butanone,


Figure 10.5: Experiment illustrating the KL divergence value of an $\ell_{1}$-regularized ordinal regression model as reference (dashed black line) in comparison to the mean (blue line) and range (shaded) of the KL divergence values of $\ell_{1}$-regularized ordinal regression models with integrated ranking information of Atlantic brown shrimp.
acetone, methyl mercaptan, sulphur hydride, dimethyl disulphide, ethyl acetate, acetic acid and ammonia [241. However, in this dissertation, it was observed that 2,3-butanediol, 3-methyl-1-butanol, dimethylamine and trimethylamine were considered relevant. Therefore, since we considered different MAP conditions and temperatures, it could be that these compounds are relevant, regardless of the applied atmosphere.

### 10.5.4. Atlantic salmon

First, the feature representations of 72 salmon samples are obtained through the use of selected-ion flow-tube mass spectrometry (SIFT-MS) to measure concentrations of 20 VOCs and are summarized in Table B. 13 . Subsequently, different groups of panellists were asked to give their appreciation of these salmon samples by evaluating them, on the ordinal scale used to evaluate cod and shrimp samples. These evaluations are gathered in Table 4.11. In addition, the feature representations of anther 72 salmon samples are gathered in Table B.11. Subsequently, different groups of panellists were asked to rank different sets of these salmon samples. These evaluations are gathered in Table A.10. These samples are summarized in Table 10.4. In each ranking test, $n_{r}$ panellists perform one ranking task of $n_{o}$ samples. Note that the labelled salmon samples are different from the ranked salmon samples, hence, resulting in different sets of feature vectors.

The ranking data in Table 10.4 are integrated with the labels provided by trained

| Ranking test | $R_{1}$ | $R_{2}$ | $R_{3}$ | $R_{4}$ | $R_{5}$ | $R_{6}$ | $R_{7}$ | $R_{8}$ | $R_{9}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n_{r}$ | 9 | 12 | 9 | 10 | 9 | 8 | 8 | 12 | 5 |
| $n_{o}$ | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Ranking test | $R_{10}$ | $R_{11}$ | $R_{12}$ | $R_{13}$ | $R_{14}$ | $R_{15}$ | $R_{16}$ | $R_{17}$ | $R_{18}$ |
| $n_{r}$ | 9 | 9 | 7 | 8 | 9 | 10 | 8 | 9 | 9 |
| $n_{o}$ | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |

Table 10.4: Ranking tests on salmon samples with different numbers of ranking tasks $\left(n_{r}\right)$ and samples ( $n_{o}$ ) per task.
panellists to fit an augmented ordinal regression model. In this case, the model is learned on 72 samples, which are quite sufficient to create large partitions for training and testing without losing significant training or testing capabilities. Here, nested 10 -fold cross validation is performed using 10 folds to determine the optimal values of the tuning parameters $\alpha_{1}$ and $\alpha_{2}$ for solving Eq. 10.6 and to estimate the general performance of the model.

To illustrate the general impact of exploiting the ranking information, we monitor the KL divergence while gradually increasing the number of ranking tests in Table 10.4 . For each $m \in\{1, \ldots, 18\}$, we randomly sample 10 different sets of $m$ ranking tests, and compute the mean KL divergence. In Figure 10.6, we show the mean KL divergence after integrating ranking information from 10 different sets of $m$ ranking tests, and we see a decrease in the mean KL divergence value. In other words, as the amount of additional ranking information increases, the performance of the model improves. Moreover, to illustrate the variation in the KL divergence at every number of ranking tests, we plot the possible values between the maximum and the minimum KL divergence values as shaded. It is clear that as we increase the number of ranking tests, the variation in the KL divergence values decreases.

In several studies, alcohols, ketones, dimethyl sulfide and hydrogen sulfide were considered as relevant compounds [53, 77]. However, it was observed in this dissertation that 3-methyl-1-butanol, dimethyl amine, ethyl acetate, methyl mercaptan, 3-methylbutanal, acetone, butanone, ammonia and carbon disulfide were considered relevant. Therefore, it is suggested that monitoring these additional VOCs could provide a better understanding at different MAP conditions and temperatures.


Figure 10.6: Experiment illustrating the KL divergence value of an $\ell_{1}$-regularized ordinal regression model as reference (dashed black line) in comparison to the mean (blue line) and range (shaded) of the KL divergence values of $\ell_{1}$-regularized ordinal regression models with integrated ranking information of Atlantic salmon.

### 10.6. Conclusions

In this chapter, we have presented an ordinal regression strategy that allows to combine absolute evaluations from trained panellists and relative evaluations from untrained panellists in the form of ordinal labels and rankings, respectively. This was done by applying the methods presented in Chapters 8 and 9 and rewriting the ordinal regression model as a constrained non-linear optimization problem that includes $\ell_{1}$ regularization and ranking information in the form of constraints. We have demonstrated a strategy of augmenting ordinal regression models that can be applied in several fields of science that adopt evaluation tests with different types of data. This strategy is useful, especially for problems where the number of samples is very small in comparison to the total number of features. We showed through simulation and real-world experimentation on the sensory and instrumental data in Chapter 4 that this strategy is efficient and works well for small datasets. Simulation studies showed that additional data in the form of rankings improved the performance of $\ell_{1}$-regularized ordinal regression models. Moreover, results of the studies on the sensory data were consistent with the results of the synthetic studies.

## PART V

EPILOGUE

## 11 General conclusions and perspectives

This dissertation is motivated by the context of intelligent food packaging for providing companies in the food industry with essential information on the quality status of packaged food products. Such information can help companies improve food logistics and traceability, increase convenience towards consumers and reduce food waste. Therefore, intelligent food packaging plays an essential role in improving processing, logistic operations and quality assurance measures. Currently, to determine characteristics of a food sample, sensory evaluation is performed. However, performing and analysing sensory evaluation can be complex, requiring a considerable amount of time and effort. Therefore, in this dissertation, we propose methods for analysing sensory evaluations to better determine the overall sensory characteristics of food samples. In addition, the proposed methods can help in achieving cost reduction and time efficiency. In this dissertation, we have further provided methods to be implemented alongside intelligent food packaging for determining the current quality status of a packaged food product. This enables the monitoring of all packages by gathering instrumental data of food samples, which on the short run diminishes and eradicates sampling plans for quality control, and on the long run predicts actual expiry dates, increases the margin of food safety, indicates freshness level and detects early spoilage.

### 11.1. Analysis of sensory evaluations

In Part III of this dissertation, we focused on answering the following questions:

1. How can we assign (joint) consensus ordinal labels?
2. How can we assign a consensus ranking?

Answering these questions is useful in sensory evaluation for determining the overall sensory quality or preference of food samples. It is a whole process that begins with data collection and ends ultimately with assigning a consensus evaluation.

In many sensory evaluation settings, trained panellists are asked to provide ordinal labels of a set of food samples for describing the overall sensory quality of the samples. In such settings, where only ordinal labels on the samples have been gathered, it can be assumed that the main goal is to assign a consensus label that describes the overall quality of a food sample. Therefore, the first question needs to be addressed.

In Chapter 5 ordinal labels were gathered from trained panellists to assign a consensus label that describes the overall quality of a food sample. We proposed a method for aggregating ordinal labels to obtain a consensus labelling. These ordinal labels represent abstract concepts, and dealing with perceptions is not an easy task. Therefore, we used monotonicity of the assigned ordinal labels for the aggregation of ordinal labels, which resulted in a method that simultaneously exploited all the information expressed by the trained panellists.

In many other sensory evaluation settings, untrained (or less trained) panellists are asked to provide a preference or ranking of food samples for describing a consensus ranking of the samples. In such settings, where only rankings on the samples have been gathered, it can be assumed that the main goal is to assign a consensus ranking that describes the relative sensory quality of the food samples. Therefore, the second question needs to be addressed.

In Chapter 6 rankings were gathered from untrained panellists to assign a consensus ranking that describes the relative sensory quality of food samples. We proposed a method for aggregating rankings to obtain a consensus ranking. We advocated for monotonicity of a profile of rankings for the aggregation of ordinal labels, which resulted in a method that simultaneously exploited all the information expressed by the untrained panellists.

In answering the first question, companies in the food industry are typically required to hire trained panellists or provide extensive training to the panellists, which can be costly, time consuming and resource intensive. As a result, they may find themselves facing the problem of having a very small number of panellists (i.e., a small number of evaluations). Therefore, they may risk describing wrong absolute quality of the food samples. Interestingly, since untrained panellists require less to no training, they may be an additional source of information on the food samples. Additionally, other types of information on the food samples may also be known and exploited.

In other settings, some companies are interested in determining the preference of one product over another. Therefore, the second question needs to be answered. To target a broad group of consumers, companies need to hire a very large number of untrained panellists, which can be costly, time consuming and resource intensive. As a result, they may find themselves facing the problem of having a very small number of panellists (i.e., a small number of evaluations), and may risk describing wrong relative quality of the food samples. Interestingly, trained panellists may be an additional source of information of absolute evaluations.

In Chapter 7. we dealt with the above-described problems of having a very small number of panellists. Scores and rankings were gathered from trained and untrained panellists, respectively, and combined to assign a consensus label/ranking that describes the absolute/relative quality of food samples. To tackle the problem of having a small number of trained panellists, we proposed an approach for assigning
consensus scores to food samples, while integrating rankings. To tackle the problem of having a small number of untrained panellists, we proposed an approach for assigning a consensus ranking of food samples, while integrating scores. These two approaches resulted in assigning improved consensus scores/ranking of the samples. Therefore, we advocated considering these two methods to answer the following question:
3. How can we combine scores and rankings to improve the assessment of food samples?

It has been realized that, when answering the third question, companies may not always have additional information in the form of rankings. However, they may have other types of (relative) information on the samples, such as the knowledge of storage days, the results of a clustering analysis or other sensory evaluation tests. Therefore, we have proposed a method for incorporating such additional information in the form of constraints, to improve the quality of the consensus scores of the samples. Applying this method helped us deduce possible consensus vectors of scores of the samples in settings where the consensus preference was not clear.

The methods presented to address the third question can be seen as a starting point to combining different types of absolute and relative information. The following step would be to apply the novel methods that addressed the first two questions for combining ordinal labels and rankings. It would be interesting to express the property of monotonicity when dealing with ordinal labels and rankings, at the same time. Accordingly, this opens up several problems:
A. Integrating rankings: Obtaining the joint consensus labelling of multiple samples by simultaneously exploiting all the information expressed by the trained (labelling) and untrained (ranking) panellists.
B. Integrating labels: Obtaining the consensus ranking of samples by simultaneously exploiting all the information expressed by the trained (labelling) and untrained (ranking) panellists.
C. Incorporating information:
i. Obtaining the joint consensus labelling of multiple samples by exploiting all the information expressed by the trained (labelling) panellists and incorporating additional information on the samples.
ii. Obtaining the consensus ranking of samples by incorporating additional information on the samples.

Moreover, it would be interesting to develop a combined approach of A and C(i), where the problem is to obtain the joint consensus labelling of multiple samples by simultaneously exploiting all the information expressed by the trained (labelling)
and untrained (ranking) panellists, and, at the same time, incorporating additional information on the samples.

Furthermore, in some cases, a panellist may not be able to provide a label/score to a subset of the samples provided to them. In other cases, there may be a limited number of samples, due to lack of proper planning or due to mechanical or technical failures of storage means, and, as a result provides a subset of the samples to a number of panellists. In all these cases, there will be missing information for some food samples. Therefore, a researcher may either consider the panellists' partial evaluations as insufficient and remove them, or, since the number of evaluations (i.e., the number of panellists and samples) is small, the best and simplest solution is to impute the missing values, such as by considering the mean/median. This can also be seen as a problem in which one wants to predict the value of a pair (panellist, sample), therefore, pairwise learning can be used 242 .

These proposed methods offer support to companies in search of appropriate solutions to problems, such as consumer rejection, food waste, etc. For instance, the methods provide improved consensus evaluations of food samples, and, thus, allow for cost reduction and time-saving measures during quality control of food samples. Furthermore, combining absolute evaluations with relative informations could provide a more accurate understanding of the general appreciation of a food sample. For instance, while expert wine tasters are often used as quality indicators, research has shown that there is an increasing demand for a more consumer-orientated system of sensory evaluation of luxury food and wine 243]. Given the fact that wine experts and consumers perceive wine quality differently, it would be interesting for wine producers to combine scores and rankings from the experts and the consumers, respectively, to provide labelled wine bottles that facilitate their customers' decision-making. As a result, the customers would appreciate the food product more, and, in effect, reduce their food waste.

### 11.2. Prediction of sensory evaluations

In Part IV of this dissertation, we focused on settings where instrumental information about the samples was collected. In these settings, we dealt with the following research questions:
4. How can we predict an ordinal label?

## 5. How can we predict rankings?

Answering these questions is useful in the development of intelligent food packages for determining the overall sensory quality or preference of food samples. It is
a whole process that begins with the collection of sensory evaluation data and chemical data of packaged food.

As has already been discussed, trained panellists are typically asked to provide ordinal labels for a set of food samples for describing the overall sensory quality of the samples. Many companies within the food industry have shown increased interest in gathering direct information of the quality status of a packaged food product. Therefore, it is of great importance to study the influence of VOC concentrations in the food packages on the sensory quality of the samples. In such a setting, measured VOC concentrations of samples and their respective assigned ordinal labels are gathered. Consequently, it would be interesting to predict an ordinal label given a measurement of the VOC concentrations of a new sample in a food package. Therefore, the fourth question needs to be addressed.

In Chapter 8 the concentrations of VOCs in packaged food samples were measured. Furthermore, trained panellists were asked to evaluate the sensory quality of the food samples and assign ordinal labels to the food samples. Researchers looking to answer the fourth question may encounter several methods that take advantage of the ordinal nature of the sensory evaluation data. However, as far as we know, very few of these methods provide a solution to having a large number of features (i.e., VOCs) compared to the number of instances (i.e., evaluated samples). Therefore, we proposed an ordinal regression model that takes instrumental data (i.e., features of the studied samples) as input and sensory data (i.e., assigned ordinal labels by trained panellists) as output, while including $\ell_{1}$ regularisation. It was shown that including $\ell_{1}$ regularisation resulted in feature selection, where features that are more influential on the sensory evaluation were selected from the large number of features. It was then concluded that this approach improved the performance of ordinal regression models.

As has already been discussed, untrained (or less trained) panellists are typically asked to provide a preference or ranking of food samples. A researcher may also be interested in studying the influence of VOC concentrations in food packages on the preference of the samples. In such a setting, where a researcher having gathered measurements of VOC concentrations of samples and rankings of these samples, it would be interesting to be able to predict a ranking given measurements of the VOC concentrations of a new set of samples. Therefore, the fifth question needs to be addressed.

In Chapter 9, apart from measuring the concentration of VOCs in packaged food samples, untrained panellists were asked to provide a ranking or preference of the food samples. Researchers looking to answer the fifth question may encounter several methods, however, not all of these methods take advantage of the ordinal nature of the sensory evaluation data. Furthermore, as far as we know, very few of these methods provide a solution to having a large number of features (i.e., VOCs) compared to the number of instances (i.e., evaluated samples). Therefore, we
proposed a model to take instrumental data (i.e., features of the studied samples) as input and sensory data (i.e, assigned rankings by untrained panellists) as output, while including $\ell_{1}$ regularisation. It was then concluded that this approach improved the performance of the ordinal regression models.

In addition to gathering sensory evaluations from trained panellists, which can be costly and time consuming, chemical analysis on a large number of packaged foods needs to be performed, which can further be resource intensive. These additional costs can have a negative impact on the size and quality of the generated dataset (i.e., a small number of studied samples and a small number of trained panellists providing an evaluation for the samples), and may risk building a weak model. Interestingly, since untrained panellists require less to no training, they may be considered to be a good source of additional information of preferences/rankings on the food samples.

In Chapter 10 we dealt with the above-described problem of having a small dataset to build an interpretable model. We proposed an approach for predicting an ordinal label of a new sample in a food package, while including $\ell_{1}$ regularisation and integrating preferences/rankings. Interestingly, the proposed approach is not limited to gathering preferences/rankings on the same samples for which ordinal labels are provided by trained panellists. Therefore, this allows flexibility in gathering more data. It was concluded that this strategy resulted in augmented ordinal regression models and was important especially for cases where the number of samples is very small in comparison to the total number of features. Therefore, we advocated considering this method to answer the following question:

## 6. How can we combine ordinal labels and rankings to improve the prediction of an ordinal label?

The methods presented to address the above questions can be seen as a starting point to incorporating different types of relative information in ordinal regression models. Moreover, it would be interesting to consider a combination of the likelihood functions in Chapters 8 and 9 . This has similarities with the method of integrating scores and rankings in Chapter 10. Furthermore, it would be interesting to consider prediction problems that involve a time component, such as predicting the sensory quality of a food sample at a later time, which can be applied to predict the actual expiry date or shelf-life of a food sample. However, while the time component adds additional information, it also makes time series problems more difficult to handle, compared to ordinal regression models.

These proposed methods offer support to companies in search of appropriate solutions for gathering essential information on the quality status of packaged food products to tackle problems, such as traceability, food-waste, etc. Firstly, the methods provide early-on detection of packaged food that does not meet predefined quality standards. For instance, logistic operations along the food supply chain bear the risk of mechanical product damages. These damages may
have different consequences, starting with simple package deformation and loss of package integrity, which lead to a compromise on the quality and safety of the food. Therefore, companies using the methods proposed in this dissertation can monitor food products along the supply chain and obtain information on the evolution of the sensory quality of the packaged food product. As a result, the companies can uncover weak points in the logistic operations and determine the reasons for these weak points, such as equipment defects, human errors, etc. Secondly, the methods allow companies to better objectify the sensory quality of food products. For instance, a food processing company can provide product certification to retailers and ensure the correct classification of the food product. Furthermore, the companies can take consumer preferences into consideration when providing product certification.

As intelligent food packaging technology is bound to replace human decision-making, this gives rise to many liability questions, particularly when the machine makes an incorrect decision. A complicating factor is that machine learning is difficult for users of the intelligent food packaging technology to understand, and, thus, they can claim ignorance and consider it a 'black box'. This is important because the liability questions often revolve around who the responsible 'party' is and what the responsible 'party' knew, or should have known, at the time of the liability. Another complication in the liability questions is determining whether the machine learning software is part of the intelligent food product. Therefore, in an event that the technology fails to make a correct decision, at least one of the following 'parties' can be held accountable: the software producer, the hardware producer, the package producer and the packaged food producer. Thus, the companies that supplies the end product (i.e., the packaged food producer) should establish a duty of care to ensure and prove that reasonable care has been taken to avoid events of harm to the consumer, for example, through the process of testing and evaluating the technology on the specific food product. Regulating intelligent food packages may enforced to ensure that duty of care is established. It seems that regulating intelligent food packaging technology would not be aimed at resolving the legal liability questions, but rather at reassuring the consumers that this technology has been developed with public safety in mind, and in a way which allows problems to be identified and rectified.

This work has implications for analysing and predicting the sensory quality of food samples. We believe that this work will open doors to many new approaches in different fields of study. It is important for the reader to note that having optimal results from the use of the proposed methods depends on the quality of the gathered data. Thus, it is first recommended to collect asessments of samples according to the general guidelines and practices in the field of study, such as Sensory Evalution Practices [16] in the field of food science. Second, in the field of food science, it is recommended to perform accurate measurements of the VOC profiles. Otherwise there exists the potential risk of inaccurately quantifying the spoilage metabolites
in packaged food. Furthermore, it is important to consider the recommendations provided in Chapter 7 on choosing an optimal source of additional information. For instance, based on accurate measurements of the VOC profiles, clustering analysis should be performed carefully.

In conclusion, precise questioning based on the findings of this dissertation could help in further studies to find incentives for which researchers and companies are ready to apply improved quality assurance measures and further develop intelligent food packaging for improving food logistics and traceability, increasing convenience towards consumers and reducing food waste.

## Bibliography

[1] H. T. Lawless and H. Heymann, Sensory Evaluation of Food. Food Science Text Series, New York: Springer, 2nd ed., 2010.
[2] M. A. Van Boekel, "Kinetic modeling of food quality: A critical review," Comprehensive Reviews in Food Science and Food Safety, vol. 7, pp. 144-158, 2008.
[3] G. Slavica, G. Radoslav, and K. Karmela, "Effects of modified atmosphere packaging on quality and safety of fresh meat," Quality Of Life, vol. 1, no. 2-4, pp. 121-133, 2010.
[4] J. Mattheis and J. K. Fellman, "Impacts of modified atmosphere packaging and controlled atmospheres on aroma, flavor, and quality of horticultural commodities," HortTechnology, vol. 10, no. 3, pp. 507-510, 2000.
[5] FAO, "Global food losses and food waste - extent, causes and prevention," (Rome), 2011.
[6] "European Commision. Preparatory Study on Food Waste Across EU 27, Technical Report 2010-054," (European Commission, Brussels), 2011.
[7] K. B. Biji, C. N. Ravishankar, C. O. Mohan, and T. K. Srinivasa Gopal, "Smart packaging systems for food applications: A review," Journal of Food Science and Technology, vol. 52, no. 10, pp. 6125-6135, 2015.
[8] J.-E. Haugen and K. Kvaal, "Electronic nose and artificial neural network," Meat Science, vol. 49, no. 98, pp. S273-S286, 1998.
[9] P. Singham, P. Birwal, and B. K. Yadav, "Importance of objective and subjective measurement of food quality and their inter-relationship," Journal of Food Process Technology, vol. 6, no. 9, pp. 1-7, 2015.
[10] M. Vanderroost, P. Ragaert, F. Devlieghere, and B. De Meulenaer, "Intelligent food packaging: The next generation," Trends in Food Science 8 Technology, vol. 39, no. 1, pp. 47-62, 2014.
[11] R. I. Masel, Pinciples of Adsdorption and Reaction on Solid Surfaces. New York, NY, USA: John Wiley and Sons Ltd, 1996.
[12] R. A. Matuszko, "Theoretical Aspects of Multicomponent Adsorption Equilibria," Technical Report, Washington, DC, 1997.
[13] C. Tien, Multicomponent and adsorption calculations, vol. 3. Newton, MA: Butterworth-Heinemann, 1994.
[14] P. Gramatica, "Principles of QSAR models validation: Internal and external," QSAR and Combinatorial Science, vol. 26, no. 5, pp. 694-701, 2007.
[15] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C. P. Varghese, and R. K. Agrawal, "Validation of QSAR models - strategies and importance," International Journal of Drug Design and Disocovery, vol. 2, no. 3, pp. 511-519, 2011.
[16] H. Stone and J. L. Sidel, Sensory Evaluation Practices. California: Elsevier, 3rd ed., 2004.
[17] C. Miller and K. Swift, The Handbook of Nonsexist Writing for Writers, Editors and Speakers. London: The Women's Press, 3rd ed., 1995.
[18] Á. Labella, Y. Liu, R. Rodríguez, and L. Martínez, "Analyzing the performance of classical consensus models in large scale group decision making: A comparative study," Applied Soft Computing, vol. 67, pp. 677-690, 2018.
[19] Y. Liu, C. Liang, F. Chiclana, and J. Wu, "A trust induced recommendation mechanism for reaching consensus in group decision making," KnowledgeBased Systems, vol. 119, pp. 221-231, 2017.
[20] J. Wu, F. Chiclana, H. Fujita, and E. Herrera-Viedma, "A visual interaction consensus model for social network group decision making with trust propagation," Knowledge-Based Systems, vol. 122, pp. 39-50, 2017.
[21] G. Ares and P. Varela, "Trained vs. consumer panels for analytical testing: Fueling a long lasting debate in the field," Food Quality and Preference, vol. 61, pp. 79-86, 2017.
[22] L. Gram, L. Ravn, M. Rasch, J. B. Bruhn, A. B. Christensen, and M. Givskov, "Food spoilage-interactions between food spoilage bacteria," International Journal of Food Microbiology, vol. 78, no. 1-2, pp. 79-97, 2002.
[23] R. Pérez-Fernández, M. Sader, and B. De Baets, "Joint consensus evaluation of multiple objects on an ordinal scale: An approach driven by monotonicity," Information Fusion, vol. 42, pp. 64-74, 2018.
[24] M. A. Amerine, R. N. Pangborn, and E. B. Poessler, Principles of Sensory Evaluation of Food. New York: Academic Press, 1965.
[25] D. C. Parkes and S. Seuken, "Social choice and rank aggregation," in Economics and Computation, ch. 13, pp. 323-367, Cambridge University Press, 2013.
[26] K. Y. Lam and P. H. Franses, "Analyzing preference rankings when there are too many alternatives," Technical Report, Econometric Institute, Rotterdam, 2008.
[27] R. Pérez-Fernández, M. Rademaker, and B. De Baets, "Monometrics and their role in the rationalisation of ranking rules," Information Fusion, vol. 34, pp. 16-27, 2017.
[28] M. Sader, R. Pérez-Fernández, L. Kuuliala, F. Devlieghere, and B. De Baets, "A combined scoring and ranking approach for determining overall food quality," International Journal of Approximate Reasoning, vol. 100, pp. 161-176, 2018.
[29] M. Sader, R. Pérez-Fernández, and B. De Baets, "Combining Absolute and Relative Information in Studies on Food Quality," in Information Processing and Management of Uncertainty in Knowledge-Based Systems - Theory and Foundations (J. et al. Medina, ed.), (Cádiz, Spain), pp. 1-10, Springer International Publishing AG, part of Springer Nature, 2018.
[30] M. Sader, J. Verwaeren, and B. De Baets, "Integrating expert and novice evaluations for augmenting ordinal regression models," Information Fusion, vol. 51, pp. 1-9, 2019.
[31] M. C. Meilgaard, G. V. Civille, and B. T. Carr, Sensory Evaluation Techniques. Boca Raton: CRC Press, 4th ed., 2006.
[32] H. R. Moskowitz, J. H. Beckley, and A. V. Resurreccion, Sensory and Consumer Research in Food Product Design and Development. Ames, Iowa: Blackwell, 2006.
[33] J. A. Bower, Statistical Methods for Food Science. Edinburgh: John Wiley \& Sons, 2nd ed., 2013.
[34] J. Lim, "Hedonic scaling: A review of methods and theory," Food Quality and Preference, vol. 22, no. 8, pp. 733-747, 2011.
[35] D. Peryam and F. Pilgrim, "Hedonic scale method of measuring food preferences," Food Technology, vol. 11, no. S1, pp. 9-14, 1957.
[36] M. C. J. Gacula, J. Singh, J. Bi, and S. Altan, Statistical Methods in Food and Consumer Research. New York: Academic Press, 2nd ed., 2009.
[37] A. Casaburi, P. Piombino, G.-J. E. Nychas, F. Villani, and D. Ercolini, "Bacterial populations and the volatilome associated to meat spoilage," Food Microbiology, vol. 45, pp. 83-102, 2015.
[38] L. Kuuliala, E. Abatih, A.-G. Ioannidis, M. Vanderroost, B. De Meulenaer, P. Ragaert, and F. Devlieghere, "Multivariate statistical analysis for the identification of potential seafood spoilage indicators," Food Control, vol. 84, pp. 49-60, 2018.
[39] E. Jaffrès, V. Lalanne, S. Macé, J. Cornet, M. Cardinal, T. Sérot, X. Dousset, and J.-J. Joffraud, "Sensory characteristics of spoilage and volatile compounds
associated with bacteria isolated from cooked and peeled tropical shrimps using SPME-GC-MS analysis," International Journal of Food Microbiology, vol. 147, no. 3, pp. 195-202, 2011.
[40] J. Lim and T. Fujimaru, "Evaluation of the labeled Hedonic scale under different experimental conditions," Food Quality and Preference, vol. 21, no. 5, pp. 521-530, 2010.
[41] J. L. García-Lapresta and R. González del Pozo, "An ordinal multi-criteria decision-making procedure under imprecise linguistic assessments," European Journal of Operational Research, vol. 279, pp. 159-167, 2019.
[42] J. L. García-Lapresta, R. González del Pozo, and D. Pérez-Román, "Metrizable ordinal proximity measures and their aggregation," Information Sciences, vol. 49, pp. 149-163, 2018.
[43] J. L. García-Lapresta and D. Pérez-Román, "Ordinal proximity measures in the context of unbalanced qualitative scales and some applications to consensus and clustering," Applied Soft Computing Journal, vol. 35, pp. 864872, 2015.
[44] J. L. García-Lapresta and D. Pérez-Román, "A consensus reaching process in the context of non-uniform ordered qualitative scales," Fuzzy Optimization and Decision Making, vol. 16, no. 4, pp. 449-461, 2017.
[45] J. L. García-Lapresta and D. Pérez-Román, "Aggregating opinions in nonuniform ordered qualitative scales," Applied Soft Computing Journal, vol. 67, pp. 652-657, 2018.
[46] L. Nicolas, C. Marquilly, and M. O'Mahony, "The 9-point hedonic scale: Are words and numbers compatible?," Food Quality and Preference, vol. 21, no. 8, pp. 1008-1015, 2010.
[47] J. H. J. Huis in't Veld, "Microbial and biochemical spoilage of foods: An overview," International Journal of Food Microbiology, vol. 33, no. 1, pp. 1-18, 1996.
[48] J. M. Jay, Modern Food Microbiology. Gaithersburg, Maryland: Aspen Publishers, 5th ed., 1998.
[49] C. J. Stannard, A. P. Williams, and P. A. Gibbs, "Temperature/growth relationships for psychrotrophic food-spoilage bacteria," Food Microbiology, vol. 2, no. 2, pp. 115-122, 1985.
[50] I. S. Arvanitoyannis and K. Kotsanopoulos, "Safety and quality control of modified atmosphere packaging products," in Modified Atmosphere and Active Packaging Technologies (I. S. Arvanitoyannis, ed.), ch. 3, pp. 59-144, Boca Raton: CRC Press, 2012.
[51] G. L. Robertson, Food Packaging Principles and Practice. New York: CRC Press, 3rd ed., 2012.
[52] L. Kuuliala, Y. Al Hage, A.-G. Ioannidis, M. Sader, F. M. Kerckhof, M. Vanderroost, N. Boon, B. De Baets, B. De Meulenaer, P. Ragaert, and F. Devlieghere, "Microbiological, chemical and sensory spoilage analysis of raw Atlantic cod (Gadus morhua) stored under modified atmospheres," Food Microbiology, vol. 70, pp. 232-244, 2018.
[53] L. Kuuliala, M. Sader, A. Solimeo, R. Pérez-fernández, M. Vanderroost, and B. D. Baets, "Spoilage evaluation of raw Atlantic salmon (Salmo salar) stored under modified atmospheres by multivariate statistics and augmented ordinal regression," International Journal of Food Microbiology, vol. 303, pp. 46-57, 2019.
[54] S. Bruckner, A. Albrecht, B. Petersen, and J. Kreyenschmidt, "Characterization and comparison of spoilage processes in fresh pork and poultry," Journal of Food Quality, vol. 35, pp. 372-382, 2012.
[55] A. Casaburi, P. Piombino, G.-J. E. Nychas, F. Villani, and D. Ercolini, "Bacterial populations and the volatilome associated to meat spoilage," Food Microbiology, vol. 45, pp. 83-102, 2014.
[56] G. J. Dickes, "The application of gas chromatography to food analysis," Talanta, vol. 26, pp. 1065-1099, 1979.
[57] M. H. Gordon, Principles and Applications of Gas Chromatography in Food Analysis. Boston: Springer, 1990.
[58] S. J. Lehotay and J. Hajšlová, "Application of gas chromatography in food analysis," Trends in Analytical Chemistry, vol. 21, no. 9-10, pp. 686-697, 2002.
[59] D. Smith and P. Španěl, "Selected ion flow tube mass spectrometry (SIFTMS) for on-line trace gas analysis," Mass Spectrometry Reviews, vol. 24, no. 5, pp. 661-700, 2005.
[60] B. M. Davis and M. J. McEwan, "Determination of olive oil oxidative status by selected ion flow tube mass spectrometry," Journal of Agricultural and Food Chemistry, vol. 55, no. 9, pp. 3334-3338, 2007.
[61] B. Noseda, P. Ragaert, D. Pauwels, T. Anthierens, H. Van Langenhove, J. Dewulf, and F. Devlieghere, "Validation of selective ion flow tube mass spectrometry for fast quantification of volatile bases produced on atlantic cod (Gadus morhua)," Journal of Agricultural and Food Chemistry, vol. 58, no. 9, pp. 5213-5219, 2010.
[62] A. Olivares, K. Dryahina, J. L. Navarro, D. Smith, P. Španěl, and M. Flores, "SPME-GC-MS versus selected ion flow tube mass spectrometry (SIFT-MS)
analyses for the study of volatile compound generation and oxidation status during dry fermented sausage processing," Journal of Agricultural and Food Chemistry, vol. 59, no. 5, pp. 1931-1938, 2011.
[63] S. Van Kerrebroeck, J. Vercammen, R. Wuyts, and L. De Vuyst, "Selected ion flow tube-mass spectrometry for online monitoring of submerged fermentations: A case study of sourdough fermentation," Journal of Agricultural and Food Chemistry, vol. 63, no. 3, pp. 829-835, 2015.
[64] E. R. Vieira, "Quality and sensory evaluation of food," in Elementary Food Science (D. R. Heldman, ed.), ch. 5, pp. 92-99, Dordrecht: Springer, 4th ed., 1996.
[65] B. Noseda, M. T. Islam, M. Eriksson, M. Heyndrickx, K. De Reu, H. Van Langenhove, and F. Devlieghere, "Microbiological spoilage of vacuum and modified atmosphere packaged Vietnamese Pangasius hypophthalmus fillets," Food Microbiology, vol. 30, no. 2, pp. 408-419, 2012.
[66] A. Olivares, K. Dryahina, P. Spaněl, and M. Flores, "Rapid detection of lipid oxidation in beef muscle packed under modified atmosphere by measuring volatile organic compounds using SIFT-MS," Food Chemistry, vol. 135, no. 3, pp. 1801-1808, 2012.
[67] L. Rokach and O. Maimon, "Clustering methods," in Data Mining and Knowledge Discovery Handbook (R. Maimon O., ed.), ch. 15, pp. 321-352, Boston: Springer, 2005.
[68] A. A. Abdullah, M. Altaf-Ul-Amin, N. Ono, T. Sato, T. Sugiura, A. H. Morita, T. Katsuragi, A. Muto, T. Nishioka, and S. Kanaya, "Development and mining of a volatile organic compound database," BioMed Research International, vol. 2015, pp. 1-13, 2015.
[69] A. A. Abdullah, M. Altaf-Ul-Amin, N. Ono, N. Yusuf, A. Zakaria, T. Nishioka, and S. Kanaya, "Comparison of clustering methods in the context of chemical structure similarity based classification of VOCs," Procedia Chemistry, vol. 20, pp. 40-44, 2016.
[70] N. Kaurt and P. J. Kaurt, "Cluster quality based performance evaluation of hierarchical clustering method," in 1st International Conference on Next Generation Computing Technologies, (Dehradun, India), pp. 649-653, 2015.
[71] S. Saitta, B. Raphael, and I. Smith, "A comprehensive validity index for clustering," Intelligent Data Analysis, vol. 12, no. 6, pp. 529-548, 2008.
[72] R. R. Sokal and F. J. Rohlf, "The comparison of dendrograms by objective methods," Taxon, vol. 11, no. 2, pp. 33-40, 1962.
[73] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 63, no. 2, pp. 411-423, 2001.
[74] F. Zaidi, D. Archambault, and G. Melançon, "Evaluating the quality of clustering algorithms using cluster path lengths," in Advances in Data Mining. Applications and Theoretical Aspects (P. Perner, ed.), (Berlin), pp. 42-56, Springer, 2010.
[75] D. I. Ellis and D. Broadhurst, "Rapid and quantitative detection of the microbial spoilage of meat by Fourier transform infrared spectroscopy and machine learning," Applied and Environmental Microbiology, vol. 68, no. 6, pp. 2822-2828, 2002.
[76] D. Wu and D.-W. Sun, "Potential of time series-hyperspectral imaging (TSHSI) for non-invasive determination of microbial spoilage of salmon flesh," Talanta, vol. 111, pp. 39-46, 2013.
[77] R. Jónsdóttir, G. Ólafsdóttir, E. Chanie, and J. E. Haugen, "Volatile compounds suitable for rapid detection as quality indicators of cold smoked salmon (Salmo salar)," Food Chemistry, vol. 109, no. 1, pp. 184-195, 2008.
[78] M. Mataragas, P. Skandamis, G.-J. E. Nychas, and E. H. Drosinos, "Modeling and predicting spoilage of cooked, cured meat products by multivariate analysis," Meat Science, vol. 77, no. 3, pp. 348-356, 2007.
[79] E. Z. Panagou, O. Papadopoulou, J. M. Carstensen, and G.-J. E. Nychas, "Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage," International Journal of Food Microbiology, vol. 174, pp. 1-11, 2014.
[80] S. Panigrahi, S. Balasubramanian, H. Gu, C. Logue, and M. Marchello, "Design and development of a metal oxide based electronic nose for spoilage classification of beef," Sensors and Actuators B: Chemical, vol. 119, no. 1, pp. 2-14, 2006.
[81] S. Panigrahi, S. Balasubramanian, H. Gu, C. Logue, and M. Marchello, "Neural-network-integrated electronic nose system for identification of spoiled beef," LWT - Food Science and Technology, vol. 39, pp. 135-145, mar 2006.
[82] O. Papadopoulou, E. Z. Panagou, C. C. Tassou, and G.-J. E. Nychas, "Contribution of Fourier transform infrared (FTIR) spectroscopy data on the quantitative determination of minced pork meat spoilage," Food Research International, vol. 44, pp. 3264-3271, dec 2011.
[83] A. A. Argyri, E. Z. Panagou, P. A. Tarantilis, M. Polysiou, and G.-J. Nychas, "Rapid qualitative and quantitative detection of beef fillets spoilage based on

Fourier transform infrared spectroscopy data and artificial neural networks," Sensors and Actuators B: Chemical, vol. 145, no. 1, pp. 146-154, 2010.
[84] A. A. Argyri, R. M. Jarvis, D. Wedge, Y. Xu, E. Z. Panagou, R. Goodacre, and G. J. E. Nychas, "A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage," Food Control, vol. 29, no. 2, pp. 461-470, 2013.
[85] K. F. Schulbach, R. L. Rouseff, and C. A. Sims, "Relating descriptive sensory analysis to gas chromatography/olfactometry ratings of fresh strawberries using partial least squares regression," Journal of Food Science, vol. 69, no. 7, pp. 273-277, 2004.
[86] R. Zhi, L. Zhao, B. Shi, H. Wang, Z. Li, J. Zhang, X. Xi, and Y. Jin, "Predicting sensory quality of longjing tea on the basis of physiochemical data," Sensors and Materials, vol. 25, no. 4, pp. 269-284, 2013.
[87] G. Birkhoff, Lattice Theory, vol. 25. New York: American Mathematical Society, 1940.
[88] B. A. Davey and H. A. Priestley, Introduction to Lattices and Order. Cambridge University Press, 2nd ed., 2002.
[89] F. S. Roberts and B. Tesman, Applied Combinatorics. Boca Raton: CRC Press, 2nd ed., 2009.
[90] M. Ehrgott, Multicriteria Optimization. Berlin, Heidelberg: Springer, 2nd ed., 2005.
[91] G. Anandalingam and T. L. Friesz, "Hierarchical optimization: an introduction," Annals of Operations Research, vol. 34, pp. 1-11, 1992.
[92] T. R. Rockafellar, Convex Analysis. Princeton: Princeton University Press, 1970.
[93] S. Boyd and L. Vandenberghe, Convex Optimization, vol. 25. Cambridge: Cambridge University Press, 2010.
[94] J. Nocedal and S. J. Wright, Numerical Optimization. New York: Springer, 2nd ed., 2006.
[95] C. H. Papadimitriou and K. Steiglitz, Combinatorial Optimization: Algorithms and Complexity. Englewood Cliffs: Prentice-Hall, 1984.
[96] V. N. Vapnik, "An overview of statistical learning theory," IEEE Transactions on Neural Networks, vol. 10, no. 5, pp. 988-999, 1999.
[97] B. Schölkopf and A. J. Smola, Learning with Kernels. Cambridge: The MIT Press, 2002.
[98] C. M. Bishop, Pattern Recognition and Machine Learning. New York: Springer, 2006.
[99] O. Bousquet, S. Boucheron, and G. Lugosi, "Introduction to statistical learning theory," Biological Cybernetics, vol. 3176, no. 1, pp. 169-207, 2004.
[100] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," Statistics Surveys, vol. 4, pp. 40-79, 2010.
[101] A. E. Hoerl and R. W. Kennard, "Ridge regression: Application to nonorthogonal problems," Technometrics, vol. 12, no. 1, pp. 69-82, 1970.
[102] R. G. Brereton and G. R. Lloyd, "Support vector machines for classification and regression.," The Analyst, vol. 135, no. 2, pp. 230-267, 2010.
[103] V. N. Vapnik, The Nature of Statistical Learning Theory. New York: Springer, 2000.
[104] R. Tibshirani, "Regression selection and shrinkage via the lasso," Journal of the Royal Statistical Society B (Methodological), vol. 58, no. 1, pp. 267-288, 1996.
[105] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," Foundations and Trends in Machine Learning, vol. 4, no. 1, pp. 1-106, 2011.
[106] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method forlarge-scale L1-regularized least squares," IEEE Journal of Selected Topics in Signal Processing, vol. 1, no. 4, pp. 606-617, 2007.
[107] K. Koh, S.-J. Kim, and S. Boyd, "An efficient method for large-scale L1regularized convex loss minimization," in Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, pp. 223-230, 2007.
[108] G.-X. Yuan, K.-W. Chang, C.-J. Hsieh, and C.-J. Lin, "A comparison of optimization methods and software for large-scale L1-regularized linear classification," The Journal of Machine Learning Research, vol. 11, pp. 3183-3234, 2010.
[109] S. S.-I. Lee, H. Lee, P. Abbeel, and A. Y. Ng, "Efficient L1 regularized logistic regression," American Association for Artificial Intelligence, vol. 21, no. 1, p. 401, 2004.
[110] A. Agresti, Categorical Data Analysis. New York: John Wiley \& Sons, 2nd ed., 2002.
[111] L. Kuuliala, Multidisciplinary Quality Characterization for the Development of Active and Intelligent Packaging Technologies for Muscle foods. PhD Thesis. Tampere University of Technology; Ghent University, 2018.
[112] T. Meskanen and H. Nurmi, "Distance from consensus: A theme and variations," in Mathematics and Democracy (B. Simeone and F. Pukelsheim, eds.), ch. 9, pp. 117-132, Berlin: Springer, 2006.
[113] R. Pérez-Fernández, P. Alonso, I. Díaz, S. Montes, and B. De Baets, "Monotonicity as a tool for differentiating between truth and optimality in the aggregation of rankings," Journal of Mathematical Psychology, vol. 77, pp. 19, 2017.
[114] S. E. Kemp, T. Hollowood, and J. Hort, Sensory Evaluation: A Practical Handbook. Chichester: John Wiley \& Sons, 2009.
[115] S. Wichchukit and M. O’Mahony, "The 9-point hedonic scale and hedonic ranking in food science: some reappraisals and alternatives," Journal of the Science of Food and Agriculture, vol. 95, no. 11, pp. 2167-2178, 2015.
[116] M. Pasch, Vorlesungen über neuere Geometrie. Leipzig: Teubner, 1882.
[117] E. Pitcher and M. F. Smiley, "Transitivities of betweenness," Transactions of the American Mathematical Society, vol. 52, no. 1, pp. 95-114, 1942.
[118] P. C. Fishburn, "Betweenness, orders and interval graphs," Journal of Pure and Applied Algebra, vol. 1, no. 2, pp. 159-178, 1971.
[119] A. C. Little, B. C. Jones, and L. M. DeBruine, "Facial attractiveness: evolutionary based research," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 366, pp. 1638-1659, 2011.
[120] M. A. S. Nunes and R. Hu, "Personality-based recommender systems," in Proceedings of the 6th ACM Conference on Recommender Systems, (Dublin), pp. 5-6, ACM Press, 2012.
[121] R. S. Horton, "Similarity and attractiveness in social perception: differentiating between biases for the self and the beautiful," Self and Identity, vol. 2, no. 2, pp. 137-152, 2003.
[122] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," International Journal of Computer Vision, vol. 40, no. 2, pp. 99-121, 2000.
[123] G. Monge, Mémoire sur la Théorie des Déblais et des Remblais. Paris: L'Imprimrie Royale, 1781.
[124] D. Klingman and R. Russell, "Solving constrained transportation problems," Operations Research, vol. 23, no. 1, pp. 91-106, 1975.
[125] H. W. Kuhn, "The Hungarian method for the assignment problem," Naval Research Logistics Quarterly 2, vol. 1-2, pp. 83-97, 1955.
[126] D. P. Bertsekas, "The Auction algorithm: a distributed relaxation method for the assignment problem," Annals of Operations Research, vol. 14, pp. 105-123, 1988.
[127] L. Brozovsky and V. Petricek, "Recommender system for online sating service," in Proceedings of Znalosti 2007 Conference, (Ostrava, Czech Republic), pp. 112, VSB, 2007.
[128] F. M. Harper and J. A. Konstan, "The MovieLens datasets: History and context," ACM Transactions on Interactive Intelligent Systems, vol. 5, no. 4, pp. 1-19, 2015.
[129] F. Chiclana, F. Herrera, and E. Herrera-Viedma, "Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations," Fuzzy Sets and Systems, vol. 97, no. 1, pp. 33-48, 1998.
[130] F. Chiclana, F. Herrera, and E. Herrera-Viedma, "Integrating multiplicative preference relations in a multipurpose decision-making model based on fuzzy preference relations," Fuzzy Sets and Systems, vol. 122, no. 2, pp. 277-291, 2001.
[131] K. K. F. Yuen, "Combining compound linguistic ordinal scale and cognitive pairwise comparison in the rectified fuzzy TOPSIS method for group decision making," Fuzzy Optimization and Decision Making, vol. 13, no. 1, pp. 105-130, 2014.
[132] J. C. Borda, Mémoire sur les Élections au Scrutin. Paris: Histoire de l'Académie Royale des Sciences, 1781.
[133] M. Condorcet, Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix. Paris: De l'Imprimerie Royale, 1785.
[134] J. G. Kemeny, "Mathematics without numbers," Daedalus, vol. 88, no. 4, pp. 577-591, 1959.
[135] T. Meskanen and H. Nurmi, "Closeness counts in social choice," in Power, Freedom, and Voting, ch. 15, pp. 289-306, Berlin: Springer, 2008.
[136] D. Black, "Partial justification of the Borda count," Public Choice, vol. 28, no. 1, pp. 1-15, 1976.
[137] J. L. García-Lapresta and D. Pérez-Román, "Some measures of consensus generated by distances on weak orders," in Proceedings of XIV Congreso Español Sobre Tecnologías y Lógica Fuzzy, (Cuencas Mineras (Mieres - Langreo)), pp. 477-483, 2008.
[138] J. M. Colomer, "Ramon Llull: from Ars electionis' to social choice theory," Social Choice Welfare, vol. 40, pp. 317-328, 2013.
[139] D. Black, The Theory of Committees and Elections. Cambridge: Cambridge University Press, 1958.
[140] P. Jean-Charles and S. Barba-Romero, Multicriterion Decision in Management: Principles and Practice. New York: Springer Science+Business Media, 2000.
[141] M. G. Kendall, "A new measure of rank correlation," Biometrika, vol. 30, no. 1-2, pp. 81-93, 1938.
[142] J. G. Kemeny and L. J. Snell, "Preference rankings: An axiomatic approach," in Mathematical Models in the Social Sciences, ch. 2, pp. 9-23, New York: Blaisdell, 1962.
[143] J. L. García-Lapresta and D. Pérez-Román, "Consensus measures generated by weighted Kemeny distances on weak orders," in Proceedings of the 10th International Conference on Intelligent Systems Design and Applications, (Cairo), pp. 463-468, 2010.
[144] M. Dummett, "The Borda count and agenda manipulation," Social Choice and Welfare, vol. 15, pp. 289-296, 1998.
[145] H. P. Young and A. Levenglick, "A Consistent extension of Condorcet's election prrinciple," SIAM Journal on Applied Mathematics, vol. 34, no. 3, pp. 515-523, 1978.
[146] J. Bartholdi, C. A. Tovey, and M. A. Trick, "Voting schemes for which it can be difficult to tell who won the election," Social Choice and Welfare, vol. 6, no. 2, pp. 157-165, 1989.
[147] R. Pérez-Fernández, M. Rademaker, P. Alonso, I. Díaz, S. Montes, and B. De Baets, "Representations of votes facilitating monotonicity-based ranking rules: From votrix to votex," International Journal of Approximate Reasoning, vol. 73, pp. 87-107, 2016.
[148] M. Rademaker and B. De Baets, "A ranking procedure based on a natural monotonicity constraint," Information Fusion, vol. 17, no. 1, pp. 74-82, 2014.
[149] M. Rademaker and B. De Baets, "Aggregation of monotone reciprocal relations with application to group decision making," Fuzzy Sets and Systems, vol. 184, no. 1, pp. 29-51, 2011.
[150] N. G. Andjiga, A. Y. Mekuko, and I. Moyouwou, "Metric rationalization of social welfare functions," Mathematical Social Sciences, vol. 72, pp. 14-23, 2014.
[151] N. Rodrigue, M. Guillet, J. Fortin, and J.-F. Martin, "Comparing information obtained from ranking and descriptive tests of four sweet corn products," Food Quality and Preference, vol. 11, no. 1-2, pp. 47-54, 2000.
[152] T. Næs, P. B. Brockhoff, and O. Tomic, Statistics for Sensory and Consumer Science. 2010.
[153] S. Ovadia, "Ratings and rankings: Reconsidering the structure of values and their measurement," International Journal of Social Research Methodology, vol. 7, no. 5, pp. 403-414, 2004.
[154] H. van Herk and M. van de Velden, "Insight into the relative merits of rating and ranking in a cross-national context using three-way correspondence analysis," Food Quality and Preference, vol. 18, no. 8, pp. 1096-1105, 2007.
[155] D. P. Bolhuis, A. Costanzo, and R. S. Keast, "Preference and perception of fat in salty and sweet foods," Food Quality and Preference, vol. 64, pp. 131-137, 2017.
[156] T. L. Bowman and S. Barringer, "Analysis of factors affecting volatile compound formation in roasted pumpkin seeds with selected ion flow tube-mass spectrometry (SIFT-MS) and sensory analysis.," Journal of Food Science, vol. 71, no. 1, pp. C51-C60, 2012.
[157] A. Øvrum, F. Alfnes, V. L. Almli, and K. Rickertsen, "Health information and diet choices: Results from a cheese experiment," Food Policy, vol. 37, no. 5, pp. 520-529, 2012.
[158] H. R. Moskowitz, "Experts versus consumers: A comparison," Journal of Sensory Studies, vol. 11, no. 1, pp. 19-37, 1996.
[159] E. P. Cox III, "The optimal number of response alternatives for a scale: A review," Journal of Marketing Research, vol. 17, no. 4, pp. 407-422, 1980.
[160] J. P. D'Angelo and D. B. West, Mathematical Thinking: Problem-Solving and Proofs. Upper Saddle River: Prentice Hall, 2nd ed., 2001.
[161] R. W. Bailey, "The number of weak orderings of a finite set," Social Choice and Welfare, vol. 15, pp. 559-562, 1998.
[162] C.-L. Hwang and A. S. M. Masud, Multiple Objective Decision Making - Methods and Applications, vol. 164 of Lecture Notes in Economics and Mathematical Systems. Berlin: Springer, 1979.
[163] K. Miettinen, Nonlinear Multiobjective Optimization, vol. 12 of International Series in Operations Research \& Management Science. Boston: Springer, 1998.
[164] A. V. Zykina, "A lexicographic optimization algorithm," Automation and Remote Control, vol. 65, no. 3, pp. 363-368, 2004.
[165] C.-L. Hwang and A. Masud, Multiple Objective Decision Making Methods and Applications. Berlin, Heidelberg: Springer-Verlag, 1st ed., 1979.
[166] K. Miettinen, Nonlinear Multiobjective Optimization. New York: Springer US, 1st ed., 1998.
[167] C. A. C. Coello, G. B. Lamont, D. A. V. Veldhuizen, D. E. Goldberg, and J. R. Koza, Evolutionary Algorithms for Solving Multi-Objective Problems. Springer US, 2nd ed., 2007.
[168] G. Butler, L. M. Poste, D. A. Mackie, and A. Jones, "Time-intensity as a tool for the measurement of meat tenderness," Food Quality and Preference, vol. 7, no. 3-4, pp. 193-204, 1996.
[169] N. François, C. Guyot-Declerck, B. Hug, D. Callemien, B. Govaerts, and S. Collin, "Beer astringency assessed by time-intensity and quantitative descriptive analysis: Influence of pH and accelerated aging," Food Quality and Preference, vol. 17, no. 6, pp. 445-452, 2006.
[170] A. C. Noble, "Application of time-intensity procedures for the evaluation of taste and mouthfeel," American Journal of Enology and Viticulture, vol. 46, no. 1, pp. 128-133, 1995.
[171] C. R. Calkins and J. M. Hodgen, "A fresh look at meat flavor," Meat Science, vol. 77, no. 1, pp. 63-80, 2007.
[172] G. A. Sullivan and C. R. Calkins, "Ranking beef muscles for Warner-Bratzler shear force and trained sensory panel ratings from published literature," Journal of Food Quality, vol. 34, no. 3, pp. 195-203, 2011.
[173] J. Bi, Sensory Discrimination Tests and Measurements: Sensometrics in Sensory Evaluation. Chichester: John Wiley \& Sons, 2nd ed., 2015.
[174] M. L. Harwood, G. R. Ziegler, and J. E. Hayes, "Rejection thresholds in chocolate milk: Evidence for segmentation," Food Quality and Preference, vol. 26, no. 1, pp. 128-133, 2012.
[175] S. Saraçli, N. Doan, and I. Doan, "Comparison of hierarchical cluster analysis methods by cophenetic correlation," Journal of Inequalities and Applications, vol. 2013, pp. 1-8, 2013.
[176] J. Ignacio, F. E. Cabrera, and L. G. Vargas, "Knowledge-based systems estimating the importance of consumer purchasing criteria in digital ecosystems," Knowledge-Based Systems, vol. 162, pp. 252-264, 2018.
[177] J. I. Peláez, R. Bernal, and M. Karanik, "Majority OWA operator for opinion rating in social media," Soft Computing, vol. 20, pp. 1047-1055, 2016.
[178] D. U. Ahn, C. Jo, and D. Olson, "Analysis of volatile components and the sensory characteristics of irradiated raw pork," Meat Science, vol. 54, no. 3, pp. 209-215, 2000.
[179] S. P. Heenan, J.-P. Dufour, N. Hamid, W. Harvey, and C. M. Delahunty, "Characterisation of fresh bread flavour: Relationships between sensory characteristics and volatile composition," Food Chemistry, vol. 116, no. 1, pp. 249257, 2009.
[180] T. Lavilla, J. Puy, M. L. López, I. Recasens, and M. Vendrell, "Relationships between volatile production, fruit quality, and sensory evaluation in Granny Smith apples stored in different controlled-atmosphere treatments by means of multivariate analysis," Journal of Agricultural and Food Chemistry, vol. 47, no. 9, pp. 3791-3803, 1999.
[181] Y. Liu, B. G. Lyon, W. R. Windham, C. E. Lyon, and E. M. Savage, "Prediction of physical, color, and sensory characteristics of broiler breasts by visible/near infrared reflectance spectroscopy," Poultry Science, vol. 83, no. 8, pp. 1467-1474, 2004.
[182] L. Shepard, R. E. Miracle, P. Leksrisompong, and M. Drake, "Relating sensory and chemical properties of sour cream to consumer acceptance," Journal of Dairy Science, vol. 96, no. 9, pp. 5435-5454, 2013.
[183] J. Tománková, "Volatile organic compounds as biomarkers of the freshness of poultry meat packaged in a modified atmosphere," Czech Journal of Food Sciences, vol. 30, no. 5, pp. 395-403, 2012.
[184] D. I. Ellis, D. Broadhurst, and R. Goodacre, "Rapid and quantitative detection of the microbial spoilage of beef by Fourier transform infrared spectroscopy and machine learning," Analytica Chimica Acta, vol. 514, no. 2, pp. 193-201, 2004.
[185] S. Limbo, L. Torri, N. Sinelli, L. Franzetti, and E. Casiraghi, "Evaluation and predictive modeling of shelf life of minced beef stored in high-oxygen modified atmosphere packaging at different temperatures," Meat Science, vol. 84, no. 1, pp. 129-36, 2010.
[186] H. Nilsen and M. Esaiassen, "Predicting sensory score of cod (Gadus morhua) from visible spectroscopy," LWT - Food Science and Technology, vol. 38, no. 1, pp. 95-99, 2005.
[187] Q. Ouyang, J. Zhao, and Q. Chen, "Instrumental intelligent test of food sensory quality as mimic of human panel test combining multiple crossperception sensors and data fusion," Analytica Chimica Acta, vol. 841, pp. 6876, 2014.
[188] P. McCullagh, "Regression models for ordinal data," Journal of the Royal Statistical Society. Series B (Methodological), vol. 42, no. 2, pp. 109-142, 1980.
[189] W. Chu and S. S. Keerthi, "Support vector ordinal regression," Neural Computation, vol. 19, no. 3, pp. 792-815, 2007.
[190] B. Nguyen, C. Morell, and B. De Baets, "Distance metric learning for ordinal classification based on triplet constraints," Knowledge-Based Systems, vol. 142, pp. 17-28, 2017.
[191] A. Albert and J. A. Anderson, "On the existence of maximum likelihood estimates in logistic regression models," Biometrika, vol. 71, no. 1, pp. 1-10, 1984.
[192] J. A. Anderson and P. R. Philips, "Regression, discrimination and measurement models for ordered categorical variables," Journal of the Royal Statistical Society, vol. 30, no. 1, pp. 22-31, 1981.
[193] W. Chu and Z. Ghahramani, "Gaussian processes for ordinal regression," Journal of Machine Learning Research, vol. 6, pp. 1019-1041, 2005.
[194] C. Croux, G. Haesbroeck, and C. Ruwet, "Robust estimation for ordinal regression," Journal of Statistical Planning and Inference, vol. 143, no. 9, pp. 1486-1499, 2013.
[195] T. Kanamori, "Statistical models and learning algorithms for ordinal regression problems," Information Fusion, vol. 14, no. 2, pp. 199-207, 2013.
[196] S. Kramer, G. Widmer, B. Pfahringer, and M. de Groeve, "Prediction of ordinal classes using regression trees," in Foundations of Intelligent Systems (Z. W. Raś and S. Ohsuga, eds.), vol. 47, pp. 426-434, Berlin: Springer, 2000.
[197] C. Winship and R. D. Mare, "Regression models with ordinal variables," American Sociological Review, vol. 49, no. 4, pp. 512-525, 1984.
[198] S. Cagnone, S. Mignani, and I. Moustaki, "Latent variable models for ordinal data," in Statistical Methods for the Evaluation of Educational Services and Quality of Products (M. Bini, B. Bertaccini, and S. Bacci, eds.), ch. 2, pp. 17-28, Berlin: Springer, 2009.
[199] S. Cagnone and P. Monari, "Latent variable models for ordinal data by using the adaptive quadrature approximation," Computational Statistics, vol. 28, no. 2, pp. 597-619, 2013.
[200] E. Frank and M. Hall, "A simple approach to ordinal classification," in Proceedings of the 12th European Conference on Machine Learning (L. De Raedt and P. Flach, eds.), (London, UK), pp. 145-156, Springer-Verlag, 2001.
[201] N. R. Parsons, M. L. Costa, J. Achten, and N. Stallard, "Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package R," Computational Statistics and Data Analysis, vol. 53, no. 3, pp. 632-641, 2009.
[202] G. Tutz and K. Hechenbichler, "Aggregating classifiers with ordinal response structure," Journal of Statistical Computation and Simulation, vol. 75, no. 5, pp. 391-408, 2005.
[203] J. Aitchison and S. Silvey, "The generalization of probit analysis to the case of multiple responses," Biometrika, vol. 44, pp. 131-140, 1957.
[204] G. M. Fitzmaurice, N. M. Laird, and J. H. Ware, Applied Longitudinal Analysis. New York: John Wiley \& Sons, 2nd ed., 2011.
[205] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, Numerical Optimization. Berlin: Springer, 2nd ed., 2006.
[206] C. T. Kelley, Iterative Methods for Optimization. Philadelphia: Society for Industrial and Applied Mathematics, 1999.
[207] F. E. Harrell Jr., Regression Modeling Strategies. Springer Series in Statistics, New York: Springer International Publishing, 2nd ed., 2015.
[208] W. Waegeman, B. De Baets, and L. Boullart, "On the scalability of ordered multi-class ROC analysis," Computational Statistics and Data Analysis, vol. 52, no. 7, pp. 3371-3388, 2008.
[209] W. Waegeman, B. De Baets, and L. Boullart, "ROC analysis in ordinal regression learning," Pattern Recognition Letters, vol. 29, no. 1, pp. 1-9, 2008.
[210] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," International Statistical Review, vol. 70, no. 3, pp. 419-435, 2002.
[211] G. C. Cawley and N. L. C. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," Journal of Machine Learning Research, vol. 11, pp. 2079-2107, 2010.
[212] K. P. Murphy, Machine learning: a probabilistic perspective. Cambridge, Massachusetts: MIT Press, 2012.
[213] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, Bayesian Data Analysis. 2nd ed., 2003.
[214] W. W. Cohen, R. E. Schapire, and Y. Singer, "Learning to order things," Journal of Artificial Intelligence Research, vol. 10, no. 1, pp. 243-270, 1999.
[215] T.-Y. Liu, "Learning to rank for information retrieval," Foundations and Trends in Information Retrieval, vol. 3, no. 3, pp. 225-331, 2007.
[216] J. A. Aledo, J. A. Gámez, and D. Molina, "Tackling the supervised label ranking problem by bagging weak learners," Information Fusion, vol. 35, pp. 38-50, 2017.
[217] J. Fürnkranz and E. Hüllermeier, "Pairwise preference learning and ranking," in Proceedings of the 14th European Conference on Machine Learning, pp. 145156, Springer-Verlag, 2003.
[218] J. Fürnkranz and E. Hüllermeier, eds., Preference Learning. Berlin: Springer, 2010.
[219] K. Crammer and Y. Singer, "Pranking with ranking," Neural Information Processing Systems, pp. 641-647, 2001.
[220] K. Crammer and Y. Singer, "Loss bounds for online category ranking," in Conference on Learning Theory (P. Auer and R. Meir, eds.), (Berlin), pp. 48-62, Springer, 2005.
[221] R. Herbrich, T. Graepel, and K. Obermayer, "Support vector learning for ordinal regression," in Proceedings of the Ninth International Conference on Artificial Neural Networks, no. 470, pp. 97-102, 1999.
[222] C. He, C. Wanq, Y. X. Zhonq, and R. F. Li, "A survey on learning to rank," in Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, vol. 3, (Kunming, China), pp. 1734-1739, 2008.
[223] A. Shashua and A. Levin, "Ranking with large margin principle: Two approaches," in Advances in Neural Information Processing Systems 15 (S. Becker, S. Thrun, and K. Obermayer, eds.), pp. 937-944, 2002.
[224] W. Chu and Z. Ghahramani, "Preference learning with Gaussian processes," in Proceedings of the 22nd International Conference on Machine learning, pp. 137-144, 2005.
[225] J. Fodor and M. Roubens, Fuzzy Preference Modelling and Multicriteria Decision Support. Dordrecht: Springer, 1994.
[226] E. Hüllermeier and J. Fürnkranz, "Comparison of ranking procedures in pairwise preference learning," in Proceedings of the 10th International Conference on Information Processing and Management of Uncertainty in KnowledgeBased Systems, (Perugia, Italy), 2004.
[227] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," Journal of Machine Learning Research, vol. 5, pp. 975-1005, 2004.
[228] C. R. Kerth, ed., The Science of Meat Quality. Oxford: John Wiley \& Sons, 2013.
[229] P. Ye and D. Doermann, "Combining preference and absolute judgements in a crowd-sourced setting," in Proceedings of Inernational Conference on Machine Learning, pp. 1-7, 2013.
[230] S. Corrente, S. Greco, M. Kadziński, and R. Słowiński, "Robust ordinal regression in preference learning and ranking," Machine Learning, vol. 93, no. 2-3, pp. 381-422, 2013.
[231] W. Chu and S. S. Keerthi, "New approaches to support vector ordinal regression," in ICML '05: Proceedings of the 22nd International Conference on Machine Learning, pp. 145-152, 2005.
[232] R. P. Singh, "Scientific principles of shelf life evaluation," in Shelf Life Evaluation of Foods (C. M. D. Man and A. A. Jones, eds.), ch. 1, pp. 3-26, Boston: Springer, 1994.
[233] M. L. Hernández-Macedo, C. J. Contreras-Castillo, S. M. Tsai, S. H. Da Cruz, C. I. G. L. Sarantopoulos, M. Padula, and C. T. S. Dias, "Gases and volatile compounds associated with micro-organisms in blown pack spoilage of Brazilian vacuum-packed beef," Letters in Applied Microbiology, vol. 55, no. 6, pp. 467-475, 2012.
[234] M. Kosowska, M. A. Majcher, and T. Fortuna, "Volatile compounds in meat and meat products," Food Science and Technology, vol. 37, no. 1, pp. 1-7, 2017.
[235] A. Watanabe, Y. Ueda, M. Higuchi, and N. Shiba, "Analysis of volatile compounds in beef fat by dynamic-headspace solid-phase microextraction combined with gas chromatographymass spectrometry," Journal of Food Science, vol. 73, no. 5, pp. C420-C425, 2008.
[236] E. Levina, P. Bickel, and B. Ca, "The Earth Mover's distance is the Mallows distance: some insights from statistics," Proceedings Eighth IEEE International Conference on Computer Vision, vol. 2, pp. 251-256, 2001.
[237] A. D'hoore, Examination of Microbial Spoilage Processes in Chicken Breast Fillets under Modified Atmosphere Packaging (MAP). Master thesis, Ghent University, 2016.
[238] L. Kuuliala, Y. Al Hage, A.-G. Ioannidis, M. Sader, F.-M. Kerckhof, M. Vanderroost, N. Boon, B. De Baets, B. De Meulenaer, P. Ragaert, and F. Devlieghere, "Microbiological, chemical and sensory spoilage analysis of raw Atlantic cod (Gadus morhua) stored under modified atmospheres," Food Microbiology, vol. 70, pp. 232-244, 2017.
[239] N. Qualities, "Evaluation of shelf life of superchilled cod (Gadus morhua) fillets and the influence of temperature fluctuations during storage on microbial and chemical quality indicators," Journal of Food Science, vol. 71, no. 2, pp. S97-S109, 2006.
[240] E. Reynisson, H. L. Lauzon, H. Magnússon, R. Jónsdóttir, G. Ólafsdóttir, V. Marteinsson, and G. Ó. Hreggvisson, "Bacterial composition and succes-
sion during storage of North-Atlantic cod (Gadus morhua) at superchilled temperatures," BMC Microbiology, vol. 12, pp. 1-12, 2009.
[241] K. Broekaert, B. Noseda, M. Heyndrickx, G. Vlaemynck, and F. Devlieghere, "Volatile compounds associated with Psychrobacter spp. and Pseudoalteromonas spp., the dominant microbiota of brown shrimp (Crangon crangon) during aerobic storage," International Journal of Food Microbiology, vol. 166, no. 3, pp. 487-493, 2013.
[242] M. Stock, Exact and Efficient Algorithms for Pairwise Learning. PhD thesis, Ghent Univeristy, 2017.
[243] J. Schiefer and C. Fische, "The gap between wine expert ratings and consumer preferences: measures, determinants and marketing implications," International Journal of Wine Business Research, vol. 20, no. 4, pp. 335-351, 2008.

## A Appendix

## A.1. Sensory evaluation for chicken breasts

## Labelling

| List of labels | H4 |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $a^{0}$ | $a^{5}$ | $a^{7}$ | $a^{8}$ | $a^{9}$ | $a^{11}$ | $a^{13}$ | $a^{15}$ |
| $\mathrm{z}_{1}$ | F | F | SP | F | SP | SP | SP | SP |
| $\mathrm{z}_{2}$ | F | F | S | SP | S | S | SP | SP |
| $\mathrm{z}_{3}$ | S | F | S | S | F | F | SP | SP |
| $\mathrm{z}_{4}$ | F | S | F | F | S | F | S | SP |
| $\mathrm{z}_{5}$ | S | F | SP | F | F | S | S | SP |
| $\mathrm{z}_{6}$ | F | SP | F | F | S | S | S | SP |
| $\mathrm{z}_{7}$ | S | F | F | S | F | SP | SP | SP |
| $\mathrm{z}_{8}$ | F | F | F | F | F | F | SP | SP |
| $\mathrm{z}_{9}$ | F | SP | S | F | S | F | SP | SP |
| $\mathrm{z}_{10}$ | S | F | F | S | F | S | S | S |
| $\mathrm{z}_{11}$ | F | S | F | S | S | S | SP | SP |
| $\mathrm{z}_{12}$ | F | S | F | F | F | S | SP | SP |
| $\mathrm{z}_{13}$ | F | F | F | S | S | F | SP | S |
| $\mathrm{z}_{14}$ | F | S | F | F | F | F | SP | SP |
| $\mathbf{z}_{15}$ | F | F | F | F | S | S | SP | SP |
| $\mathbf{z}_{16}$ | S | F | SP | S | S | S | SP | SP |
| $\mathrm{z}_{17}$ | F | F | F | F | S | S | S | S |
| $\mathrm{z}_{18}$ | F | F | SP | F | SP | S | S | SP |
| $\mathbf{z}_{19}$ | F | S | SP | S | F | S | SP | SP |
| $\mathrm{z}_{20}$ | F | F | F | F | F | S | SP | SP |
| $\mathrm{z}_{21}$ | S | SP | SP | F | F | F | S | SP |
| $\mathrm{z}_{22}$ | F | F | F | S | F | F | SP | SP |
| $\mathrm{z}_{23}$ | F | SP | F | S | F | S | F | SP |
| $\mathrm{z}_{24}$ | F | F | S | SP | S | F | SP | SP |
| $\mathrm{z}_{25}$ | F | F | S | F | F | S | SP | SP |
| $\mathbf{z}_{26}$ | F | F | F | S | S | S | SP | SP |
| $\mathrm{z}_{27}$ | S | SP | SP | F | F | S | S | F |
| $\mathrm{z}_{28}$ | S | S | S | F | F | SP | S | SP |
| $\mathbf{z}_{29}$ | F | F | S | F | F | F | SP | SP |
| $\mathrm{z}_{30}$ | F | F | F | F | F | F | S | SP |
| $\mathrm{z}_{31}$ | SP | F | S | F | F | S | SP | SP |
| $\mathrm{z}_{32}$ | F | F | SP | F | F | F | SP | S |
| $\mathrm{z}_{33}$ | F | F | F | S | S | S | S | S |

Table A.1: Labels assigned to chicken samples by panellists in storage experiment H4 described in Table 4.2 .

## Ranking

| Ranking | Session 1 |  | Session 2 |  |
| :---: | :---: | :---: | :---: | :---: |
|  | L4 | H4 | L4 | H4 |
| $\prec_{1}^{1}$ | $a^{9} \prec a^{5} \prec a^{7} \prec a^{0}$ | $a^{7} \prec a^{0} \prec a^{5} \prec a^{9}$ | $a^{15} \prec a^{13} \prec a^{11}$ | $a^{11} \prec a^{13} \prec a^{15}$ |
| $\prec \frac{1}{2}$ | $a^{9} \prec a^{7} \prec a^{5} \prec a^{0}$ | $a^{9} \prec a^{0} \prec a^{7} \prec a^{5}$ | $a^{15} \prec a^{13} \prec a^{11}$ | $a^{11} \prec a^{13} \prec a^{15}$ |
| $\prec \frac{1}{3}$ | $a^{9} \prec a^{5} \prec a^{7} \prec a^{0}$ | $a^{9} \prec a^{7} \prec a^{0} \prec a^{5}$ | $a^{15} \prec a^{11} \prec a^{13}$ | $a^{11} \prec a^{13} \prec a^{15}$ |
| $\prec_{4}^{1}$ |  | $a^{9} \prec a^{0} \prec a^{7} \prec a^{5}$ | $a^{15} \prec a^{13} \prec a^{11}$ |  |
| $\prec_{1}^{2}$ | $a^{7} \prec a^{11} \prec a^{5} \prec a^{0}$ | $a^{11} \prec a^{0} \prec a^{7} \prec a^{5}$ | $a^{15} \prec a^{9} \prec a^{13}$ | $a^{13} \prec a^{15} \prec a^{9}$ |
| $\prec_{2}^{2}$ | $a^{11} \prec a^{5} \prec a^{7} \prec a^{0}$ | $a^{11} \prec a^{5} \prec a^{7} \prec a^{0}$ | $a^{13} \prec a^{15} \prec a^{9}$ | $a^{13} \prec a^{15} \prec a^{9}$ |
| $\prec_{4}^{2}$ | $a^{11} \prec a^{5} \prec a^{0} \prec a^{7}$ | $a^{11} \prec a^{5} \prec a^{7} \prec a^{0}$ | $a^{13} \prec a^{15} \prec a^{9}$ | $a^{15} \prec a^{9} \prec a^{13}$ |
| $\prec_{4}^{2}$ | $a^{11} \prec a^{5} \prec a^{7} \prec a^{0}$ | $a^{11} \prec a^{0} \prec a^{7} \prec a^{5}$ | $a^{15} \prec a^{13} \prec a^{9}$ | $a^{13} \prec a^{9} \prec a^{15}$ |
| $\prec_{1}^{3}$ | $a^{9} \prec a^{15} \prec a^{0} \prec a^{7}$ | $a^{9} \prec a^{15} \prec a^{7} \prec a^{0}$ | $a^{11} \prec a^{13} \prec a^{5}$ | $a^{11} \prec a^{13} \prec a^{5}$ |
| $\prec_{2}^{3}$ | $a^{9} \prec a^{15} \prec a^{0} \prec a^{7}$ | $a^{7} \prec a^{15} \prec a^{0} \prec a^{9}$ | $a^{11} \prec a^{13} \prec a^{5}$ | $a^{11} \prec a^{13} \prec a^{5}$ |
| $\prec_{3}^{3}$ | $a^{9} \prec a^{7} \prec a^{15} \prec a^{0}$ |  | $a^{13} \prec a^{11} \prec a^{5}$ |  |
| $\prec_{1}^{4}$ | $a^{13} \prec a^{11} \prec a^{5} \prec a^{7}$ | $a^{11} \prec a^{13} \prec a^{7} \prec a^{5}$ | $a^{15} \prec a^{9} \prec a^{0}$ | $a^{9} \prec a^{15} \prec a^{0}$ |
| $\prec_{2}^{4}$ | $a^{13} \prec a^{11} \prec a^{5} \prec a^{7}$ | $a^{13} \prec a^{11} \prec a^{5} \prec a^{7}$ | $a^{15} \prec a^{9} \prec a^{0}$ | $a^{9} \prec a^{15} \prec a^{0}$ |
| $\prec_{3}^{4}$ | $a^{13} \prec a^{11} \prec a^{5} \prec a^{7}$ | $a^{11} \prec a^{13} \prec a^{7} \prec a^{5}$ | $a^{9} \prec a^{15} \prec a^{0}$ | $a^{0} \prec a^{9} \prec a^{15}$ |
| $\prec_{4}^{4}$ | $a^{13} \prec a^{11} \prec a^{5} \prec a^{7}$ | $a^{11} \prec a^{13} \prec a^{7} \prec a^{5}$ | $a^{0} \prec a^{9} \prec a^{15}$ | $a^{15} \prec a^{9} \prec a^{0}$ |

Table A.2: Rankings of chicken samples expressed by the panellists in each group (represented by the corresponding superindex) and for every session (1 and 2) in storage experiments L4 and H4 described in Table 4.3

| Ranking | L8 | H8 |
| :---: | :---: | :---: |
| $\prec_{1}$ | $a^{5} \prec a^{6} \prec a^{0} \prec a^{2} \prec a^{4}$ | $a^{6} \prec a^{5} \prec a^{0} \prec a^{2} \prec a^{4}$ |
| $\prec_{2}$ | $a^{5} \prec a^{6} \prec a^{4} \prec a^{2} \prec a^{0}$ | $a^{6} \prec a^{5} \prec a^{2} \prec a^{4} \prec a^{0}$ |
| $\prec_{3}$ | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ | $a^{6} \prec a^{0} \prec a^{5} \prec a^{4} \prec a^{2}$ |
| $\prec_{4}$ | $a^{6} \prec a^{2} \prec a^{5} \prec a^{4} \prec a^{0}$ | $a^{5} \prec a^{6} \prec a^{0} \prec a^{4} \prec a^{2}$ |
| $\prec_{5}$ | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ |
| $\prec_{6}$ | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ | $a^{4} \prec a^{5} \prec a^{6} \prec a^{0} \prec a^{2}$ |
| $\prec_{7}$ | $a^{6} \prec a^{5} \prec a^{4} \prec a^{2} \prec a^{0}$ | $a^{4} \prec a^{0} \prec a^{6} \prec a^{5} \prec a^{2}$ |
| $\prec_{8}$ | $a^{2} \prec a^{5} \prec a^{6} \prec a^{4} \prec a^{0}$ | $a^{5} \prec a^{6} \prec a^{0} \prec a^{2} \prec a^{4}$ |
| $\prec_{9}$ | $a^{5} \prec a^{6} \prec a^{0} \prec a^{4} \prec a^{2}$ | $a^{6} \prec a^{5} \prec a^{0} \prec a^{2} \prec a^{4}$ |
| $\prec_{10}$ | $a^{6} \prec a^{5} \prec a^{0} \prec a^{4} \prec a^{2}$ | $a^{6} \prec a^{5} \prec a^{2} \prec a^{0} \prec a^{4}$ |
| $\prec_{11}$ | $a^{2} \prec a^{5} \prec a^{4} \prec a^{6} \prec a^{0}$ | $a^{6} \prec a^{2} \prec a^{4} \prec a^{5} \prec a^{0}$ |
| $\prec_{12}$ | $a^{5} \prec a^{6} \prec a^{4} \prec a^{2} \prec a^{0}$ | $a^{6} \prec a^{4} \prec a^{5} \prec a^{2} \prec a^{0}$ |
| $\prec_{13}$ | $a^{5} \prec a^{6} \prec a^{0} \prec a^{4} \prec a^{2}$ | $a^{5} \prec a^{4} \prec a^{6} \prec a^{2} \prec a^{0}$ |
| $\prec_{14}$ | $a^{6} \prec a^{5} \prec a^{0} \prec a^{2} \prec a^{4}$ | $a^{6} \prec a^{0} \prec a^{2} \prec a^{4} \prec a^{5}$ |

Table A.3: Rankings of chicken samples expressed by all the 14 panellists in storage experiments L8 and H8 described in Table 4.3

## A.2. Sensory evaluation for Atlantic cod

## Labelling

| List of labels | Company A |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | L4 |  |  |  | L8 |  |  |  | H4 |  |  |  |
|  | $a^{0}$ | $a^{4}$ | $a^{8}$ | $a^{13}$ | $a^{0}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ | $a^{0}$ | $a^{4}$ | $a^{8}$ | $a^{13}$ |
| $\mathrm{z}_{1}$ | VF | SP | S | M | F | F | M | S | VF | F | S | SP |
| $\mathrm{z}_{2}$ | VF | SP | S | SP | M | SP | M | SP | VF | F | M | SP |
| $\mathrm{z}_{3}$ | F | M | S | SP | VF | SP | SP | SP | VF | S | S | M |
| $\mathrm{z}_{4}$ | VF | S | F | M | VF | M | S | M | VF | S | M | M |
| $\mathrm{z}_{5}$ | VF | S | F | M | VF | SP | SP | SP | VF | S | SP | M |
| $\mathrm{z}_{6}$ | F | S | M | M | VF | F | SP | S | VF | S | SP | SP |
| $\mathbf{z}_{7}$ | F | M | S | S | VF | M | F | SP | VF | S | SP | SP |
| $\mathrm{z}_{8}$ | VF | F | SP | M | VF | S | M | S | VF | M | F | S |
| $\mathrm{z}_{9}$ |  |  |  |  | F | VF | M | SP | F | SP | S | SP |
| $\mathrm{z}_{10}$ |  |  |  |  | VF | F | S | S | S | F | M | SP |

Table A.4: Labels assigned by panellists to cod samples from company A in storage experiments L4, L8 and H4 described in Table 4.5 .

| Experiment | Samples | Company A |  |  |  |  | Company B |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | SP | M | S | F | VF | SP | M | S | F | VF |
| H4 | $a^{0}$ | $\frac{8}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | 0 | 0 | $\frac{6}{10}$ | $\frac{3}{10}$ | 0 | $\frac{1}{10}$ | 0 |
|  | $a^{4}$ | 0 | $\frac{3}{10}$ | $\frac{5}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{6}{10}$ | $\frac{2}{10}$ | $\frac{1}{10}$ | 0 |
|  | $a^{8}$ | 0 | $\frac{1}{10}$ | $\frac{3}{10}$ | $\frac{3}{10}$ | $\frac{3}{10}$ | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{2}{10}$ | $\frac{2}{10}$ | $\frac{4}{10}$ |
|  | $a^{13}$ | 0 | 0 | $\frac{1}{10}$ | $\frac{3}{10}$ | $\frac{6}{10}$ | 0 | $\frac{1}{10}$ | $\frac{2}{10}$ | $\frac{3}{10}$ | $\frac{4}{10}$ |
| H8 | $a^{0}$ | $\frac{8}{12}$ | $\frac{3}{12}$ | $\frac{1}{12}$ | 0 | 0 | $\frac{8}{10}$ | $\frac{2}{10}$ | 0 | 0 | 0 |
|  | $a^{3}$ | 0 | $\frac{3}{12}$ | $\frac{2}{12}$ | $\frac{4}{12}$ | $\frac{3}{12}$ | 0 | $\frac{5}{10}$ | $\frac{3}{10}$ | $\frac{2}{10}$ | 0 |
|  | $a^{5}$ | 0 | $\frac{1}{12}$ | $\frac{2}{12}$ | $\frac{4}{12}$ | $\frac{5}{12}$ | 0 | 0 | $\frac{7}{10}$ | $\frac{3}{10}$ | 0 |
|  | $a^{7}$ | $\frac{1}{12}$ | 0 | $\frac{13}{12}$ | $\frac{5}{12}$ | $\frac{3}{12}$ | 0 | $\frac{2}{10}$ | 0 | $\frac{3}{10}$ | $\frac{5}{10}$ |
| A4 | $a^{0}$ | $\frac{2}{10}$ | $\frac{3}{10}$ | $\frac{4}{10}$ | $\frac{1}{10}$ | 0 | $\frac{4}{10}$ | $\frac{5}{10}$ | $\frac{1}{10}$ | 0 | 0 |
|  | $a^{1}$ | $\frac{5}{10}$ | $\frac{4}{10}$ | 0 | $\frac{1}{10}$ | 0 | $\frac{5}{10}$ | $\frac{2}{10}$ | $\frac{1}{10}$ | $\frac{2}{10}$ | 0 |
|  | $a^{2}$ | $\frac{1}{10}$ | $\frac{3}{10}$ | $\frac{3}{10}$ | $\frac{3}{10}$ | 0 | 0 | $\frac{1}{10}$ | $\frac{3}{10}$ | $\frac{5}{10}$ | $\frac{1}{10}$ |
|  | $a^{3}$ | 0 | 0 | $\frac{4}{10}$ | $\frac{5}{10}$ | $\frac{1}{10}$ | 0 | $\frac{1}{10}$ | $\frac{1}{10}$ | $\frac{2}{10}$ | $\frac{6}{10}$ |
| L4 | $a^{0}$ | $\frac{5}{8}$ | $\frac{2}{8}$ | 0 | 0 | $\frac{1}{8}$ |  |  |  |  |  |
|  | $a^{4}$ | 0 | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{2}{8}$ | $\frac{2}{8}$ |  |  |  |  |  |
|  | $a^{8}$ | 0 | $\frac{2}{8}$ | $\frac{3}{8}$ | $\frac{2}{8}$ | $\frac{1}{8}$ |  |  |  |  |  |
|  | $a^{13}$ | 0 | 0 | $\frac{2}{8}$ | $\frac{5}{8}$ | $\frac{1}{8}$ |  |  |  |  |  |
| L8 | $a^{0}$ | $\frac{7}{10}$ | $\frac{2}{10}$ | 0 | $\frac{1}{10}$ | 0 |  |  |  |  |  |
|  | $a^{3}$ | $\frac{1}{10}$ | $\frac{3}{10}$ | $\frac{1}{10}$ | $\frac{2}{10}$ | $\frac{3}{10}$ |  |  |  |  |  |
|  | $a^{5}$ | 0 | $\frac{1}{10}$ | $\frac{2}{10}$ | $\frac{4}{10}$ | $\frac{3}{10}$ |  |  |  |  |  |
|  | $a^{7}$ | 0 | 0 | $\frac{4}{10}$ | $\frac{1}{10}$ | $\frac{5}{10}$ |  |  |  |  |  |

Table A.5: Relative frequency of assigned labels to cod samples in each storage experiment described in Table 4.5 .

## Ranking

## Session 1

| Ranking | Session 1 |  |  |
| :---: | :---: | :---: | :---: |
|  | L4 | H4 | H8 |
| $\prec_{1}$ | $a^{13} \prec a^{4} \prec a^{8} \prec a^{0}$ | $a^{8} \prec a^{13} \prec a^{6} \prec a^{4}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |
| $\prec_{2}$ | $a^{13} \prec a^{8} \prec a^{0} \prec a^{4}$ | $a^{8} \prec a^{6} \prec a^{13} \prec a^{4}$ | $a^{5} \prec a^{7} \prec a^{0} \prec a^{3}$ |
| $\prec_{3}$ | $a^{13} \prec a^{8} \prec a^{4} \prec a^{0}$ | $a^{8} \prec a^{13} \prec a^{4} \prec a^{6}$ | $a^{5} \prec a^{7} \prec a^{3} \prec a^{0}$ |
| $\prec_{4}$ | $a^{8} \prec a^{13} \prec a^{4} \prec a^{0}$ | $a^{6} \prec a^{13} \prec a^{8} \prec a^{4}$ | $a^{5} \prec a^{7} \prec a^{3} \prec a^{0}$ |
| $\prec_{5}$ | $a^{13} \prec a^{8} \prec a^{0} \prec a^{4}$ | $a^{8} \prec a^{13} \prec a^{6} \prec a^{4}$ | $a^{5} \prec a^{3} \prec a^{7} \prec a^{0}$ |
| $\prec_{6}$ | $a^{13} \prec a^{8} \prec a^{4} \prec a^{0}$ | $a^{8} \prec a^{13} \prec a^{6} \prec a^{4}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |
| $\prec_{7}$ | $a^{13} \prec a^{8} \prec a^{4} \prec a^{0}$ | $a^{8} \prec a^{13} \prec a^{6} \prec a^{4}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |
| $\prec_{8}$ | $a^{13} \prec a^{8} \prec a^{0} \prec a^{4}$ | $a^{13} \prec a^{8} \prec a^{6} \prec a^{4}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |
| $\prec_{9}$ | $a^{13} \prec a^{8} \prec a^{0} \prec a^{4}$ |  |  |
| $\prec_{10}$ | $a^{13} \prec a^{8} \prec a^{4} \prec a^{0}$ |  |  |
|  | Session 2 |  |  |
| Ranking | L4 | H4 | H8 |
| $\prec_{1}$ | $a^{8} \prec a^{7} \prec a^{6} \prec a^{5}$ | $a^{11} \prec a^{6} \prec a^{7} \prec a^{8}$ | $a^{5} \prec a^{6} \prec a^{4} \prec a^{3}$ |
| $\prec_{2}$ | $a^{8} \prec a^{6} \prec a^{5} \prec a^{7}$ | $a^{8} \prec a^{11} \prec a^{7} \prec a^{6}$ | $a^{5} \prec a^{6} \prec a^{3} \prec a^{4}$ |
| $\prec_{3}$ | $a^{8} \prec a^{5} \prec a^{7} \prec a^{6}$ | $a^{11} \prec a^{6} \prec a^{8} \prec a^{7}$ | $a^{5} \prec a^{6} \prec a^{3} \prec a^{4}$ |
| $\prec_{4}$ | $a^{5} \prec a^{8} \prec a^{6} \prec a^{7}$ | $a^{11} \prec a^{7} \prec a^{8} \prec a^{6}$ | $a^{4} \prec a^{5} \prec a^{6} \prec a^{3}$ |
| $\prec_{5}$ | $a^{8} \prec a^{7} \prec a^{6} \prec a^{5}$ | $a^{7} \prec a^{11} \prec a^{8} \prec a^{6}$ | $a^{6} \prec a^{5} \prec a^{3} \prec a^{4}$ |
| $\prec_{6}$ | $a^{8} \prec a^{5} \prec a^{7} \prec a^{6}$ | $a^{11} \prec a^{8} \prec a^{6} \prec a^{7}$ | $a^{5} \prec a^{3} \prec a^{6} \prec a^{4}$ |
| $\prec_{7}$ | $a^{8} \prec a^{7} \prec a^{6} \prec a^{5}$ | $a^{11} \prec a^{7} \prec a^{8} \prec a^{6}$ | $a^{5} \prec a^{6} \prec a^{4} \prec a^{3}$ |
| $\prec_{8}$ | $a^{8} \prec a^{7} \prec a^{6} \prec a^{5}$ | $a^{11} \prec a^{8} \prec a^{7} \prec a^{6}$ | $a^{5} \prec a^{4} \prec a^{6} \prec a^{3}$ |
| $\prec_{9}$ |  | $a^{11} \prec a^{8} \prec a^{7} \prec a^{6}$ |  |
| Ranking | L8 | A4 |  |
| $\prec_{1}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{3} \prec a^{0} \prec a^{1} \prec a^{2}$ |  |
| $\prec_{2}$ | $a^{0} \prec a^{3} \prec a^{5} \prec a^{7}$ | $a^{3} \prec a^{2} \prec a^{0} \prec a^{1}$ |  |
| $\prec_{3}$ | $a^{7} \prec a^{3} \prec a^{5} \prec a^{0}$ | $a^{0} \prec a^{1} \prec a^{3} \prec a^{2}$ |  |
| $\prec_{4}$ | $a^{5} \prec a^{0} \prec a^{3} \prec a^{7}$ | $a^{3} \prec a^{2} \prec a^{0} \prec a^{1}$ |  |
| $\prec_{5}$ | $a^{5} \prec a^{7} \prec a^{3} \prec a^{0}$ | $a^{3} \prec a^{2} \prec a^{0} \prec a^{1}$ |  |
| $\prec_{6}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{3} \prec a^{1} \prec a^{2} \prec a^{0}$ |  |
| $\prec_{7}$ | $a^{5} \prec a^{7} \prec a^{3} \prec a^{0}$ | $a^{3} \prec a^{0} \prec a^{2} \prec a^{1}$ |  |
| $\prec_{8}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{3} \prec a^{2} \prec a^{1} \prec a^{0}$ |  |
| $\prec_{9}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |  |  |

Table A.6: Rankings of cod samples expressed by the panellists in each storage experiment described in Table 4.6

## A.3. Sensory evaluation for Atlantic brown shrimp

## Labelling

| List of labels | Session 1 |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | L4 |  |  |  | H4 |  |  |  |
|  | $a^{0}$ | $a^{3}$ | $a^{5}$ | $a^{10}$ | $a^{0}$ | $a^{3}$ | $a^{7}$ | $a^{12}$ |
| $\mathrm{z}_{1}$ | M | F | F | S | F | F | M | SP |
| $\mathrm{z}_{2}$ | F | S | F | M | F | F | SP | SP |
| $\mathrm{z}_{3}$ | F | S | M | SP | VF | F | SP | M |
| $\mathrm{z}_{4}$ | S | F | VF | M | S | F | M | SP |
| $\mathrm{z}_{5}$ | VF | F | S | SP | M | S | SP | SP |
| $\mathrm{z}_{6}$ | VF | VF | S | M | F | F | SP | SP |
| $\mathrm{z}_{7}$ | F | VF | VF | SP | M | F | SP | SP |
| $\mathrm{z}_{8}$ | VF | VF | S | SP | VF | M | SP | SP |
| $\mathrm{z}_{9}$ | F | VF | S | SP | F | VF | SP | SP |
| $\mathbf{z}_{10}$ | VF | F | S | M |  |  |  |  |
| List of labels | Session 1 |  |  |  |  |  |  |  |
|  | L4 |  |  |  | H4 |  |  |  |
|  | $a^{0}$ | $a^{3}$ | $a^{5}$ | $a^{10}$ | $a^{0}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ |
| $\mathrm{z}_{1}$ | VF | S | F | SP | F | F | SP | SP |
| $\mathrm{z}_{2}$ | VF | S | S | SP | S | F | SP | SP |
| $\mathrm{z}_{3}$ | M | S | F | SP | VF | S | M | M |
| $\mathrm{z}_{4}$ | F | F | F | S | VF | F | SP | M |
| $\mathrm{z}_{5}$ | F | S | VF | M | S | F | M | M |
| $\mathrm{z}_{6}$ | F | M | S | SP | VF | F | S | M |
| $\mathrm{z}_{7}$ | F | F | F | SP | VF | F | SP | M |
| $\mathrm{z}_{8}$ | VF | VF | VF | SP | M | S | M | SP |
| $\mathrm{z}_{9}$ | S | F | M | SP | VF | VF | S | F |
| $\mathbf{z}_{10}$ | F | F | S | SP | S | F | SP | SP |

Table A.7: Labels assigned by panellists to shrimp samples in each storage experiment described in Table 4.8 .

## Ranking

| Ranking | Session 1 |  |
| :---: | :---: | :---: |
|  | L4 | H4 |
| $\prec_{1}$ | $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{12} \prec a^{3} \prec a^{7} \prec a^{0}$ |
| $\prec_{2}$ | $a^{10} \prec a^{5} \prec a^{0} \prec a^{3}$ | $a^{12} \prec a^{7} \prec a^{3} \prec a^{0}$ |
| $\prec_{3}$ | $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{7} \prec a^{12} \prec a^{0} \prec a^{3}$ |
| $\prec_{4}$ | $a^{10} \prec a^{5} \prec a^{0} \prec a^{3}$ | $a^{12} \prec a^{3} \prec a^{7} \prec a^{0}$ |
| $\prec_{5}$ | $a^{10} \prec a^{3} \prec a^{5} \prec a^{0}$ | $a^{7} \prec a^{12} \prec a^{3} \prec a^{0}$ |
| $\prec_{6}$ | $a^{10} \prec a^{3} \prec a^{5} \prec a^{0}$ | $a^{7} \prec a^{12} \prec a^{3} \prec a^{0}$ |
| $\prec_{7}$ | $a^{10} \prec a^{5} \prec a^{0} \prec a^{3}$ | $a^{12} \prec a^{7} \prec a^{3} \prec a^{0}$ |
| $\prec_{8}$ | $a^{10} \prec a^{0} \prec a^{5} \prec a^{3}$ | $a^{12} \prec a^{7} \prec a^{3} \prec a^{0}$ |
| $\prec_{9}$ | $a^{10} \prec a^{3} \prec a^{0} \prec a^{5}$ | $a^{12} \prec a^{7} \prec a^{3} \prec a^{0}$ |
|  | Session 2 |  |
| Ranking | L4 | H4 |
| $\prec_{1}$ | $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{5} \prec a^{7} \prec a^{0} \prec a^{3}$ |
| $\prec_{2}$ | $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |
| $\prec_{3}$ | $a^{10} \prec a^{5} \prec a^{0} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |
| $\prec_{4}$ | $a^{10} \prec a^{5} \prec a^{0} \prec a^{3}$ | $a^{7} \prec a^{0} \prec a^{5} \prec a^{3}$ |
| $\prec_{5}$ | $a^{10} \prec a^{0} \prec a^{3} \prec a^{5}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |
| $\prec_{6}$ | $a^{10} \prec a^{5} \prec a^{3} \prec a^{0}$ | $a^{5} \prec a^{7} \prec a^{3} \prec a^{0}$ |
| $\prec_{7}$ | $a^{10} \prec a^{3} \prec a^{0} \prec a^{5}$ | $a^{5} \prec a^{7} \prec a^{0} \prec a^{3}$ |
| $\prec_{8}$ | $a^{10} \prec a^{0} \prec a^{3} \prec a^{5}$ | $a^{5} \prec a^{7} \prec a^{3} \prec a^{0}$ |
| $\prec 9$ | $a^{10} \prec a^{0} \prec a^{3} \prec a^{5}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{0}$ |
| $\prec_{10}$ | $a^{10} \prec a^{3} \prec a^{5} \prec a^{0}$ | $a^{7} \prec a^{0} \prec a^{5} \prec a^{3}$ |

Table A.8: Rankings of shrimp samples expressed by the panellists in each experiment described in Table 4.9 .

## A.4. Sensory evaluation for Atlantic salmon

## Labelling

| List of labels | Session 1 |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | H4 |  |  |  | AN4 |  |  |  | ANH4 |  |  |  |
|  | $a^{1}$ | $a^{5}$ | $a^{7}$ | $a^{11}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ |
| $\mathrm{z}_{1}$ | VF | F | S | SP | VF | S | M | SP | VF | M | SP | S |
| $\mathrm{z}_{2}$ | VF | F | VF | F | VF | F | M | SP | F | SP | M | SP |
| $\mathrm{z}_{3}$ | F | M | S | SP | VF | F | VF | S | S | S | SP | SP |
| $\mathrm{z}_{4}$ | VF | F | M | SP | VF | M | S | SP | VF | M | M | SP |
| $\mathrm{z}_{5}$ | VF | S | M | SP | VF | F | M | SP | VF | F | SP | S |
| $\mathrm{z}_{6}$ | F | VF | M | SP | F | M | M | SP | M | M | SP | SP |
| $\mathrm{z}_{7}$ | VF | VF | M | SP | F | M | S | SP | VF | F | SP | S |
| $\mathrm{z}_{8}$ | F | S | M | SP | VF | S | M | SP | F | VF | M | SP |
| $\mathrm{z}_{9}$ |  |  |  |  | F | S | M | SP | F | M | SP | SP |
| $\mathrm{z}_{10}$ |  |  |  |  | F | SP | S | SP |  |  |  |  |
| $\mathrm{z}_{11}$ |  |  |  |  | F | M | F | SP |  |  |  |  |
| $\mathrm{z}_{12}$ |  |  |  |  | M | SP | S | SP |  |  |  |  |
|  | Session 1 |  |  |  |  |  |  |  |  |  |  |  |
|  | A4 |  |  |  | L4 |  |  |  | M4 |  |  |  |
| List of labels | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{13}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ |
| $\mathrm{z}_{1}$ | F | S | SP | M | VF | VF | S | M | VF | VF | SP | SP |
| $\mathrm{z}_{2}$ | VF | F | S | M | S | M | SP | SP | VF | F | SP | SP |
| $\mathrm{z}_{3}$ | VF | S | SP | M | F | VF | S | M | F | M | M | S |
| $\mathrm{z}_{4}$ | S | SP | SP | SP | VF | F | S | M | F | S | M | S |
| $\mathrm{z}_{5}$ | VF | S | M | M | S | SP | F | M | F | F | M | M |
| $\mathrm{z}_{6}$ | F | S | M | M | F | VF | S | M | F | F | S | S |
| $\mathbf{z}_{7}$ | VF | S | M | SP | F | S | S | M | VF | VF | SP | M |
| $\mathrm{z}_{8}$ | VF | F | SP | S | F | M | M | SP | F | VF | SP | M |
| $\mathrm{z}_{9}$ | S | M | SP | SP | M | S | M | SP |  |  |  |  |
| $\mathbf{z}_{10}$ | F | SP | SP | S |  |  |  |  |  |  |  |  |
| $\mathrm{z}_{11}$ | VF | S | SP | M |  |  |  |  |  |  |  |  |

Table A.9: Labels assigned by panellists to salmon samples in each storage experiment described in Table 4.11

| List of labels | Session 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | H4 |  |  |  | AN4 |  |  |  | ANH4 |  |  |  | A4 |  |  |  |
|  | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ | $a^{3}$ | $a^{7}$ | $a^{9}$ | $a^{11}$ | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ |
| $\mathrm{z}_{1}$ | S | F | S | M | VF | SP | M | SP | F | F | S | M | VF | F | S | SP |
| $\mathbf{z}_{2}$ | VF | F | F | S | S | M | M | SP | F | VF | M | M | SP | F | SP | SP |
| $\mathbf{z}_{3}$ | VF | VF | S | S | VF | M | F | SP | M | VF | S | SP | F | S | M | SP |
| $\mathrm{z}_{4}$ | VF | VF | VF | M | VF | M | SP | SP | F | F | F | SP | VF | F | S | M |
| $\mathbf{z}_{5}$ | VF | S | F | M | VF | S | S | M | S | F | F | M | VF | F | M | SP |
| $\mathrm{z}_{6}$ | VF | F | S | M | F | M | SP | SP |  |  |  |  | VF | VF | M | SP |
| $\mathbf{z}_{7}$ | F | VF | F | M | VF | M | M | SP |  |  |  |  | F | S | S | SP |
| $\mathbf{z}_{8}$ | VF | F | S | S | F | M | S | M |  |  |  |  | F | F | M | SP |
| $\mathrm{z}_{9}$ |  |  |  |  | VF | M | M | M |  |  |  |  | VF | VF | M | SP |
| $\mathbf{z}_{10}$ |  |  |  |  | F | SP | M | S |  |  |  |  | VF | VF | F | SP |
| $\mathbf{z}_{11}$ |  |  |  |  | S | SP | S | SP |  |  |  |  |  |  |  |  |
| $\mathrm{z}_{12}$ |  |  |  |  | VF | M | S | S |  |  |  |  |  |  |  |  |
| Session 3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| List of | H4 |  |  |  | AN4 |  |  |  | ANH4 |  |  |  | A4 |  |  |  |
| labels | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ | $a^{1}$ | $a^{5}$ | $a^{9}$ | $a^{11}$ |
| $\mathrm{z}_{1}$ | VF | VF | M | M | VF | S | M | M | VF | F | M | S | F | M | SP | SP |
| $\mathrm{z}_{2}$ | VF | F | SP | M | F | SP | M | SP | VF | VF | VF | M | VF | F | S | M |
| $\mathrm{z}_{3}$ | F | F | M | M | VF | F | M | S | VF | F | SP | SP | VF | VF | S | M |
| $\mathrm{z}_{4}$ | F | VF | S | M | S | M | S | VF | VF | F | S | S | VF | M | SP | SP |
| $\mathrm{z}_{5}$ | S | VF | F | SP | S | SP | F | M | M | S | SP | SP | F | SP | S | SP |
| $\mathrm{z}_{6}$ | VF | VF | S | S | F | F | M | SP | VF | S | M | M | S | F | M | M |
| $\mathrm{z}_{7}$ | VF | VF | SP | M | VF | F | M | SP | VF | F | S | SP | VF | F | M | F |
| $\mathrm{z}_{8}$ | F | VF | M | F | VF | F | SP | SP | VF | F | S | F | VF | VF | M | F |
| $\mathrm{z}_{9}$ | F | F | M | M |  |  |  |  |  |  |  |  | VF | S | M | SP |
| Session 4 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| List of | H4 |  |  |  | AN4 |  |  |  | ANH4 |  |  |  | A4 |  |  |  |
| labels | $a^{3}$ | $a^{5}$ | $a^{7}$ | $a^{9}$ | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ | $a^{1}$ | $a^{3}$ | $a^{5}$ | $a^{7}$ |
| $\mathrm{z}_{1}$ | S | F | F | S | F | VF | M | SP | VF | VF | VF | VF | F | F | SP | SP |
| $\mathrm{z}_{2}$ | S | F | M | SP | F | S | SP | M | F | F | F | F | F | VF | M | SP |
| $\mathbf{z}_{3}$ | F | S | S | S | VF | VF | S | S | VF | F | F | VF | F | VF | S | SP |
| $\mathrm{z}_{4}$ | M | F | VF | M | VF | VF | SP | SP | S | F | F | VF | VF | VF | VF | SP |
| $\mathrm{z}_{5}$ | VF | VF | F | M | F | F | S | SP | S | VF | F | M | S | F | F | SP |
| $\mathbf{z}_{6}$ | M | F | M | F | F | VF | M | S | M | F | M | S | F | VF | VF | S |
| $\mathbf{z}_{7}$ | M | S | F | M | VF | VF | S | F | VF | VF | VF | VF | F | F | M | M |
| $\mathrm{z}_{8}$ | F | F | S | S | VF | VF | S | F | VF | VF | F | F | S | F | S | M |
| $\mathrm{z}_{9}$ | VF | VF | VF | F | VF | F | SP | M | VF | VF | F | S | F | F | S | SP |
| $\mathbf{z}_{10}$ | F | F | F | F |  |  |  |  |  |  |  |  |  |  |  |  |

Table A.9: (Continued) Labels assigned by panellists to salmon samples in each storage experiment described in Table 4.11 .

## Ranking

| Ranking | Session 1 |  |  |
| :---: | :---: | :---: | :---: |
|  | H4 | AN4 | ANH4 |
| $\prec_{1}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ | $a^{9} \prec a^{5} \prec a^{1} \prec a^{13}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ |
| $\prec_{2}$ | $a^{11} \prec a^{9} \prec a^{1} \prec a^{5}$ | $a^{13} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| $\prec_{3}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{9} \prec a^{13} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| $\prec_{4}$ | $a^{11} \prec a^{9} \prec a^{1} \prec a^{5}$ | $a^{13} \prec a^{9} \prec a^{1} \prec a^{5}$ | $a^{11} \prec a^{5} \prec a^{1} \prec a^{9}$ |
| $\prec_{5}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ | $a^{13} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{1} \prec a^{5}$ |
| $\prec_{6}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{9} \prec a^{5} \prec a^{13} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| $\prec_{7}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{13} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{9} \prec a^{11} \prec a^{1} \prec a^{5}$ |
| $\prec_{8}$ | $a^{9} \prec a^{11} \prec a^{1} \prec a^{5}$ | $a^{9} \prec a^{13} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| $\prec 9$ | $a^{9} \prec a^{11} \prec a^{1} \prec a^{5}$ | $a^{13} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{1} \prec a^{5}$ |
| $\prec_{10}$ |  | $a^{9} \prec a^{13} \prec a^{5} \prec a^{1}$ |  |
| $\prec_{11}$ |  | $a^{9} \prec a^{13} \prec a^{5} \prec a^{1}$ |  |
| $\prec_{12}$ |  | $a^{13} \prec a^{5} \prec a^{9} \prec a^{1}$ |  |
|  | Session 1 |  |  |
| Ranking | A4 | L4 | M4 |
| $\prec_{1}$ | $a^{11} \prec a^{5} \prec a^{9} \prec a^{1}$ | $a^{7} \prec a^{11} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{1} \prec a^{5}$ |
| $\prec_{2}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{7} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| $\prec_{3}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{7} \prec a^{1} \prec a^{5}$ | $a^{11} \prec a^{9} \prec a^{1} \prec a^{5}$ |
| $\prec_{4}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{5} \prec a^{1} \prec a^{7}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ |
| $\prec_{5}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{7} \prec a^{11} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| $\prec_{6}$ | $a^{9} \prec a^{5} \prec a^{11} \prec a^{1}$ | $a^{11} \prec a^{7} \prec a^{1} \prec a^{5}$ | $a^{9} \prec a^{1} \prec a^{11} \prec a^{5}$ |
| $\prec_{7}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{5} \prec a^{7} \prec a^{1} \prec a^{11}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| $\prec_{8}$ | $a^{11} \prec a^{9} \prec a^{1} \prec a^{5}$ | $a^{7} \prec a^{11} \prec a^{1} \prec a^{5}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| $\prec_{9}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{1} \prec a^{7} \prec a^{5}$ |  |
| $\prec_{10}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ |  |  |

Table A.10: Rankings of salmon samples expressed by the panellists in each experiment described in Table 4.12.

Session 2

| Ranking | Session 2 |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | H4 | AN4 | ANH4 | A4 |
| $\prec_{1}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ | $a^{7} \prec a^{3} \prec a^{1} \prec a^{5}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| $\prec_{2}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ | $a^{3} \prec a^{5} \prec a^{1} \prec a^{7}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| $\prec_{3}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ | $a^{7} \prec a^{1} \prec a^{3} \prec a^{5}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| $\prec_{4}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ | $a^{7} \prec a^{3} \prec a^{1} \prec a^{5}$ | $a^{7} \prec a^{3} \prec a^{5} \prec a^{1}$ |
| $\prec_{5}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{11} \prec a^{7} \prec a^{9} \prec a^{3}$ | $a^{1} \prec a^{5} \prec a^{3} \prec a^{7}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ |
| $\prec_{6}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{7} \prec a^{9} \prec a^{11} \prec a^{3}$ |  | $a^{7} \prec a^{3} \prec a^{1} \prec a^{5}$ |
| $\prec_{7}$ | $a^{7} \prec a^{3} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ |  | $a^{7} \prec a^{3} \prec a^{1} \prec a^{5}$ |
| $\prec_{8}$ | $a^{7} \prec a^{3} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ |  | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ |
| $\prec_{9}$ |  | $a^{11} \prec a^{7} \prec a^{9} \prec a^{3}$ |  | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| $\prec_{10}$ |  | $a^{9} \prec a^{7} \prec a^{11} \prec a^{3}$ |  |  |
| $\prec_{11}$ |  | $a^{11} \prec a^{7} \prec a^{9} \prec a^{3}$ |  |  |
| $\prec_{12}$ |  | $a^{11} \prec a^{9} \prec a^{7} \prec a^{3}$ |  |  |

Session 3

| Ranking | H4 | AN4 | ANH4 | A4 |
| :---: | :---: | :---: | :---: | :---: |
| $\prec_{1}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{1} \prec a^{5}$ |
| $\prec_{2}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{5} \prec a^{9} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| $\prec_{3}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{1} \prec a^{5}$ | $a^{1} \prec a^{11} \prec a^{9} \prec a^{5}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ |
| $\prec_{4}$ | $a^{5} \prec a^{9} \prec a^{11} \prec a^{1}$ | $a^{1} \prec a^{5} \prec a^{9} \prec a^{11}$ | $a^{5} \prec a^{1} \prec a^{9} \prec a^{11}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |
| $\prec_{5}$ | $a^{11} \prec a^{9} \prec a^{1} \prec a^{5}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{5} \prec a^{1} \prec a^{9}$ | $a^{9} \prec a^{5} \prec a^{11} \prec a^{1}$ |
| $\prec_{6}$ | $a^{1} \prec a^{11} \prec a^{9} \prec a^{5}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{5} \prec a^{9} \prec a^{1}$ | $a^{9} \prec a^{11} \prec a^{5} \prec a^{1}$ |
| $\prec_{7}$ | $a^{11} \prec a^{1} \prec a^{5} \prec a^{9}$ | $a^{9} \prec a^{11} \prec a^{1} \prec a^{5}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ | $a^{11} \prec a^{9} \prec a^{1} \prec a^{5}$ |
| $\prec_{8}$ | $a^{11} \prec a^{9} \prec a^{5} \prec a^{1}$ |  | $a^{11} \prec a^{5} \prec a^{1} \prec a^{9}$ | $a^{5} \prec a^{9} \prec a^{11} \prec a^{1}$ |
| $\prec_{9}$ | $a^{5} \prec a^{9} \prec a^{11} \prec a^{1}$ |  | $a^{9} \prec a^{5} \prec a^{11} \prec a^{1}$ |  |

Session 4

| Ranking | H4 | AN4 | ANH4 | A4 |
| :---: | :---: | :---: | :---: | :---: |
| $\prec_{1}$ | $a^{9} \prec a^{7} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{7} \prec a^{3} \prec a^{5} \prec a^{1}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ |
| $\prec_{2}$ | $a^{9} \prec a^{7} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{7} \prec a^{3} \prec a^{5} \prec a^{1}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| $\prec_{3}$ | $a^{9} \prec a^{7} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ | $a^{7} \prec a^{1} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| $\prec_{4}$ | $a^{9} \prec a^{3} \prec a^{5} \prec a^{7}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ | $a^{7} \prec a^{3} \prec a^{5} \prec a^{1}$ |
| $\prec_{5}$ | $a^{7} \prec a^{9} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{3} \prec a^{5} \prec a^{1}$ | $a^{1} \prec a^{3} \prec a^{7} \prec a^{5}$ | $a^{7} \prec a^{5} \prec a^{1} \prec a^{3}$ |
| $\prec_{6}$ | $a^{9} \prec a^{7} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{1} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ | $a^{7} \prec a^{3} \prec a^{5} \prec a^{1}$ |
| $\prec_{7}$ | $a^{9} \prec a^{7} \prec a^{3} \prec a^{5}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ | $a^{7} \prec a^{1} \prec a^{5} \prec a^{3}$ | $a^{3} \prec a^{5} \prec a^{7} \prec a^{1}$ |
| $\prec_{8}$ | $a^{9} \prec a^{7} \prec a^{3} \prec a^{5}$ | $a^{7} \prec a^{1} \prec a^{5} \prec a^{3}$ | $a^{7} \prec a^{1} \prec a^{3} \prec a^{5}$ | $a^{5} \prec a^{7} \prec a^{1} \prec a^{3}$ |
| $\prec_{9}$ | $a^{9} \prec a^{7} \prec a^{3} \prec a^{5}$ |  | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ | $a^{7} \prec a^{5} \prec a^{3} \prec a^{1}$ |
| $\prec_{10}$ | $a^{9} \prec a^{5} \prec a^{7} \prec a^{3}$ |  |  |  |

Table A.10: (Continued) Rankings of salmon samples expressed by the panellists in each experiment described in Table 4.12 .

## Scoring

| Vector of scores | Group 1 <br> Tuesday |  |  |  | Group 2 <br> Thursday |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{A}^{1}$ | $\mathrm{B}^{2}$ | $\mathrm{C}^{3}$ | $\mathrm{D}^{4}$ | $\mathrm{A}^{2}$ | $\mathrm{B}^{3}$ | $\mathrm{C}^{4}$ | $\mathrm{D}^{5}$ |  |  |  |  |
| $\mathrm{S}_{1}$ | 4 | 4 | 3 | 2 | 5 | 4 | 5 | 3 |  |  |  |  |
| $\mathrm{s}_{2}$ | 5 | 4 | 3 | 4 | 4 | 4 | 4 | 4 |  |  |  |  |
| $\mathrm{s}_{3}$ | 5 | 4 | 3 | 3 | 4 | 4 | 4 | 2 |  |  |  |  |
| $\mathrm{s}_{4}$ | 3 | 3 | 1 | 2 | 4 | 4 | 3 | 2 |  |  |  |  |
| $\mathrm{s}_{5}$ | 5 | 4 | 3 | 4 | 5 | 4 | 3 | 2 |  |  |  |  |
| $\mathrm{s}_{6}$ | 5 | 5 | 5 | 4 | 5 | 3 | 4 | 2 |  |  |  |  |
| $\mathrm{s}_{7}$ | 4 | 4 | 3 | 4 | 5 | 4 | 4 | 2 |  |  |  |  |
| $\mathrm{s}_{8}$ | 3 | 4 | 2 | 4 | 5 | 5 | 3 | 2 |  |  |  |  |
| $\mathrm{s}_{9}$ | 5 | 5 | 5 | 4 | 4 | 5 | 3 | 1 |  |  |  |  |
| Vector of scores | Group 3 <br> Monday |  |  |  | Group 4 <br> Wednesday |  |  |  | Group 5 <br> Friday |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | $\mathrm{A}^{3}$ | $\mathrm{B}^{4}$ | $\mathrm{C}^{5}$ | $\mathrm{D}^{6}$ | $\mathrm{A}^{4}$ | $\mathrm{B}^{5}$ | $\mathrm{C}^{6}$ | $\mathrm{D}^{7}$ | $\mathrm{A}^{5}$ | $\mathrm{B}^{6}$ | $\mathrm{C}^{7}$ | $\mathrm{D}^{8}$ |
| $\mathrm{s}_{1}$ | 5 | 5 | 3 | 2 | 5 | 5 | 2 | 3 | 5 | 4 | 5 | 4 |
| $\mathrm{s}_{2}$ | 3 | 2 | 5 | 5 | 4 | 2 | 1 | 1 | 5 | 5 | 2 | 1 |
| $\mathrm{s}_{3}$ | 4 | 5 | 4 | 5 | 5 | 1 | 5 | 2 | 2 | 3 | 2 | 1 |
| $\mathrm{s}_{4}$ | 1 | 3 | 4 | 4 | 2 | 2 | 2 | 1 | 4 | 3 | 2 | 2 |
| $\mathrm{s}_{5}$ | 5 | 4 | 2 | 3 | 3 | 4 | 2 | 2 | 4 | 2 | 3 | 1 |
| $\mathrm{s}_{6}$ | 5 | 5 | 1 | 4 | 3 | 2 | 4 | 2 | 4 | 5 | 2 | 3 |
| $\mathrm{s}_{7}$ | 4 | 5 | 3 | 4 | 5 | 5 | 3 | 2 | 3 | 3 | 4 | 2 |
| $\mathrm{s}_{8}$ | 5 | 4 | 2 | 4 | 5 | 4 | 2 | 3 | 4 | 3 | 2 | 2 |
| $\mathrm{s}_{9}$ | 3 | 4 | 4 | 1 | 2 | 3 | 2 | 3 | 3 | 3 | 2 | 5 |
| $\mathrm{s}_{10}$ | 4 | 4 | 2 | 3 | 4 | 3 | 2 | 1 | 4 | 5 | 2 | 2 |

Table A.11: Vectors of scores assigned to groups of samples of salmon fillets (A, B, C, D) by the trained panellists on each day of the week in the order shown in Table 4.14

## Ranking with ties

| Ranking | Group 1 | Group 2 | Group 3 |
| :---: | :---: | :---: | :---: |
|  | Tuesday | Thursday | Monday |
| $\precsim_{1}$ | $\mathrm{B}^{2} \prec \mathrm{D}^{4} \prec \mathrm{~A}^{1} \sim \mathrm{C}^{3}$ | $\mathrm{D}^{5} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2} \prec \mathrm{C}^{4}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~A}^{3} \prec \mathrm{~B}^{4}$ |
| $\precsim_{2}$ | $\mathrm{A}^{1} \sim \mathrm{C}^{3} \prec \mathrm{D}^{4} \sim \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{~A}^{2} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \sim \mathrm{~A}^{3} \sim \mathrm{~B}^{4}$ |
| $\precsim_{3}$ | $\mathrm{D}^{4} \prec \mathrm{~A}^{1} \sim \mathrm{C}^{3} \prec \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{~B}^{3} \prec \mathrm{C}^{4} \prec \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{~A}^{3} \prec \mathrm{D}^{6} \prec \mathrm{~B}^{4}$ |
| $\precsim_{4}$ | $\mathrm{D}^{4} \prec \mathrm{~B}^{2} \prec \mathrm{C}^{3} \prec \mathrm{~A}^{1}$ | $\mathrm{B}^{3} \prec \mathrm{D}^{5} \prec \mathrm{C}^{4} \sim \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{~A}^{3} \prec \mathrm{D}^{6} \prec \mathrm{~B}^{4}$ |
| $\precsim_{5}$ | $\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~A}^{1} \sim \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{~A}^{2} \prec \mathrm{~B}^{3} \prec \mathrm{C}^{4}$ | $\mathrm{A}^{3} \prec \mathrm{D}^{6} \prec \mathrm{~B}^{4} \prec \mathrm{C}^{5}$ |
| $\precsim_{6}$ | $\mathrm{D}^{4} \prec \mathrm{C}^{3} \prec \mathrm{~B}^{2} \prec \mathrm{~A}^{1}$ | $\mathrm{A}^{2} \prec \mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3}$ | $\mathrm{B}^{4} \prec \mathrm{~A}^{3} \prec \mathrm{D}^{6} \prec \mathrm{C}^{5}$ |
| $\precsim_{7}$ | $\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~B}^{2} \prec \mathrm{~A}^{1}$ | $\mathrm{D}^{5} \prec \mathrm{~A}^{2} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \sim \mathrm{~A}^{3} \prec \mathrm{~B}^{4}$ |
| $\precsim_{8}$ | $\mathrm{A}^{1} \prec \mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2} \prec \mathrm{C}^{4}$ | $\mathrm{C}^{5} \prec \mathrm{~A}^{3} \prec \mathrm{D}^{6} \prec \mathrm{~B}^{4}$ |
| $\precsim_{9}$ | $\mathrm{C}^{3} \sim \mathrm{D}^{4} \prec \mathrm{~B}^{2} \prec \mathrm{~A}^{1}$ | $\mathrm{A}^{2} \prec \mathrm{~B}^{3} \sim \mathrm{C}^{4} \prec \mathrm{D}^{5}$ | $\mathrm{C}^{5} \prec \mathrm{~A}^{3} \prec \mathrm{~B}^{4} \prec \mathrm{D}^{6}$ |
| $\precsim_{10}$ | $\mathrm{C}^{3} \prec \mathrm{~B}^{2} \prec \mathrm{D}^{4} \prec \mathrm{~A}^{1}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{~A}^{3} \prec \mathrm{D}^{6} \prec \mathrm{~B}^{4}$ |
| $\precsim_{11}$ | $\mathrm{D}^{4} \prec \mathrm{~A}^{1} \prec \mathrm{C}^{3} \prec \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ | $\mathrm{B}^{4} \prec \mathrm{D}^{6} \prec \mathrm{~A}^{3} \prec \mathrm{C}^{5}$ |
| $\precsim_{12}$ | $\mathrm{C}^{3} \prec \mathrm{~A}^{1} \sim \mathrm{~B}^{2} \sim \mathrm{D}^{4}$ | $\mathrm{D}^{5} \prec \mathrm{~A}^{2} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3}$ | $\mathrm{D}^{6} \prec \mathrm{C}^{5} \prec \mathrm{~A}^{3} \sim \mathrm{~B}^{4}$ |
| $\precsim_{13}$ | $\mathrm{D}^{4} \prec \mathrm{C}^{3} \sim \mathrm{~A}^{1} \prec \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{~B}^{3} \prec \mathrm{C}^{4} \sim \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{~B}^{4} \prec \mathrm{~A}^{3} \prec \mathrm{D}^{6}$ |
| $\precsim_{14}$ | $\mathrm{C}^{3} \prec \mathrm{~A}^{1} \prec \mathrm{D}^{4} \prec \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~A}^{2} \prec \mathrm{~B}^{3}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \sim \mathrm{~A}^{3} \prec \mathrm{~B}^{4}$ |
| $\precsim_{15}$ | $\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~A}^{1} \prec \mathrm{~B}^{2}$ | $\mathrm{B}^{3} \prec \mathrm{C}^{4} \prec \mathrm{~A}^{2} \prec \mathrm{D}^{5}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~A}^{3} \prec \mathrm{~B}^{4}$ |
| $\precsim_{16}$ | $\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~B}^{2} \prec \mathrm{~A}^{1}$ | $\mathrm{C}^{4} \prec \mathrm{~A}^{2} \prec \mathrm{D}^{5} \prec \mathrm{~B}^{3}$ | $\mathrm{B}^{4} \prec \mathrm{~A}^{3} \prec \mathrm{D}^{6} \prec \mathrm{C}^{5}$ |
| $\precsim_{17}$ | $\mathrm{C}^{3} \prec \mathrm{~A}^{1} \prec \mathrm{D}^{4} \prec \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~A}^{3} \prec \mathrm{~B}^{4}$ |
| $\precsim_{18}$ | $\mathrm{D}^{4} \prec \mathrm{C}^{3} \prec \mathrm{~B}^{2} \prec \mathrm{~A}^{1}$ | $\mathrm{D}^{5} \prec \mathrm{~A}^{2} \prec \mathrm{~B}^{3} \prec \mathrm{C}^{4}$ | $\mathrm{D}^{6} \prec \mathrm{C}^{5} \prec \mathrm{~A}^{3} \sim \mathrm{~B}^{4}$ |
| $\precsim_{19}$ | $\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~A}^{1} \sim \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~B}^{4} \prec \mathrm{~A}^{3}$ |
| $\precsim_{20}$ | $\mathrm{A}^{1} \sim \mathrm{C}^{3} \sim \mathrm{~B}^{2} \prec \mathrm{D}^{4}$ | $\mathrm{D}^{5} \prec \mathrm{~B}^{3} \sim \mathrm{~A}^{2} \prec \mathrm{C}^{4}$ | $\mathrm{C}^{5} \prec \mathrm{~B}^{4} \prec \mathrm{D}^{6} \prec \mathrm{~A}^{3}$ |
| $\precsim_{21}$ | $\mathrm{D}^{4} \prec \mathrm{~A}^{1} \prec \mathrm{C}^{3} \prec \mathrm{~B}^{2}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \sim \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{~A}^{3} \prec \mathrm{D}^{6} \prec \mathrm{~B}^{4}$ |
| $\precsim_{22}$ | $\mathrm{D}^{4} \prec \mathrm{~B}^{2} \prec \mathrm{C}^{3} \prec \mathrm{~A}^{1}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~B}^{3} \prec \mathrm{~A}^{2}$ | $\mathrm{C}^{5} \prec \mathrm{~A}^{3} \sim \mathrm{~B}^{4} \prec \mathrm{D}^{6}$ |
| $\precsim_{23}$ | $\mathrm{C}^{3} \prec \mathrm{~B}^{2} \sim \mathrm{D}^{4} \prec \mathrm{~A}^{1}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~A}^{2} \sim \mathrm{~B}^{3}$ | $\mathrm{D}^{6} \prec \mathrm{C}^{5} \prec \mathrm{~A}^{3} \sim \mathrm{~B}^{4}$ |
| $\precsim_{24}$ | $\mathrm{D}^{4} \prec \mathrm{~A}^{1} \prec \mathrm{~B}^{2} \prec \mathrm{C}^{3}$ | $\mathrm{D}^{5} \prec \mathrm{C}^{4} \prec \mathrm{~A}^{2} \prec \mathrm{~B}^{3}$ | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~B}^{4} \prec \mathrm{~A}^{3}$ |
| $\precsim_{25}$ | $\mathrm{D}^{4} \prec \mathrm{~A}^{1} \sim \mathrm{~B}^{2} \sim \mathrm{C}^{3}$ |  | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~B}^{4} \prec \mathrm{~A}^{3}$ |
| $\precsim_{26}$ | $\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~B}^{2} \prec \mathrm{~A}^{1}$ |  | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~A}^{3} \prec \mathrm{~B}^{4}$ |
| $\precsim_{27}$ | $\mathrm{C}^{3} \prec \mathrm{D}^{4} \prec \mathrm{~A}^{1} \prec \mathrm{~B}^{2}$ |  | $\mathrm{C}^{5} \prec \mathrm{D}^{6} \prec \mathrm{~B}^{4} \prec \mathrm{~A}^{3}$ |
| $\precsim_{28}$ |  |  |  |

Table A.12: Rankings with ties of samples of salmon fillets (A, B, C, D) expressed by the untrained panellists on each day of the week in the order shown in Table 4.14

| Ranking | Group 4 | Group 5 |
| :---: | :---: | :---: |
|  | Wednesday | Friday |
| $\precsim_{1}$ | $\mathrm{B}^{5} \sim \mathrm{D}^{7} \prec \mathrm{C}^{6} \sim \mathrm{~A}^{4}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| $\precsim_{2}$ | $\mathrm{D}^{7} \prec \mathrm{C}^{6} \prec \mathrm{~B}^{5} \prec \mathrm{~A}^{4}$ | $\mathrm{B}^{6} \prec \mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~A}^{5}$ |
| $\precsim_{3}$ | $\mathrm{D}^{7} \prec \mathrm{~B}^{5} \prec \mathrm{C}^{6} \sim \mathrm{~A}^{4}$ | $\mathrm{C}^{7} \prec \mathrm{~B}^{6} \prec \mathrm{~A}^{5} \sim \mathrm{D}^{8}$ |
| $\precsim_{4}$ | $\mathrm{A}^{4} \prec \mathrm{D}^{7} \prec \mathrm{C}^{6} \prec \mathrm{~B}^{5}$ | $\mathrm{C}^{7} \prec \mathrm{~A}^{5} \prec \mathrm{D}^{8} \prec \mathrm{~B}^{6}$ |
| $\precsim_{5}$ | $\mathrm{A}^{4} \sim \mathrm{D}^{7} \prec \mathrm{C}^{6} \prec \mathrm{~B}^{5}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| $\precsim_{6}$ | $\mathrm{D}^{7} \prec \mathrm{C}^{6} \sim \mathrm{~A}^{4} \sim \mathrm{~B}^{5}$ | $\mathrm{C}^{7} \sim \mathrm{D}^{8} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| $\precsim_{7}$ | $\mathrm{D}^{7} \sim \mathrm{C}^{6} \prec \mathrm{~A}^{4} \sim \mathrm{~B}^{5}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~B}^{6} \prec \mathrm{~A}^{5}$ |
| $\precsim_{8}$ | $\mathrm{D}^{7} \prec \mathrm{~A}^{4} \sim \mathrm{C}^{6} \prec \mathrm{~B}^{5}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| $\precsim 9$ | $\mathrm{A}^{4} \prec \mathrm{C}^{6} \sim \mathrm{~B}^{5} \prec \mathrm{D}^{7}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~B}^{6} \sim \mathrm{~A}^{5}$ |
| $\precsim_{10}$ | $\mathrm{B}^{5} \prec \mathrm{D}^{7} \prec \mathrm{~A}^{4} \prec \mathrm{C}^{6}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \sim \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| $\precsim_{11}$ | $\mathrm{D}^{7} \prec \mathrm{~B}^{5} \prec \mathrm{C}^{6} \prec \mathrm{~A}^{4}$ | $\mathrm{D}^{8} \prec \mathrm{C}^{7} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| $\precsim_{12}$ | $\mathrm{D}^{7} \prec \mathrm{~A}^{4} \prec \mathrm{C}^{6} \sim \mathrm{~B}^{5}$ | $\mathrm{C}^{7} \prec \mathrm{~A}^{5} \prec \mathrm{D}^{8} \prec \mathrm{~B}^{6}$ |
| $\precsim_{13}$ | $\mathrm{B}^{5} \sim \mathrm{D}^{7} \sim \mathrm{C}^{6} \prec \mathrm{~A}^{4}$ | $\mathrm{D}^{8} \prec \mathrm{C}^{7} \prec \mathrm{~A}^{5} \sim \mathrm{~B}^{6}$ |
| $\precsim_{14}$ | $\mathrm{B}^{5} \prec \mathrm{D}^{7} \prec \mathrm{~A}^{4} \prec \mathrm{C}^{6}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~B}^{6} \prec \mathrm{~A}^{5}$ |
| $\precsim_{15}$ | $\mathrm{B}^{5} \prec \mathrm{~A}^{4} \prec \mathrm{D}^{7} \prec \mathrm{C}^{6}$ | $\mathrm{C}^{7} \prec \mathrm{~A}^{5} \sim \mathrm{D}^{8} \prec \mathrm{~B}^{6}$ |
| $\precsim_{16}$ | $\mathrm{D}^{7} \prec \mathrm{C}^{6} \prec \mathrm{~A}^{4} \sim \mathrm{~B}^{5}$ | $\mathrm{C}^{7} \prec \mathrm{D}^{8} \prec \mathrm{~B}^{6} \prec \mathrm{~A}^{5}$ |
| $\precsim_{17}$ | $\mathrm{D}^{7} \prec \mathrm{C}^{6} \sim \mathrm{~B}^{5} \prec \mathrm{~A}^{4}$ | $\mathrm{D}^{8} \prec \mathrm{C}^{7} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| $\precsim_{18}$ | $\mathrm{D}^{7} \prec \mathrm{C}^{6} \sim \mathrm{~B}^{5} \prec \mathrm{~A}^{4}$ | $\mathrm{D}^{8} \prec \mathrm{C}^{7} \prec \mathrm{~A}^{5} \sim \mathrm{~B}^{6}$ |
| $\precsim_{19}$ | $\mathrm{B}^{5} \prec \mathrm{D}^{7} \prec \mathrm{C}^{6} \prec \mathrm{~A}^{4}$ | $\mathrm{D}^{8} \prec \mathrm{C}^{7} \prec \mathrm{~B}^{6} \prec \mathrm{~A}^{5}$ |
| $\precsim_{20}$ | $\mathrm{D}^{7} \prec \mathrm{C}^{6} \prec \mathrm{~B}^{5} \prec \mathrm{~A}^{4}$ | $\mathrm{D}^{8} \prec \mathrm{C}^{7} \prec \mathrm{~B}^{6} \sim \mathrm{~A}^{5}$ |
| $\precsim_{21}$ | $\mathrm{A}^{4} \sim \mathrm{D}^{7} \prec \mathrm{C}^{6} \sim \mathrm{~B}^{5}$ | $\mathrm{D}^{8} \prec \mathrm{~A}^{5} \prec \mathrm{C}^{7} \prec \mathrm{~B}^{6}$ |
| $\precsim_{22}$ | $\mathrm{A}^{4} \sim \mathrm{~B}^{5} \prec \mathrm{C}^{6} \prec \mathrm{D}^{7}$ | $\mathrm{D}^{8} \prec \mathrm{C}^{7} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| $\precsim_{23}$ | $\mathrm{D}^{7} \prec \mathrm{C}^{6} \prec \mathrm{~B}^{5} \prec \mathrm{~A}^{4}$ | $\mathrm{D}^{8} \prec \mathrm{C}^{7} \prec \mathrm{~A}^{5} \prec \mathrm{~B}^{6}$ |
| $\precsim_{24}$ | $\mathrm{D}^{7} \prec \mathrm{~A}^{4} \prec \mathrm{~B}^{5} \prec \mathrm{C}^{6}$ |  |
| $\precsim_{25}$ | $\mathrm{D}^{7} \prec \mathrm{~B}^{5} \prec \mathrm{C}^{6} \prec \mathrm{~A}^{4}$ |  |
| $\precsim_{26}$ | $\mathrm{A}^{4} \prec \mathrm{D}^{7} \prec \mathrm{C}^{6} \prec \mathrm{~B}^{5}$ |  |
| $\precsim_{27}$ | $\mathrm{D}^{7} \prec \mathrm{~B}^{5} \prec \mathrm{C}^{6} \prec \mathrm{~A}^{4}$ |  |
| $\precsim_{28}$ | $\mathrm{D}^{7} \prec \mathrm{~B}^{5} \prec \mathrm{C}^{6} \prec \mathrm{~A}^{4}$ |  |

Table A.12: (Continued) Rankings with ties of samples of salmon fillets (A, B, C, D) expressed by the untrained panellists on each day of the week in the order shown in Table 4.14

## B Appendix

## B.1. Quantification of VOCs in chicken breasts

| VOC | Precursor | m/z | b(\%) | k | Product ion |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Acids |  |  |  |  |  |
| Acetic acid | $\mathrm{NO}^{+}$ | 90 | 100 | 9.0 E-10 | $\mathrm{NO}^{+} . \mathrm{CH}_{3} \mathrm{COOH}^{*}$ |
|  | $\mathrm{NO}^{+}$ | 108 |  | $9.0 \mathrm{E}-10$ | $\mathrm{NO}^{+} \cdot \mathrm{CH}_{3} \mathrm{COOH} \cdot \mathrm{H}_{2} \mathrm{O}$ |
| Hexanoic acid | $\mathrm{NO}^{+}$ | 146 | 90 | $2.5 \mathrm{E}-9$ | $\mathrm{C}_{6} \mathrm{H}_{12} \mathrm{O}_{2} \cdot \mathrm{NO}^{+*}$ |
| Alcohols |  |  |  |  |  |
| 1-Octen-3-ol | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 111 | 100 | 3.1 E-9 | $\mathrm{C}_{8} \mathrm{H}_{15}^{+}$* |
| 1-Pentanol | $\mathrm{O}_{2}^{+}$ | 42 | 35 | $2.8 \mathrm{E}-9$ | $\mathrm{C}_{3} \mathrm{H}_{6}^{+}$* |
| 2,3-Butanediol | $\mathrm{NO}^{+}$ | 89 | 100 | 2.3 E-9 | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{O}_{2}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 107 |  | $2.3 \mathrm{E}-9$ | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{O}_{2}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
| 2-Propanol | $\mathrm{NO}^{+}$ | 59 | 100 | $2.4 \mathrm{E}-9$ | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{O}^{+*}$ |
| 3-Methyl-1-butanol | $\mathrm{O}_{2}^{+}$ | 59 | 85 | $2.1 \mathrm{E}-9$ | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{O}^{+*}$ |
| Ethanol | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 47 | 100 | $2.7 \mathrm{E}-9$ | $\mathrm{C}_{2} \mathrm{H}_{7} \mathrm{O}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 65 |  | $2.7 \mathrm{E}-9$ | $\mathrm{C}_{2} \mathrm{H}_{7} \mathrm{O}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 93 |  | 2.7 E-9 | $\left(\mathrm{C}_{2} \mathrm{H}_{6} \mathrm{O}^{+}\right)_{2} \cdot \mathrm{H}^{+}$ |
| Aldehydes |  |  |  |  |  |
| Hexanal | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 101 | 50 | 3.7 E-9 | $\mathrm{C}_{6} \mathrm{H} 13 \mathrm{O}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 119 |  | $3.7 \mathrm{E}-9$ | $\mathrm{C}_{6} \mathrm{H13O}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 137 |  | $3.7 \mathrm{E}-9$ | $\mathrm{C}_{6}{\mathrm{H} 13 \mathrm{O}^{+} \cdot 2 \mathrm{H}_{2} \mathrm{O}}$ |
| Nonanal | $\mathrm{NO}^{+}$ | 141 | 100 | 2.7 E-9 | $\mathrm{C}_{8} \mathrm{H}_{17} \mathrm{O}^{+*}$ |
| Octanal | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 129 | 85 | $3.8 \mathrm{E}-9$ | $\mathrm{C}_{8} \mathrm{H}_{17} \mathrm{O}^{+}$ |
|  | $\mathrm{NO}^{+}$ | 127 | 100 | $3.0 \mathrm{E}-9$ | $\mathrm{C}_{8} \mathrm{H}_{15} \mathrm{O}^{+*}$ |
| Pentanal | $\mathrm{NO}^{+}$ | 85 | 100 | $3.2 \mathrm{E}-9$ | $\mathrm{C}_{5} \mathrm{H}_{9} \mathrm{O}^{+*}$ |
| Ketones |  |  |  |  |  |
| 1-Penten-3-one | $\mathrm{NO}^{+}$ | 114 | 100 | 2.5 E-9 | $\mathrm{C}_{5} \mathrm{H}_{8} \mathrm{O} . \mathrm{NO}^{+*}$ |
| 2,3-Butanedione | $\mathrm{NO}^{+}$ | 86 | 65 | 1.3 E-9 | $\mathrm{C}_{4} \mathrm{H}_{6} \mathrm{O}_{2}^{+*}$ |
| Acetoin | $\mathrm{NO}^{+}$ | 118 | 100 | 2.5 E-9 | $\mathrm{C}_{4} \mathrm{H}_{8} \mathrm{O}_{2} \cdot \mathrm{NO}^{+*}$ |
| Acetone | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 59 | 100 | $3.9 \mathrm{E}-9$ | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{O}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 77 |  | $3.9 \mathrm{E}-9$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{CO} . \mathrm{H}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
| Butanone | $\mathrm{NO}^{+}$ | 102 | 100 | $2.8 \mathrm{E}-9$ | $\mathrm{NO}^{+} . \mathrm{C}_{4} \mathrm{H}_{8} \mathrm{O}^{*}$ |
| Sulfur compounds |  |  |  |  |  |
| Dimethyl disulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 95 | 100 | $2.6 \mathrm{E}-9$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}_{2} \cdot \mathrm{H}^{+*}$ |
| Dimethyl sulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 63 | 100 | $2.5 \mathrm{E}-9$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S} \cdot \mathrm{H}^{+*}$ |
| Dimethyl trisulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 127 | 100 | 2.8 E-9 | $\mathrm{C}_{2} \mathrm{H}_{6} \mathrm{~S}_{3} \mathrm{H}^{+*}$ |
| Hydrogen sulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 35 | 100 | $1.6 \mathrm{E}-9$ | $\mathrm{H}_{3} \mathrm{~S}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 53 |  | 1.6 E-9 | $\mathrm{H}_{3} \mathrm{~S}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
| Esters |  |  |  |  |  |
| Ethyl acetate | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 89 | 100 | 2.9 E-9 | $\mathrm{CH}_{3} \mathrm{COOC}_{2} \mathrm{H}_{5} \cdot \mathrm{H}^{+*}$ |
| Amines |  |  |  |  |  |
| Ammonia | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 18 | 100 | $2.6 \mathrm{E}-9$ | $\mathrm{NH}^{+} 4^{*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 36 |  | $2.6 \mathrm{E}-9$ | $\mathrm{NH}_{4}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |

Table B.1: Volatile organic compounds (VOCs) quantified in packaged chicken samples with SIFT-MS: product ions, mass to charge ratios ( $\mathrm{m} / \mathrm{z}$ ), branching ratios (b) and reaction rate coefficients (k). Product ions denoted with * were selected for quantifying the respective VOC.

|  | $\mathrm{H} 4: 30 / 70$ |  |  |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\left(\mathrm{CO}_{2}, \mathrm{O}_{2}\right) 4{ }^{\circ} \mathrm{C}$ |  |  |  |  |  |  |  |
| VOC | 0 | 5 | 7 | 8 | 9 | 11 | 13 | 15 |
| 1-octen-3-ol | 36.58 | 32.05 | 37.91 | 33.25 | 43.51 | 35.82 | 35.88 | 27.99 |
| 1-pentanol | 363.75 | 327.60 | 340.59 | 321.18 | 368.28 | 321.97 | 415.04 | 328.20 |
| 1-penten-3-one | 667.55 | 675.47 | 652.26 | 629.36 | 662.75 | 655.54 | 665.80 | 467.13 |
| 2,3-butanediol | 56.46 | 59.16 | 75.90 | 46.02 | 101.45 | 78.32 | 74.02 | 80.96 |
| 2,3-butanedione | 988.49 | 859.11 | 898.61 | 681.28 | 959.44 | 728.42 | 1055.32 | 942.19 |
| 2-propanol | 24.85 | 25.91 | 29.68 | 19.38 | 39.15 | 26.68 | 26.50 | 52.10 |
| 3-methyl-1-butanol | 443.22 | 544.18 | 626.23 | 444.83 | 797.01 | 565.21 | 583.24 | 635.10 |
| Acetic acid | 5.46 | 12.37 | 11.48 | 10.37 | 13.29 | 11.03 | 7.31 | 11.63 |
| Acetoin | 39.54 | 94.60 | 74.68 | 60.41 | 100.99 | 68.94 | 69.82 | 89.23 |
| Acetone | 6.50 | 10.96 | 17.43 | 12.35 | 14.18 | 15.07 | 188.67 | 282.88 |
| Ammonia | 730.77 | 789.53 | 941.83 | 638.41 | 1298.34 | 994.21 | 923.62 | 1046.57 |
| Butanone | 23.39 | 29.33 | 35.87 | 26.47 | 44.88 | 33.81 | 39.60 | 40.92 |
| Dimethyl disulfide | 44.36 | 41.26 | 38.21 | 37.39 | 68.09 | 48.31 | 54.41 | 64.13 |
| Dimethyl sulfide | 5.41 | 12.35 | 14.41 | 10.45 | 15.64 | 13.07 | 10.69 | 11.00 |
| Dimethyl trisulfide | 191.59 | 61.44 | 68.26 | 358.25 | 85.44 | 139.33 | 132.13 | 275.88 |
| Ethanol | 417.47 | 419.71 | 340.30 | 377.25 | 383.88 | 384.18 | 393.39 | 255.97 |
| Ethyl acetate | 13.16 | 17.10 | 16.67 | 18.36 | 15.45 | 12.58 | 91.52 | 135.11 |
| Hexanal | 619.29 | 628.42 | 668.39 | 808.22 | 730.18 | 734.92 | 1246.35 | 1326.51 |
| Hexanoic acid | 16.37 | 30.22 | 19.69 | 30.75 | 19.74 | 15.60 | 16.14 | 20.99 |
| Hydrogen sulfide | 2.82 | 6.75 | 5.45 | 5.55 | 5.44 | 4.70 | 4.05 | 4.17 |
| Nonanal | 547.33 | 569.59 | 438.85 | 535.17 | 526.54 | 544.37 | 547.87 | 355.45 |
| Octanal | 1417.06 | 1506.58 | 1308.88 | 1324.56 | 1409.15 | 1434.23 | 1428.80 | 994.07 |
| Pentanal | 119.47 | 184.40 | 170.51 | 147.44 | 180.61 | 177.80 | 328.59 | 469.12 |

Table B.2: Measured concentrations ( $\mathrm{\mu g} / \mathrm{m}^{3}$ ) of VOCs detected with SIFT-MS on each day in the headspace of chicken samples used for labelling tests.

| VOC | L4: $40 / 30 / 30\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ |  |  |  |  |  |  | L8: $40 / 30 / 30\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 8^{\circ} \mathrm{C}$ |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0 | 5 | 7 | 9 | 11 | 13 | 15 | 0 | 2 | 4 | 5 | 6 | 7 |
| 1-octen-3-ol | 33.72 | 27.66 | 28.39 |  | 43.49 | 39.46 | 34.38 | 22.12 | 14.79 | 33.29 | 21.93 | 31.71 | 27.56 |
| 1-pentanol | 103.47 | 36.63 | 104.05 |  | 136.36 | 185.32 | 141.32 | 88.87 | 72.30 | 56.60 | 84.87 | 122.36 | 122.94 |
| 1-penten-3-one | 225.77 | 9.85 | 182.52 |  | 225.93 | 206.63 | 128.69 | 198.68 | 133.05 | 116.43 | 134.95 | 218.03 | 191.96 |
| 2,3-butanediol | 30.59 | 37.01 | 64.58 |  | 297.08 | 195.95 | 98.39 | 54.36 | 51.77 | 41.84 | 59.50 | 96.47 | 106.78 |
| 2,3-butanedione | 80.82 | 10.06 | 118.44 |  | 284.88 | 313.15 | 136.24 | 71.07 | 42.77 | 44.92 | 66.44 | 159.14 | 215.72 |
| 2-propanol | 20.55 | 5.46 | 18.68 |  | 23.05 | 40.47 | 17.39 | 33.48 | 48.79 | 21.48 | 52.35 | 39.70 | 47.93 |
| 3-methyl-1-butanol | 76.62 | 51.44 | 109.23 |  | 231.82 | 388.99 | 152.72 | 238.30 | 243.97 | 120.79 | 234.92 | 425.91 | 420.86 |
| Acetic acid | 22.84 | 26.58 | 57.81 |  | 318.96 | 134.11 | 85.08 | 21.56 | 23.64 | 22.53 | 32.23 | 48.26 | 52.92 |
| Acetoin | 9.05 | 19.13 | 606.74 |  | 2353.94 | 3750.72 | 1887.87 | 6.76 | 19.84 | 12.24 | 69.74 | 358.81 | 759.01 |
| Acetone | 515.31 | 361.49 | 614.99 |  | 854.88 | 1400.85 | 547.72 | 1228.39 | 1189.29 | 740.31 | 1215.27 | 2151.82 | 2163.87 |
| Ammonia | 21.04 | 14.17 | 35.75 |  | 26.99 | 17.79 | 25.43 | 24.85 | 27.23 | 12.62 | 17.09 | 18.21 | 23.88 |
| Butanone | 43.43 | 4.02 | 29.73 |  | 43.22 | 57.00 | 22.50 | 37.19 | 32.21 | 21.55 | 30.67 | 59.79 | 47.01 |
| Dimethyl disulfide | 5.62 | 6.78 | 7.71 |  | 7.42 | 7.85 | 6.68 | 4.83 | 7.62 | 5.88 | 4.63 | 5.74 | 5.73 |
| Dimethyl sulfide | 137.27 | 73.55 | 452.06 |  | 1407.77 | 1119.03 | 1000.10 | 169.61 | 94.51 | 228.44 | 229.88 | 225.47 | 388.99 |
| Dimethyl trisulfide | 187.79 | 28.23 | 165.14 |  | 155.38 | 139.73 | 86.25 | 149.24 | 96.09 | 82.97 | 97.00 | 135.40 | 129.60 |
| Ethanol | 105.83 | 110.91 | 2887.10 |  | 7039.20 | 6480.72 | 4870.86 | 103.39 | 82.64 | 65.78 | 154.45 | 441.17 | 748.15 |
| Ethyl acetate | 10.70 | 24.34 | 822.89 |  | 3566.06 | 5971.67 | 2962.07 | 6.44 | 16.24 | 13.15 | 95.83 | 519.89 | 1123.73 |
| Hexanal | 62.22 | 41.40 | 64.05 |  | 85.31 | 74.48 | 48.92 | 55.10 | 43.86 | 40.80 | 46.14 | 50.95 | 55.39 |
| Hexanoic acid | 3.70 | 5.48 | 7.35 |  | 3.59 | 5.64 | 4.52 | 3.40 | 4.50 | 4.13 | 4.34 | 3.00 | 3.83 |
| Hydrogen sulfide | 1.17 | 1.90 | 1.93 |  | 1.14 | 4.60 | 1.70 | 0.75 | 0.69 | 0.71 | 6.76 | 15.49 | 46.85 |
| Nonanal | 198.49 | 12.99 | 159.12 |  | 217.48 | 194.13 | 109.09 | 203.47 | 117.29 | 99.76 | 125.74 | 188.64 | 176.54 |
| Octanal | 4.67 | 5.07 | 5.70 |  | 3.67 | 5.32 | 3.67 | 3.84 | 3.25 | 3.81 | 3.75 | 3.45 | 3.60 |
| Pentanal | 62.54 | 5.49 | 76.18 |  | 171.35 | 161.23 | 84.58 | 54.37 | 34.74 | 35.46 | 75.79 | 167.77 | 242.67 |

Table B.3: Measured concentrations ( $\mathrm{\mu g} / \mathrm{m}^{3}$ ) of VOCs detected with SIFT-MS on each day in the headspace of chicken samples used for ranking tests.

| VOC | H4: 30/70 ( $\left.\mathrm{CO}_{2}, \mathrm{O}_{2}\right) 4^{\circ} \mathrm{C}$ |  |  |  |  |  |  | H8: $30 / 70\left(\mathrm{CO}_{2}, \mathrm{O}_{2}\right) 8^{\circ} \mathrm{C}$ |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0 | 5 | 7 | 9 | 11 | 13 | 15 | 0 | 2 | 4 | 5 | 6 | 7 |
| 1-octen-3-ol | 32.34 | 24.00 | 21.82 | 21.34 | 38.66 | 37.01 | 25.65 | 8.54 | 20.65 | 34.53 | 27.74 | 34.86 | 26.63 |
| 1-pentanol | 38.97 | 51.90 | 61.79 | 66.88 | 68.21 | 63.78 | 67.64 | 17.88 | 46.06 | 30.85 | 53.73 | 69.62 | 61.94 |
| 1-penten-3-one | 102.27 | 125.81 | 104.57 | 163.76 | 200.61 | 174.02 | 151.35 | 10.90 | 118.88 | 48.66 | 189.07 | 188.43 | 175.69 |
| 2,3-butanediol | 19.33 | 33.06 | 33.36 | 35.52 | 49.44 | 47.13 | 61.90 | 10.80 | 29.65 | 27.12 | 31.59 | 40.05 | 46.49 |
| 2,3-butanedione | 41.34 | 56.29 | 47.37 | 139.25 | 245.47 | 287.91 | 254.04 | 7.62 | 40.81 | 35.75 | 87.85 | 221.71 | 275.22 |
| 2-propanol | 19.38 | 19.77 | 15.82 | 22.98 | 20.68 | 20.75 | 30.75 | 4.04 | 9.79 | 11.03 | 15.95 | 19.61 | 40.02 |
| 3-methyl-1-butanol | 74.26 | 171.22 | 114.13 | 128.68 | 169.46 | 143.90 | 169.52 | 18.62 | 58.06 | 69.10 | 122.44 | 122.90 | 151.54 |
| Acetic acid | 17.85 | 25.78 | 22.59 | 23.68 | 45.43 | 30.63 | 71.73 | 18.80 | 19.47 | 23.06 | 19.34 | 25.49 | 29.15 |
| Acetoin | 12.13 | 23.32 | 82.62 | 252.16 | 592.65 | 798.46 | 1513.59 | 50.77 | 35.87 | 9.96 | 53.57 | 450.47 | 646.16 |
| Acetone | 340.97 | 603.22 | 602.81 | 578.56 | 787.38 | 649.30 | 740.88 | 114.15 | 314.68 | 481.77 | 642.15 | 597.11 | 680.82 |
| Ammonia | 29.95 | 23.96 | 25.91 | 34.16 | 18.18 | 20.07 | 21.73 | 15.21 | 10.22 | 13.76 | 11.44 | 18.98 | 24.80 |
| Butanone | 11.30 | 40.77 | 26.78 | 30.17 | 39.06 | 48.16 | 30.27 | 2.70 | 14.90 | 16.89 | 23.76 | 22.23 | 34.31 |
| Dimethyl disulfide | 5.72 | 7.01 | 4.68 | 5.13 | 7.63 | 6.40 | 6.17 | 2.67 | 3.92 | 5.85 | 5.12 | 6.02 | 4.09 |
| Dimethyl sulfide | 89.04 | 88.14 | 72.18 | 383.45 | 151.84 | 324.53 | 461.87 | 71.21 | 69.10 | 155.44 | 752.50 | 151.76 | 651.07 |
| Dimethyl trisulfide | 85.49 | 96.43 | 90.99 | 110.90 | 134.25 | 115.79 | 106.42 | 22.71 | 95.40 | 45.68 | 131.26 | 134.79 | 118.12 |
| Ethanol | 107.53 | 306.30 | 125.96 | 87.67 | 134.06 | 197.09 | 391.28 | 105.12 | 157.59 | 62.48 | 140.06 | 314.27 | 516.89 |
| Ethyl acetate | 9.96 | 19.22 | 80.82 | 363.38 | 858.27 | 1204.38 | 2213.07 | 41.35 | 26.37 | 12.08 | 77.30 | 671.70 | 960.43 |
| Hexanal | 39.00 | 42.88 | 37.26 | 39.24 | 45.45 | 48.89 | 45.45 | 23.66 | 40.48 | 27.25 | 46.19 | 54.40 | 45.53 |
| Hexanoic acid | 3.44 | 3.62 | 2.95 | 2.68 | 3.37 | 3.16 | 3.60 | 3.54 | 4.00 | 3.74 | 3.63 | 4.29 | 3.70 |
| Hydrogen sulfide | 0.87 | 0.86 | 0.89 | 1.08 | 1.28 | 1.04 | 1.22 | 0.37 | 0.65 | 0.52 | 0.50 | 2.90 | 4.92 |
| Nonanal | 82.24 | 96.37 | 104.01 | 137.65 | 181.01 | 161.63 | 152.23 | 15.23 | 108.51 | 52.06 | 173.90 | 167.47 | 159.70 |
| Octanal | 3.78 | 3.47 | 3.19 | 2.54 | 4.11 | 4.35 | 4.71 | 2.96 | 3.52 | 3.77 | 3.14 | 4.99 | 3.68 |
| Pentanal | 30.69 | 38.66 | 45.56 | 120.89 | 138.50 | 223.35 | 128.68 | 7.05 | 44.03 | 15.81 | 93.03 | 203.42 | 238.33 |

Table B.3: (Continued) Measured concentrations ( $\mathrm{\mu g} / \mathrm{m}^{3}$ ) of VOCs detected with SIFT-MS on each day in the headspace of chicken samples used for ranking tests.

## B.2. Quantification of VOCs in Atlantic cod

| VOC | Precursor | $\mathrm{m} / \mathrm{z}$ | b(\%) | k | Product ion |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Acids |  |  |  |  |  |
| Acetic acid | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 61 | 100 | $2.6 \mathrm{E}-09$ | $\mathrm{CH}_{3} \mathrm{COOH}_{2}+{ }^{*}$ |
|  | $\mathrm{NO}^{+}$ | 90 | 100 | $9.0 \mathrm{E}-10$ | $\mathrm{NO}^{+} . \mathrm{CH}_{3} \mathrm{COOH}$ |
|  | $\mathrm{O}_{2}^{+}$ | 60 | 50 | $2.3 \mathrm{E}-09$ | $\mathrm{CH}_{3} \mathrm{COOH}^{+}$ |
| Alcohols |  |  |  |  |  |
| Ethanol | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 47 | 100 | 2.7 E -09 | $\mathrm{C}_{2} \mathrm{H}_{7} \mathrm{O}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 65 |  | 2.7 E -09 | $\mathrm{C}_{2} \mathrm{H}_{7} \mathrm{O}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 83 |  | $2.7 \mathrm{E}-09$ | $\mathrm{C}_{2} \mathrm{H}_{7} \mathrm{O}^{+} \cdot\left(\mathrm{H}_{2} \mathrm{O}\right)_{2}$ |
| 2,3-butanediol | $\mathrm{NO}^{+}$ | 89 | 100 | $2.3 \mathrm{E}-09$ | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{O}_{2}^{+*}$ |
| 3-methyl-1-butanol | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 71 | 100 | $2.8 \mathrm{E}-09$ | $\mathrm{C}_{5} \mathrm{H}_{11}+^{*}$ |
|  | $\mathrm{NO}^{+}$ | 87 | 85 | $2.3 \mathrm{E}-09$ | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{O}^{+}$ |
| Isobutyl alcohol | $\mathrm{NO}^{+}$ | 73 | 95 | $2.4 \mathrm{E}-09$ | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{O}^{+*}$ |
|  | $\mathrm{O}_{2}^{+}$ | 33 | 50 | $2.5 \mathrm{E}-09$ | $\mathrm{CH}_{5} \mathrm{O}^{+}$ |
| Aldehydes |  |  |  |  |  |
| 2-methylpropanal | $\mathrm{O}_{2}^{+}$ | 72 | 70 | $3.0 \mathrm{E}-09$ | $\mathrm{C}_{4} \mathrm{H}_{8} \mathrm{O}^{+*}$ |
| 3-methylbutanal | $\mathrm{NO}^{+}$ | 85 | 100 | $2.4 \mathrm{E}-09$ | $\mathrm{C}_{5} \mathrm{H}_{9} \mathrm{O}^{+*}$ |
| Ketones |  |  |  |  |  |
| Acetone | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 59 | 100 | $3.9 \mathrm{E}-09$ | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{O}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 88 | 100 | $1.2 \mathrm{E}-09$ | $\mathrm{NO}^{+} . \mathrm{C}_{3} \mathrm{H}_{6} \mathrm{O}$ |
| Acetoin | $\mathrm{O}_{2}^{+}$ | 88 | 20 | $2.5 \mathrm{E}-09$ | $\mathrm{C}_{4} \mathrm{H}_{8} \mathrm{O}_{2}^{+*}$ |
| 2-pentanone | $\mathrm{NO}^{+}$ | 116 | 100 | $3.1 \mathrm{E}-09$ | $\mathrm{NO}^{+} \cdot \mathrm{C}_{5} \mathrm{H}_{10} \mathrm{O}^{+*}$ |
| Sulfur compounds |  |  |  |  |  |
| Hydrogen sulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 35 | 100 | $1.6 \mathrm{E}-09$ | $\mathrm{H}_{3} \mathrm{~S}+*$ |
|  | $\mathrm{O}_{2}^{+}$ | 34 | 100 | $1.4 \mathrm{E}-09$ | $\mathrm{H}_{2} \mathrm{~S}+$ |
| Methyl mercaptan | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 49 | 100 | 1.8 E -09 | $\mathrm{CH}_{4} \mathrm{~S} \cdot \mathrm{H}^{+*}$ |
| Dimethyl sulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 63 | 100 | $2.5 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S} \cdot \mathrm{H}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 62 | 100 | $2.2 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}+$ |
| Dimethyl disulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 95 | 100 | $2.6 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}_{2} \cdot \mathrm{H}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 94 | 100 | $2.4 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}_{2}+$ |
|  | $\mathrm{O}_{2}^{+}$ | 94 | 80 | $2.3 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}_{2}+$ |
| Dimethyl trisulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 127 | 100 | 2.8 E -09 | $\mathrm{C}_{2} \mathrm{H}_{6} \mathrm{~S}_{3} \mathrm{H}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 126 |  | $1.9 \mathrm{E}-09$ | $\mathrm{C}_{2} \mathrm{H}_{6} \mathrm{~S}_{3}+$ |
| Esters |  |  |  |  |  |
| Ethyl acetate | $\mathrm{NO}^{+}$ | 118 | 90 | 2.1 E -09 | $\mathrm{NO}^{+} . \mathrm{CH}_{3} \mathrm{COOC}_{2} \mathrm{H}_{5}{ }^{*}$ |
| Ethyl propanoate | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 103 | 95 | $2.9 \mathrm{E}-09$ | $\mathrm{C}_{2} \mathrm{H}_{5} \mathrm{COOC}_{2} \mathrm{H}_{5} \cdot \mathrm{H}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 132 | 60 | $2.5 \mathrm{E}-09$ | $\mathrm{NO}^{+} . \mathrm{C}_{2} \mathrm{H}_{5} \mathrm{COOC}_{2} \mathrm{H}_{5}$ |
| Amines |  |  |  |  |  |
| Ammonia | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 18 | 100 | 2.6 E -09 | $\mathrm{NH}_{4}+{ }^{*}$ |
|  | $\mathrm{O}_{2}^{+}$ | 17 | 100 | $2.4 \mathrm{E}-09$ | $\mathrm{NH}_{3}+$ |
| Dimethylamine | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 46 | 100 | 2.1 E -09 | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~N} \cdot \mathrm{H}^{+*}$ |
| Trimethylamine | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 60 | 90 | $2.0 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{3} \mathrm{~N} \cdot \mathrm{H}^{+}$ |
|  | $\mathrm{NO}^{+}$ | 59 | 100 | $1.6 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{3} \mathrm{~N}^{+*}$ |

Table B.4: Volatile organic compounds (VOCs) quantified in packaged cod samples with SIFT-MS: product ions, mass to charge ratios ( $\mathrm{m} / \mathrm{z}$ ), branching ratios (b) and reaction rate coefficients (k). Product ions denoted with * were selected for quantifying the respective VOC.

| VOC | L4: $60 / 5 / 35\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ Company A |  |  |  | L8: $60 / 5 / 35\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 8^{\circ} \mathrm{C}$Company A |  |  |  | $\begin{gathered} \mathrm{H} 4: 40 / 60\left(\mathrm{CO}_{2} / \mathrm{O}_{2}\right) 4^{\circ} \mathrm{C} \\ \text { Company A } \end{gathered}$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0 | 4 | 8 | 13 | 0 | 3 | 5 | 7 | 0 | 4 | 8 | 13 |
| 2,3-butanediol | 2.96 | 3.78 | 39.22 | 56.28 | 3.06 | 41.70 | 152.09 | 89.54 | 2.64 | 3.28 | 59.77 | 124.30 |
| 2-methylpropanal | 28.35 | 48.11 | 68.17 | 46.91 | 64.30 | 86.91 | 83.66 | 40.35 | 78.15 | 35.63 | 71.97 | 66.54 |
| 2-pentanone | 1.02 | 1.56 | 2.46 | 2.77 | 2.24 | 3.45 | 4.78 | 6.88 | 1.54 | 1.59 | 7.18 | 15.20 |
| 3-methyl-1-butanol | 1.18 | 6.57 | 36.89 | 20.51 | 3.02 | 66.32 | 106.17 | 76.13 | 3.69 | 2.88 | 64.68 | 136.04 |
| 3-methylbutanal | 17.15 | 24.91 | 23.52 | 20.38 | 591.28 | 756.88 | 767.69 | 143.75 | 717.43 | 16.72 | 706.35 | 95.76 |
| Acetic acid | 4.24 | 7.20 | 19.40 | 55.81 | 7.31 | 16.44 | 46.55 | 24.28 | 7.48 | 5.92 | 13.99 | 28.68 |
| Acetoin | 3.92 | 6.12 | 25.14 | 22.55 | 6.46 | 46.96 | 73.04 | 56.27 | 8.22 | 3.94 | 33.48 | 59.90 |
| Acetone | 34.63 | 16.50 | 48.64 | 39.01 | 26.95 | 28.53 | 47.24 | 88.81 | 21.04 | 30.79 | 49.65 | 73.07 |
| Ammonia | 17.47 | 11.67 | 14.77 | 17.90 | 14.93 | 13.44 | 16.55 | 22.05 | 13.35 | 13.70 | 17.37 | 13.95 |
| Dimethyl amine | 1.73 | 2.45 | 9.38 | 9.18 | 2.74 | 4.56 | 33.94 | 12.53 | 1.99 | 2.90 | 5.16 | 9.89 |
| Dimethyl disulfide | 0.71 | 1.44 | 3.42 | 15.42 | 1.72 | 1.61 | 4.61 | 9.74 | 1.11 | 0.79 | 2.41 | 5.89 |
| Dimethyl sulfide | 43.39 | 268.63 | 302.83 | 1296.44 | 221.56 | 518.77 | 1451.75 | 1964.68 | 86.14 | 142.32 | 113.84 | 137.59 |
| Dimethyl trisulfide | 10.09 | 10.14 | 9.01 | 9.12 | 19.28 | 19.83 | 17.83 | 8.40 | 18.58 | 1.81 | 17.01 | 7.74 |
| Ethanol | 1497.02 | 418.94 | 6345.53 | 19059.95 | 1721.33 | 1251.87 | 7700.55 | 13288.80 | 2617.80 | 483.19 | 3699.90 | 12562.58 |
| Ethyl acetate | 3.68 | 81.51 | 597.40 | 177.83 | 2.58 | 1019.05 | 1325.81 | 886.89 | 6.52 | 34.98 | 1026.62 | 1976.49 |
| Ethyl propanoate | 1.08 | 1.46 | 7.94 | 3.36 | 1.67 | 2.50 | 5.20 | 3.02 | 1.65 | 1.31 | 4.09 | 8.54 |
| Hydrogen sulfide | 0.22 | 3.56 | 112.40 | 730.88 | 0.38 | 4.50 | 0.24 | 0.72 | 0.43 | 1.16 | 9.95 | 6.67 |
| Isobutyl alcohol | 20.27 | 9.82 | 29.55 | 56.90 | 13.38 | 23.02 | 35.98 | 45.61 | 14.62 | 10.60 | 19.55 | 35.68 |
| Methyl mercaptan | 1.02 | 5.00 | 48.72 | 168.99 | 2.62 | 39.05 | 22.68 | 1144.07 | 1.83 | 1.74 | 2.80 | 8.11 |
| Trimethyl amine | 7.30 | 23.33 | 255.36 | 889.60 | 5.16 | 168.14 | 645.85 | 771.71 | 7.71 | 16.70 | 58.86 | 198.53 |

Table B.5: Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of cod samples used for labelling tests.

|  | $\begin{gathered} \text { H8: } 40 / 60\left(\mathrm{CO}_{2} / \mathrm{O}_{2}\right) 4^{\circ} \mathrm{C} \\ \text { Company B } \end{gathered}$ |  |  |  | $\begin{gathered} \mathrm{H} 8: 40 / 60\left(\mathrm{CO}_{2} / \mathrm{O}_{2}\right) 8^{\circ} \mathrm{C} \\ \text { Company A } \end{gathered}$ |  |  |  | $\begin{gathered} \mathrm{H} 8: 40 / 60\left(\mathrm{CO}_{2} / \mathrm{O}_{2}\right) 8^{\circ} \mathrm{C} \\ \text { Company B } \end{gathered}$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| VOC | 0 | 4 | 8 | 13 | 0 | 3 | 5 | 7 | 0 | 3 | 5 | 7 |
| 2,3-butanediol | 3.24 | 16.36 | 115.24 | 126.77 | 5.35 | 42.16 | 115.42 | 81.10 | 11.26 | 4.62 | 37.83 | 79.46 |
| 2-methylpropanal | 76.67 | 159.99 | 177.28 | 490.10 | 79.34 | 81.65 | 104.31 | 109.19 | 190.47 | 169.13 | 483.42 | 96.48 |
| 2-pentanone | 3.64 | 2.35 | 5.09 | 14.34 | 2.90 | 7.87 | 34.14 | 17.52 | 2.32 | 1.95 | 12.36 | 12.86 |
| 3-methyl-1-butanol | 4.26 | 32.16 | 134.98 | 113.53 | 6.37 | 84.89 | 183.30 | 117.95 | 6.09 | 6.37 | 51.97 | 127.78 |
| 3-methylbutanal | 651.71 | 475.48 | 762.95 | 849.55 | 722.48 | 754.64 | 945.39 | 987.92 | 456.83 | 412.98 | 792.73 | 803.21 |
| Acetic acid | 6.95 | 11.85 | 61.02 | 41.66 | 16.58 | 16.46 | 28.64 | 21.67 | 24.39 | 7.03 | 21.42 | 32.09 |
| Acetoin | 8.58 | 15.55 | 68.78 | 77.07 | 13.66 | 40.51 | 61.19 | 56.20 | 15.71 | 10.44 | 32.46 | 57.76 |
| Acetone | 21.05 | 36.80 | 42.17 | 51.84 | 32.23 | 35.30 | 48.01 | 46.92 | 29.22 | 20.72 | 41.64 | 45.75 |
| Ammonia | 18.77 | 46.79 | 20.48 | 50.52 | 42.89 | 17.04 | 14.87 | 17.37 | 21.51 | 11.93 | 14.03 | 17.46 |
| Dimethyl amine | 3.41 | 6.31 | 10.94 | 18.11 | 3.74 | 3.98 | 13.03 | 11.59 | 4.23 | 3.45 | 5.99 | 9.52 |
| Dimethyl disulfide | 1.23 | 1.00 | 2.90 | 4.21 | 1.18 | 1.01 | 2.68 | 3.02 | 2.39 | 2.46 | 1.79 | 2.87 |
| Dimethyl sulfide | 153.96 | 143.93 | 538.33 | 1224.56 | 124.58 | 182.44 | 181.08 | 481.43 | 128.56 | 126.22 | 102.01 | 210.49 |
| Dimethyl trisulfide | 16.20 | 12.23 | 15.58 | 19.04 | 23.04 | 17.47 | 20.38 | 21.62 | 17.55 | 13.21 | 17.00 | 16.95 |
| Ethanol | 585.20 | 389.70 | 3289.93 | 8101.08 | 478.89 | 1003.94 | 3540.47 | 4593.47 | 6398.25 | 8548.97 | 4408.65 | 5392.42 |
| Ethyl acetate | 8.10 | 433.39 | 2353.37 | 1768.78 | 8.84 | 1310.49 | 2152.68 | 1526.70 | 8.49 | 59.69 | 647.15 | 1667.79 |
| Ethyl propanoate | 1.76 | 2.14 | 4.90 | 4.22 | 2.51 | 5.84 | 12.02 | 6.13 | 2.53 | 1.67 | 3.85 | 5.92 |
| Hydrogen sulfide | 0.74 | 1.49 | 0.61 | 1.46 | 0.79 | 0.76 | 0.73 | 0.84 | 0.56 | 0.41 | 2.38 | 3.43 |
| Isobutyl alcohol | 19.24 | 11.72 | 27.12 | 29.91 | 13.39 | 21.13 | 34.44 | 35.55 | 15.50 | 16.77 | 15.91 | 23.90 |
| Methyl mercaptan | 0.97 | 2.12 | 3.43 | 5.51 | 2.02 | 2.86 | 4.25 | 9.68 | 3.40 | 4.45 | 5.68 | 4.40 |
| Trimethyl amine | 12.15 | 47.37 | 852.14 | 712.26 | 17.97 | 61.83 | 382.40 | 484.85 | 25.94 | 5.10 | 37.90 | 307.67 |

Table B.5: (Continued) Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of cod samples used for labelling tests.

|  | A4: Air $4{ }^{\circ} \mathrm{C}$ (Company A) |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Company A |  |  |  |  |

Table B.5: (Continued) Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of cod samples used for labelling tests.

|  | L4: $60 / 5 / 35\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ <br> Session 1 |  |  |  | L4: $60 / 5 / 35\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ Session 2 |  |  |  | L8: $60 / 5 / 35\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 8{ }^{\circ} \mathrm{C}$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| VOC | 0 | 4 | 8 | 13 | 5 | 6 | 7 | 8 | 0 | 3 | 5 | 7 |
| 2,3-butanediol | 2.09 | 4.32 | 41.20 | 66.40 | 8.01 | 19.95 | 23.27 | 39.22 | 1.80 | 33.86 | 74.44 | 103.24 |
| 2-methylpropanal | 30.14 | 28.60 | 55.65 | 40.70 | 44.94 | 28.07 | 45.45 | 68.17 | 5.25 | 57.43 | 47.84 | 89.27 |
| 2-pentanone | 1.44 | 1.18 | 39.72 | 2.88 | 2.29 | 1.79 | 2.51 | 2.46 | 1.29 | 2.84 | 4.79 | 7.16 |
| 3-methyl-1-butanol | 0.77 | 6.61 | 26.72 | 17.63 | 19.82 | 42.16 | 31.90 | 36.89 | 1.05 | 65.21 | 103.48 | 85.66 |
| 3-methylbutanal | 17.60 | 15.35 | 22.21 | 19.21 | 20.33 | 13.79 | 21.76 | 23.52 | 5.48 | 407.42 | 159.98 | 847.98 |
| Acetic acid | 4.70 | 6.89 | 40.00 | 85.69 | 6.33 | 9.57 | 13.94 | 19.40 | 6.86 | 14.54 | 23.40 | 29.35 |
| Acetoin | 3.46 | 6.17 | 19.96 | 15.53 | 15.44 | 30.20 | 19.59 | 25.14 | 3.35 | 44.77 | 61.64 | 58.10 |
| Acetone | 17.40 | 21.84 | 42.13 | 45.26 | 14.83 | 18.09 | 16.75 | 48.64 | 28.97 | 35.49 | 58.88 | 92.25 |
| Ammonia | 17.26 | 12.37 | 19.15 | 25.14 | 12.06 | 12.06 | 13.30 | 14.77 | 14.97 | 13.18 | 14.15 | 21.00 |
| Dimethyl amine | 1.99 | 2.87 | 9.22 | 10.12 | 3.26 | 3.38 | 6.53 | 9.38 | 3.32 | 4.14 | 12.74 | 12.21 |
| Dimethyl disulfide | 1.02 | 1.53 | 2.81 | 9.46 | 1.49 | 1.65 | 2.15 | 3.42 | 1.27 | 2.15 | 3.82 | 7.38 |
| Dimethyl sulfide | 136.04 | 332.66 | 203.46 | 1254.45 | 194.60 | 180.30 | 213.56 | 302.83 | 219.94 | 471.35 | 728.81 | 1124.29 |
| Dimethyl trisulfide | 11.10 | 9.09 | 8.46 | 10.50 | 7.44 | 6.29 | 8.18 | 9.01 | 1.53 | 11.66 | 7.07 | 19.04 |
| Ethanol | 797.56 | 865.49 | 6446.47 | 13808.06 | 1288.12 | 1812.01 | 3771.41 | 6345.53 | 2071.94 | 1773.51 | 6156.71 | 10822.75 |
| Ethyl acetate | 3.20 | 100.19 | 404.03 | 211.30 | 332.42 | 762.86 | 549.20 | 597.40 | 3.70 | 1023.81 | 1496.90 | 1056.48 |
| Ethyl propanoate | 1.39 | 1.68 | 9.31 | 3.81 | 2.20 | 4.55 | 7.30 | 7.94 | 1.15 | 3.81 | 6.46 | 8.76 |
| Hydrogen sulfide | 0.42 | 2.84 | 7.70 | 685.17 | 2.44 | 0.51 | 0.51 | 112.40 | 0.21 | 0.49 | 0.53 | 1.24 |
| Isobutyl alcohol | 15.03 | 14.36 | 34.49 | 49.88 | 16.35 | 21.75 | 22.37 | 29.55 | 15.94 | 23.47 | 38.10 | 48.91 |
| Methyl mercaptan | 1.61 | 2.87 | 15.33 | 246.57 | 6.63 | 6.18 | 6.06 | 48.72 | 3.95 | 41.34 | 131.40 | 909.87 |
| Trimethyl amine | 5.97 | 22.65 | 513.89 | 1081.90 | 34.80 | 61.40 | 150.19 | 255.36 | 6.52 | 116.81 | 325.13 | 813.01 |

Table B.6: Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of cod samples used for ranking tests.

| VOC | A4: Air $4^{\circ} \mathrm{C}$ |  |  |  | $\begin{gathered} \mathrm{H} 4: 40 / 60\left(\mathrm{CO}_{2} / \mathrm{O}_{2}\right) 4^{\circ} \mathrm{C} \\ \text { Session } 1 \end{gathered}$ |  |  |  | $\begin{gathered} \mathrm{H} 4: 40 / 60\left(\mathrm{CO}_{2} / \mathrm{O}_{2}\right) 4^{\circ} \mathrm{C} \\ \text { Session 2 } \end{gathered}$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0 | 1 | 2 | 3 | 4 | 6 | 8 | 13 | 6 | 7 | 8 | 11 |
| 2,3-butanediol | 4.87 | 4.15 | 3.56 | 32.86 | 4.36 | 8.65 | 47.02 | 71.73 | 8.04 | 23.83 | 36.50 | 55.47 |
| 2-methylpropanal | 65.83 | 102.03 | 101.30 | 101.23 | 79.25 | 51.84 | 58.21 | 82.27 | 89.17 | 54.84 | 52.65 | 80.19 |
| 2-pentanone | 3.03 | 2.10 | 1.54 | 2.70 | 1.71 | 3.18 | 3.57 | 7.72 | 2.81 | 4.11 | 4.58 | 9.46 |
| 3-methyl-1-butanol | 5.75 | 6.00 | 5.08 | 54.00 | 5.61 | 15.79 | 61.93 | 89.73 | 18.53 | 58.23 | 43.12 | 43.63 |
| 3 -methylbutanal | 851.98 | 1212.81 | 1042.99 | 1187.86 | 481.24 | 341.35 | 553.13 | 826.05 | 254.97 | 439.07 | 147.60 | 869.79 |
| Acetic acid | 6.73 | 5.96 | 3.10 | 21.01 | 6.40 | 6.99 | 13.59 | 19.83 | 8.06 | 10.63 | 12.49 | 17.52 |
| Acetoin | 10.68 | 10.29 | 7.17 | 35.20 | 8.18 | 9.18 | 29.18 | 43.69 | 9.99 | 23.65 | 21.48 | 25.04 |
| Acetone | 21.54 | 19.06 | 19.76 | 29.68 | 26.68 | 40.46 | 37.62 | 52.54 | 36.26 | 32.79 | 44.81 | 53.55 |
| Ammonia | 50.36 | 52.88 | 27.88 | 26.43 | 10.66 | 11.87 | 16.93 | 13.88 | 11.17 | 13.66 | 16.80 | 11.65 |
| Dimethyl amine | 4.97 | 4.78 | 4.27 | 8.47 | 2.50 | 2.66 | 4.37 | 8.35 | 2.42 | 4.15 | 3.42 | 5.59 |
| Dimethyl disulfide | 1.19 | 1.11 | 0.88 | 1.34 | 0.88 | 1.19 | 0.90 | 2.96 | 1.10 | 1.23 | 1.45 | 2.69 |
| Dimethyl sulfide | 360.23 | 479.79 | 639.22 | 726.49 | 123.30 | 127.18 | 148.85 | 226.22 | 123.98 | 141.04 | 171.49 | 122.87 |
| Dimethyl trisulfide | 29.35 | 31.07 | 26.19 | 31.26 | 10.99 | 9.28 | 10.28 | 16.07 | 6.93 | 10.42 | 7.23 | 15.91 |
| Ethanol | 1112.07 | 339.15 | 254.01 | 991.27 | 496.37 | 1311.24 | 1901.68 | 5024.17 | 862.37 | 1244.10 | 1551.29 | 7022.35 |
| Ethyl acetate | 8.29 | 5.62 | 3.32 | 807.16 | 44.75 | 241.17 | 908.91 | 1255.38 | 303.83 | 878.49 | 665.67 | 590.59 |
| Ethyl propanoate | 2.24 | 2.02 | 1.79 | 4.60 | 2.46 | 1.40 | 4.07 | 8.69 | 2.78 | 4.48 | 3.90 | 3.60 |
| Hydrogen sulfide | 0.53 | 0.59 | 0.23 | 0.49 | 0.36 | 6.98 | 2.22 | 1.10 | 1.97 | 0.68 | 2.44 | 58.38 |
| Isobutyl alcohol | 10.25 | 8.10 | 4.34 | 15.44 | 11.22 | 16.91 | 14.47 | 23.33 | 17.33 | 15.02 | 15.49 | 22.76 |
| Methyl mercaptan | 2.06 | 1.40 | 2.03 | 8.39 | 1.18 | 2.04 | 3.27 | 10.70 | 1.49 | 1.81 | 1.96 | 5.44 |
| Trimethyl amine | 10.67 | 20.24 | 6.36 | 438.65 | 9.81 | 18.89 | 60.76 | 151.07 | 8.84 | 27.05 | 30.02 | 59.38 |

Table B.6: (Continued) Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of cod samples used for ranking tests.

| VOC | H8: $40 / 60\left(\mathrm{CO}_{2} / \mathrm{O}_{2}\right) 8^{\circ} \mathrm{C}$Session 1 |  |  |  | H8: $40 / 60\left(\mathrm{CO}_{2} / \mathrm{O}_{2}\right) 8^{\circ} \mathrm{C}$Session 2 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0 | 3 | 5 | 6 | 3 | 4 | 5 | 7 |
| 2,3-butanediol | 4.95 | 43.94 | 64.83 | 72.43 | 46.25 | 24.62 | 118.87 | 99.56 |
| 2-methylpropanal | 82.88 | 94.84 | 82.53 | 81.25 | 102.99 | 89.69 | 111.93 | 94.38 |
| 2-pentanone | 2.61 | 5.71 | 8.28 | 9.81 | 6.06 | 4.16 | 19.27 | 9.14 |
| 3-methyl-1-butanol | 6.63 | 81.57 | 120.75 | 133.81 | 73.95 | 38.94 | 136.99 | 134.53 |
| 3-methylbutanal | 860.66 | 995.86 | 672.77 | 893.21 | 996.72 | 954.77 | 932.41 | 940.87 |
| Acetic acid | 10.97 | 10.95 | 24.45 | 14.04 | 25.29 | 9.37 | 34.98 | 23.17 |
| Acetoin | 10.57 | 37.11 | 42.24 | 49.80 | 35.86 | 22.48 | 56.93 | 51.82 |
| Acetone | 27.56 | 39.44 | 42.77 | 48.25 | 55.00 | 26.94 | 49.26 | 46.31 |
| Ammonia | 34.65 | 18.89 | 14.51 | 13.26 | 22.57 | 13.08 | 17.49 | 20.07 |
| Dimethyl amine | 4.34 | 4.36 | 9.86 | 8.47 | 4.19 | 4.54 | 18.90 | 12.78 |
| Dimethyl disulfide | 1.31 | 1.82 | 2.38 | 2.04 | 1.48 | 1.84 | 3.14 | 4.48 |
| Dimethyl sulfide | 164.29 | 313.99 | 216.53 | 344.99 | 178.89 | 288.01 | 319.79 | 610.33 |
| Dimethyl trisulfide | 23.79 | 21.79 | 16.61 | 17.40 | 25.56 | 19.82 | 20.33 | 19.44 |
| Ethanol | 263.45 | 480.42 | 2155.59 | 2278.02 | 917.93 | 545.31 | 5943.79 | 7424.35 |
| Ethyl acetate | 7.12 | 1229.60 | 1464.39 | 1703.78 | 1116.19 | 554.00 | 1710.15 | 1549.01 |
| Ethyl propanoate | 2.42 | 4.31 | 8.72 | 3.64 | 10.02 | 2.37 | 9.24 | 5.29 |
| Hydrogen sulfide | 0.86 | 0.41 | 0.49 | 0.57 | 1.37 | 0.65 | 0.80 | 0.70 |
| Isobutyl alcohol | 9.03 | 21.15 | 34.30 | 27.56 | 16.97 | 25.50 | 33.61 | 39.59 |
| Methyl mercaptan | 2.73 | 2.18 | 4.00 | 3.08 | 3.78 | 2.06 | 4.17 | 18.48 |
| Trimethyl amine | 10.39 | 58.29 | 289.84 | 147.81 | 214.66 | 69.76 | 717.09 | 604.65 |

Table B.6: (Continued) Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of cod samples used for ranking tests.

## B.3. Quantification of VOCs in Atlantic brown shrimp

| VOC | Precursor | $\mathrm{m} / \mathrm{z}$ | $\mathrm{b}(\%)$ | k | Product ion |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Acids |  |  |  |  |  |
| Acetic acid | $\mathrm{NO}^{+}$ | 90 | 100 | $9.0 \mathrm{E}-10$ | $\mathrm{NO}^{+} \cdot \mathrm{CH}_{3} \mathrm{COOH}^{*}$ |
|  | $\mathrm{NO}^{+}$ | 108 |  | $9.0 \mathrm{E}-10$ | $\mathrm{NO}^{+} \cdot \mathrm{CH}_{3} \mathrm{COOH} \cdot \mathrm{H}_{2} \mathrm{O}$ |
| Alcohols |  |  |  |  |  |
| 2,3-butanediol | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 91 | 100 | $3.0 \mathrm{E}-09$ | $\mathrm{C}_{4} \mathrm{H}_{10} \mathrm{O}_{2}^{+} \cdot \mathrm{H}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 89 | $100$ | $2.3 \mathrm{E}-09$ | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{O}_{2}^{+}$ |
| 2-propanol | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 43 | 80 | 2.7 E-09 | $\mathrm{C}_{3} \mathrm{H}_{7}^{+*}$ |
| 3-methyl-1-butanol | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 71 | 100 | 2.8 E-09 | $\mathrm{C}_{5} \mathrm{H}_{11}^{+} *$ |
|  | $\mathrm{NO}^{+}$ | 87 | 85 | 2.3 E-09 | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{O}^{+}$ |
| Ethanol | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 47 | 100 | $2.7 \mathrm{E}-09$ | $\mathrm{C}_{2} \mathrm{H}_{7} \mathrm{O}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 65 |  |  | $\mathrm{C}_{2} \mathrm{H}_{7} \mathrm{O}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 83 |  |  | $\mathrm{C}_{2} \mathrm{H}_{7} \mathrm{O}^{+} \cdot\left(\mathrm{H}_{2} \mathrm{O}\right)_{2}$ |
| Isobutyl alcohol | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 57 | 100 | 2.7 E-09 | $\mathrm{C}_{4} \mathrm{H}_{9}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 73 | 95 | $2.4 \mathrm{E}-09$ | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{O}^{+}$ |
|  | $\mathrm{O}_{2}^{+}$ | 33 | 50 | 2.5 E-09 | $\mathrm{CH}_{5} \mathrm{O}^{+}$ |
| Ketones |  |  |  |  |  |
| Acetone | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 59 | 100 | $3.9 \mathrm{E}-09$ | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{O}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 88 | 100 | 1.2 E-09 | $\mathrm{NO}^{+} \cdot \mathrm{C}_{3} \mathrm{H}_{6} \mathrm{O}$ |
| Acetoin | $\mathrm{O}_{2}^{+}$ | 88 | 20 | $2.5 \mathrm{E}-09$ | $\mathrm{C}_{4} \mathrm{H}_{8} \mathrm{O}_{2}^{+*}$ |
| Butanone | $\mathrm{NO}^{+}$ | 102 | 100 | $2.8 \mathrm{E}-09$ | $\mathrm{NO}^{+} \cdot \mathrm{C}_{4} \mathrm{H}_{8} \mathrm{O}^{*}$ |
| 2-pentanone | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 87 | 100 | 3.9 E-09 | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{O}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 105 |  | 3.9 E-09 | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{O}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
|  | $\mathrm{NO}^{+}$ | 116 | 100 | 3.1 E-09 | $\mathrm{NO}^{+} \cdot \mathrm{C}_{5} \mathrm{H}_{10} \mathrm{O}^{+}$ |
| Sulfur compounds |  |  |  |  |  |
| Hydrogen sulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 35 | 100 | 1.6 E-09 | $\mathrm{H}_{3} \mathrm{~S}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 53 |  | 1.6 E-09 | $\mathrm{H}_{3} \mathrm{~S}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
|  | $\mathrm{O}_{2}^{+}$ | 34 | 100 | 1.4 E-09 | $\mathrm{H}_{2} \mathrm{~S}^{+}$ |
| Carbon disulfide | $\mathrm{O}_{2}^{+}$ | 76 | 100 | 7.0 E-10 | CS2 ${ }^{+*}$ |
| Dimethyl sulfide | $\mathrm{NO}^{+}$ | 62 | 100 | $2.2 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}^{+*}$ |
| Dimethyl disulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 95 | 100 | 2.6 E-09 | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S} 2 \cdot \mathrm{H}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 94 | 100 | $2.4 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}^{+}$ |
| Methyl mercaptan | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 49 | 100 | $1.8 \mathrm{E}-09$ |  |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 67 |  | 1.8 E-09 | $\mathrm{CH}_{4} \mathrm{~S} \cdot \mathrm{H}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
| Esters |  |  |  |  |  |
| Ethyl acetate |  |  |  |  |  |
|  | $\mathrm{O}_{2}^{+}$ | 31 | 20 | $2.4 \mathrm{E}-09$ | $\mathrm{CH}_{3} \mathrm{O}^{+}$ |
| Amines |  |  |  |  |  |
| Ammonia | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 18 | 100 | 2.6 E-09 | $\mathrm{NH}_{4}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 36 |  | $2.6 \mathrm{E}-09$ | $\mathrm{NH}_{4}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
|  | $\mathrm{O}_{2}^{+}$ | 17 | 100 | $2.4 \mathrm{E}-09$ | $\mathrm{NH}_{3}^{+}$ |
| Dimethylamine | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 46 | 100 | 2.1 E-09 | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~N} \cdot \mathrm{H}^{+*}$ |
| Trimethylamine | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 58 | 10 | 2.0 E-09 | $\mathrm{C}_{3} \mathrm{H}_{8} \mathrm{~N}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 60 | 90 | 2.0 E-09 | $\left(\mathrm{CH}_{3}\right)_{3} \mathrm{~N} \cdot \mathrm{H}^{+}$ |
| Others |  |  |  |  |  |
| Ethylene oxide | $\mathrm{NO}^{+}$ | 74 | 100 | 1.0 E-10 | $\mathrm{C}_{2} \mathrm{H}_{4} \mathrm{O} \cdot \mathrm{NO}^{+*}$ |

Table B.7: Volatile organic compounds (VOCs) quantified in packaged shrimp samples with SIFT-MS: mass to charge ratios ( $\mathrm{m} / \mathrm{z}$ ), branching ratios (b), reaction rate coefficients (k) and product ions. Product ions denoted with * were selected for quantifying the respective VOC.

| VOC | Session 1 |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | L4: $30 / 0 / 70\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4{ }^{\circ} \mathrm{C}$ |  |  |  | H4: 50/0/50 $\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4{ }^{\circ} \mathrm{C}$ |  |  |  |
|  | 0 | 3 | 5 | 10 | 0 | 3 | 7 | 12 |
| 2,3-butanediol | 2.65 | 2.34 | 2.44 | 2.73 | 4.33 | 4.59 | 37.51 | 34.83 |
| 2-pentanone | 0.95 | 1.17 | 1.19 | 0.82 | 1.13 | 1.12 | 1.34 | 2.42 |
| 2-propanol | 10.52 | 8.63 | 11.51 | 7.73 | 11.36 | 12.40 | 13.89 | 16.60 |
| 3-methyl-1-butanol | 0.75 | 0.66 | 1.00 | 0.49 | 1.39 | 1.03 | 1.60 | 2.34 |
| Acetic acid | 11.99 | 9.08 | 6.36 | 7.29 | 17.95 | 19.99 | 75.33 | 51.60 |
| Acetoin | 18.48 | 3.93 | 3.16 | 4.56 | 12.11 | 5.50 | 12.98 | 31.41 |
| Acetone | 22.30 | 19.25 | 16.36 | 18.80 | 35.32 | 40.58 | 47.28 | 150.84 |
| Ammonia | 5.81 | 3.77 | 5.98 | 4.42 | 6.27 | 8.13 | 58.39 | 34.14 |
| Butanone | 155.55 | 93.34 | 137.82 | 48.47 | 226.43 | 199.70 | 168.54 | 151.97 |
| Carbon disulfide | 34.34 | 62.34 | 122.93 | 51.24 | 39.87 | 142.00 | 202.63 | 202.55 |
| Dimethyl amine | 0.91 | 1.14 | 1.82 | 2.50 | 1.28 | 1.37 | 6.98 | 5.05 |
| Dimethyl disulfide | 0.57 | 0.52 | 0.76 | 1.13 | 0.47 | 0.81 | 7.54 | 11.18 |
| Dimethyl sulfide | 3.38 | 2.44 | 7.49 | 39.92 | 9.09 | 11.16 | 150.14 | 212.95 |
| Ethanol | 994.07 | 728.73 | 1222.02 | 2737.57 | 1129.69 | 1306.72 | 16208.06 | 22266.88 |
| Ethyl acetate | 37.95 | 40.63 | 247.18 | 384.06 | 53.81 | 100.14 | 1778.90 | 948.07 |
| Ethylene oxide | 110.49 | 50.18 | 154.43 | 102.71 | 262.56 | 173.73 | 932.76 | 710.50 |
| Hydrogen sulfide | 449.10 | 341.63 | 348.01 | 56.98 | 303.01 | 517.89 | 963.57 | 1818.83 |
| Isobutyl alcohol | 5.63 | 2.60 | 3.70 | 11.49 | 10.37 | 4.04 | 19.73 | 27.80 |
| Methyl mercaptan | 35.52 | 50.54 | 91.22 | 50.69 | 14.42 | 34.90 | 145.82 | 266.79 |
| Trimethyl amine | 55.30 | 7.22 | 20.14 | 130.91 | 14.15 | 100.08 | 1405.43 | 522.92 |
| VOC | Session 2 |  |  |  |  |  |  |  |
|  | L4: $30 / 0 / 70\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4{ }^{\circ} \mathrm{C}$ |  |  |  | H4: $50 / 0 / 50\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4{ }^{\circ} \mathrm{C}$ |  |  |  |
|  | 0 | 3 | 5 | 10 | 0 | 3 | 7 | 12 |
| 2,3-butanediol | 3.05 | 2.38 | 2.19 | 3.89 | 5.03 | 4.73 | 17.80 | 23.68 |
| 2-pentanone | 1.38 | 1.28 | 1.02 | 1.20 | 1.43 | 1.62 | 1.44 | 1.38 |
| 2-propanol | 11.02 | 9.43 | 9.62 | 8.36 | 17.24 | 11.64 | 14.30 | 14.04 |
| 3-methyl-1-butanol | 1.13 | 1.00 | 0.85 | 0.87 | 0.93 | 0.92 | 1.81 | 1.90 |
| Acetic acid | 12.49 | 10.36 | 5.12 | 7.94 | 18.12 | 25.38 | 38.90 | 31.82 |
| Acetoin | 17.13 | 2.66 | 2.48 | 5.66 | 16.51 | 6.00 | 8.08 | 8.69 |
| Acetone | 19.25 | 16.81 | 18.19 | 25.32 | 42.19 | 38.45 | 65.30 | 47.60 |
| Ammonia | 6.11 | 4.21 | 5.68 | 4.68 | 7.21 | 11.07 | 10.35 | 32.66 |
| Butanone | 149.56 | 106.80 | 144.13 | 59.86 | 243.50 | 169.99 | 187.97 | 199.23 |
| Carbon disulfide | 30.61 | 55.98 | 113.27 | 52.49 | 36.71 | 117.25 | 178.67 | 214.73 |
| Dimethyl amine | 0.77 | 0.93 | 1.44 | 2.62 | 1.29 | 1.73 | 6.65 | 9.09 |
| Dimethyl disulfide | 0.57 | 0.39 | 0.38 | 2.20 | 0.61 | 0.62 | 3.15 | 2.80 |
| Dimethyl sulfide | 3.25 | 2.91 | 5.28 | 58.86 | 7.97 | 12.74 | 105.95 | 113.04 |
| Ethanol | 1090.07 | 729.80 | 851.40 | 3982.19 | 1995.45 | 1086.39 | 6710.63 | 9136.69 |
| Ethyl acetate | 35.03 | 31.72 | 73.93 | 358.54 | 41.46 | 129.84 | 1810.73 | 2560.65 |
| Ethylene oxide | 131.58 | 83.33 | 107.21 | 107.97 | 231.48 | 191.64 | 799.89 | 1063.97 |
| Hydrogen sulfide | 439.92 | 454.60 | 675.78 | 941.84 | 288.73 | 356.26 | 108.36 | 592.03 |
| Isobutyl alcohol | 6.31 | 2.10 | 3.08 | 7.97 | 13.19 | 5.40 | 7.88 | 8.42 |
| Methyl mercaptan | 35.29 | 64.70 | 86.98 | 86.71 | 17.79 | 31.62 | 88.39 | 107.96 |
| Trimethyl amine | 52.16 | 5.02 | 10.11 | 125.94 | 22.50 | 211.87 | 604.96 | 514.63 |

Table B.8: Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of shrimp samples used for labelling tests.

| VOC | Session 1 |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | L4: $30 / 0 / 70\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4{ }^{\circ} \mathrm{C}$ |  |  |  | H4: 50/0/50 $\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4{ }^{\circ} \mathrm{C}$ |  |  |  |
|  | 0 | 3 | 5 | 10 | 0 | 3 | 7 | 12 |
| 2,3-butanediol | 2.65 | 2.34 | 2.44 | 2.73 | 4.33 | 4.59 | 37.51 | 34.83 |
| 2-pentanone | 0.95 | 1.17 | 1.19 | 0.82 | 1.13 | 1.12 | 1.34 | 2.42 |
| 2-propanol | 10.52 | 8.63 | 11.51 | 7.73 | 11.36 | 12.40 | 13.89 | 16.60 |
| 3-methyl-1-butanol | 0.75 | 0.66 | 1.00 | 0.49 | 1.39 | 1.03 | 1.60 | 2.34 |
| Acetic acid | 11.99 | 9.08 | 6.36 | 7.29 | 17.95 | 19.99 | 75.33 | 51.60 |
| Acetoin | 18.48 | 3.93 | 3.16 | 4.56 | 12.11 | 5.50 | 12.98 | 31.41 |
| Acetone | 22.30 | 19.25 | 16.36 | 18.80 | 35.32 | 40.58 | 47.28 | 150.84 |
| Ammonia | 5.81 | 3.77 | 5.98 | 4.42 | 6.27 | 8.13 | 58.39 | 34.14 |
| Butanone | 155.55 | 93.34 | 137.82 | 48.47 | 226.43 | 199.70 | 168.54 | 151.97 |
| Carbon disulfide | 34.34 | 62.34 | 122.93 | 51.24 | 39.87 | 142.00 | 202.63 | 202.55 |
| Dimethyl amine | 0.91 | 1.14 | 1.82 | 2.50 | 1.28 | 1.37 | 6.98 | 5.05 |
| Dimethyl disulfide | 0.57 | 0.52 | 0.76 | 1.13 | 0.47 | 0.81 | 7.54 | 11.18 |
| Dimethyl sulfide | 3.38 | 2.44 | 7.49 | 39.92 | 9.09 | 11.16 | 150.14 | 212.95 |
| Ethanol | 994.07 | 728.73 | 1222.02 | 2737.57 | 1129.69 | 1306.72 | 16208.06 | 22266.88 |
| Ethyl acetate | 37.95 | 40.63 | 247.18 | 384.06 | 53.81 | 100.14 | 1778.90 | 948.07 |
| Ethylene oxide | 110.49 | 50.18 | 154.43 | 102.71 | 262.56 | 173.73 | 932.76 | 710.50 |
| Hydrogen sulfide | 449.10 | 341.63 | 348.01 | 56.98 | 303.01 | 517.89 | 963.57 | 1818.83 |
| Isobutyl alcohol | 5.63 | 2.60 | 3.70 | 11.49 | 10.37 | 4.04 | 19.73 | 27.80 |
| Methyl mercaptan | 35.52 | 50.54 | 91.22 | 50.69 | 14.42 | 34.90 | 145.82 | 266.79 |
| Trimethyl amine | 55.30 | 7.22 | 20.14 | 130.91 | 14.15 | 100.08 | 1405.43 | 522.92 |


| VOC | Session 2 |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | L4: $30 / 0 / 70\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ |  |  |  | H4: 50/0/50 $\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ |  |  |  |
|  | 0 | 3 | 5 | 10 | 0 | 3 | 5 | 7 |
| 2,3-butanediol | 3.05 | 2.38 | 2.19 | 3.89 | 5.03 | 4.73 | 17.80 | 23.68 |
| 2-pentanone | 1.38 | 1.28 | 1.02 | 1.20 | 1.43 | 1.62 | 1.44 | 1.38 |
| 2-propanol | 11.02 | 9.43 | 9.62 | 8.36 | 17.24 | 11.64 | 14.30 | 14.04 |
| 3-methyl-1-butanol | 1.13 | 1.00 | 0.85 | 0.87 | 0.93 | 0.92 | 1.81 | 1.90 |
| Acetic acid | 12.49 | 10.36 | 5.12 | 7.94 | 18.12 | 25.38 | 38.90 | 31.82 |
| Acetoin | 17.13 | 2.66 | 2.48 | 5.66 | 16.51 | 6.00 | 8.08 | 8.69 |
| Acetone | 19.25 | 16.81 | 18.19 | 25.32 | 42.19 | 38.45 | 65.30 | 47.60 |
| Ammonia | 6.11 | 4.21 | 5.68 | 4.68 | 7.21 | 11.07 | 10.35 | 32.66 |
| Butanone | 149.56 | 106.80 | 144.13 | 59.86 | 243.50 | 169.99 | 187.97 | 199.23 |
| Carbon disulfide | 30.61 | 55.98 | 113.27 | 52.49 | 36.71 | 117.25 | 178.67 | 214.73 |
| Dimethyl amine | 0.77 | 0.93 | 1.44 | 2.62 | 1.29 | 1.73 | 6.65 | 9.09 |
| Dimethyl disulfide | 0.57 | 0.39 | 0.38 | 2.20 | 0.61 | 0.62 | 3.15 | 2.80 |
| Dimethyl sulfide | 3.25 | 2.91 | 5.28 | 58.86 | 7.97 | 12.74 | 105.95 | 113.04 |
| Ethanol | 1090.07 | 729.80 | 851.40 | 3982.19 | 1995.45 | 1086.39 | 6710.63 | 9136.69 |
| Ethyl acetate | 35.03 | 31.72 | 73.93 | 358.54 | 41.46 | 129.84 | 1810.73 | 2560.65 |
| Ethylene oxide | 131.58 | 83.33 | 107.21 | 107.97 | 231.48 | 191.64 | 799.89 | 1063.97 |
| Hydrogen sulfide | 439.92 | 454.60 | 675.78 | 941.84 | 288.73 | 356.26 | 108.36 | 592.03 |
| Isobutyl alcohol | 6.31 | 2.10 | 3.08 | 7.97 | 13.19 | 5.40 | 7.88 | 8.42 |
| Methyl mercaptan | 35.29 | 64.70 | 86.98 | 86.71 | 17.79 | 31.62 | 88.39 | 107.96 |
| Trimethyl amine | 52.16 | 5.02 | 10.11 | 125.94 | 22.50 | 211.87 | 604.96 | 514.63 |

Table B.9: Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of shrimp samples used for ranking tests.

## B.4. Quantification of VOCs in Atlantic salmon

| VOC | Precursor | $\mathrm{m} / \mathrm{z}$ | b(\%) | k | Product ion |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Acids |  |  |  |  |  |
| Acetic acid | $\mathrm{NO}^{+}$ | 90 | 100 | $9.0 \mathrm{E}-10$ | $\mathrm{NO}^{+} \cdot \mathrm{CH}_{3} \mathrm{COOH}^{*}$ |
|  | $\mathrm{NO}^{+}$ | 108 |  | $9.0 \mathrm{E}-10$ | $\mathrm{NO}^{+} \cdot \mathrm{CH}_{3} \mathrm{COOH} \cdot \mathrm{H}_{2} \mathrm{O}$ |
| 3-methylbutanoic acid | $\mathrm{NO}^{+}$ | 132 | 70 | $2.5 \mathrm{E}-09$ | $\mathrm{C}_{5} \mathrm{H}_{10} \mathrm{O}_{2} \cdot \mathrm{NO}^{+*}$ |
| Alcohols |  |  |  |  |  |
| 2,3-butanediol | $\mathrm{NO}^{+}$ | 89 | 100 | 2.3 E-09 | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{O}_{2}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 107 |  | $2.3 \mathrm{E}-09$ | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{O}_{2}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
| Ethanol | $\mathrm{NO}^{+}$ | 45 | 100 | 1.2 E-09 | $\mathrm{C}_{2} \mathrm{H}_{5} \mathrm{O}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 63 |  | $1.2 \mathrm{E}-09$ | $\mathrm{C}_{2} \mathrm{H}_{5} \mathrm{O}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
|  | $\mathrm{NO}^{+}$ | 81 |  | 1.2 E-09 | $\mathrm{C}_{2} \mathrm{H}_{5} \mathrm{O}^{+} \cdot 2\left(\mathrm{H}_{2} \mathrm{O}\right)_{2}$ |
| 3-methyl-1-butanol | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 71 | 100 | $3.0 \mathrm{E}-09$ | $\mathrm{C}_{5} \mathrm{H}_{11}^{+} *$ |
|  | $\mathrm{NO}^{+}$ | 87 | 85 | 2.3 E-09 | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{O}^{+}$ |
| Isobutyl alcohol | $\mathrm{NO}^{+}$ | 73 | 95 | 2.4 E-09 | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{O}^{+*}$ |
|  | $\mathrm{O}_{2}^{+}$ | 33 | 50 | $2.5 \mathrm{E}-09$ | $\mathrm{CH}_{5} \mathrm{O}^{+}$ |
| Aldehydes |  |  |  |  |  |
| 3-methylbutanal | $\mathrm{NO}^{+}$ | 85 | 100 | $2.4 \mathrm{E}-09$ | $\mathrm{C}_{5} \mathrm{H}_{9} \mathrm{O}^{+*}$ |
| Aromatic Hydrocarbons |  |  |  |  |  |
| Ethyl benzene | $\mathrm{NO}^{+}$ | 106 | 100 | $2.0 \mathrm{E}-09$ | $\mathrm{C}_{8} \mathrm{H}_{10}^{+}$* |
| Propyl benzene | $\mathrm{NO}^{+}$ | 120 | 100 | $2.0 \mathrm{E}-09$ | $\mathrm{C}_{9} \mathrm{H}_{12}^{+}$* |
| Styrene | $\mathrm{NO}^{+}$ | 104 | 100 | $1.7 \mathrm{E}-09$ | $\mathrm{C}_{8} \mathrm{H}_{8}^{+*}$ |
| Ketones |  |  |  |  |  |
| Acetone | $\mathrm{NO}^{+}$ | 88 | 100 | 1.2 E-09 | $\mathrm{NO}^{+} . \mathrm{C}_{3} \mathrm{H}_{6} \mathrm{O}^{*}$ |
| Acetoin | $\mathrm{NO}^{+}$ | 118 | 100 | $2.5 \mathrm{E}-09$ | $\mathrm{C}_{4} \mathrm{H}_{8} \mathrm{O}_{2} \cdot \mathrm{NO}^{+*}$ |
| 2,3-butanedione | $\mathrm{NO}^{+}$ | 65 | 65 | $1.3 \mathrm{E}-09$ | $\mathrm{C}_{4} \mathrm{H}_{6} \mathrm{O}_{2}^{+*}$ |
| Butanone | $\mathrm{NO}^{+}$ | 102 | 100 | $2.8 \mathrm{E}-09$ | $\mathrm{NO}^{+} \cdot \mathrm{C}_{4} \mathrm{H}_{8} \mathrm{O}^{*}$ |
| Sulfur compounds |  |  |  |  |  |
| Carbon disulfide | $\mathrm{O}_{2}^{+}$ | 76 | 100 | 7.0 E-10 | $\mathrm{CS}_{2}^{+*}$ |
| Dimethyl sulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 63 | 100 | $2.5 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}^{+*}$ |
|  | $\mathrm{NO}^{+}$ | 62 | 100 | $2.2 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}^{+}$ |
|  | $\mathrm{O}_{2}^{+}$ | 47 | 25 | $2.2 \mathrm{E}-09$ | $\mathrm{CH}_{3} \mathrm{~S}^{+}$ |
|  | $\mathrm{O}_{2}^{+}$ | 62 | 60 | $2.2 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}^{+}$ |
| Dimethyl disulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 95 | 100 | $2.6 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}_{2} \cdot \mathrm{H}^{+}$ |
|  | $\mathrm{NO}^{+}$ | 94 | 100 | $2.4 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~S}_{2}^{+*}$ |
| Dimethyl trisulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 127 | 100 | 2.8 E-09 | $\mathrm{C}_{2} \mathrm{H}_{6} \mathrm{~S}_{3} \mathrm{H}^{+}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 145 |  | 2.8 E-09 | $\mathrm{C}_{2} \mathrm{H}_{6} \mathrm{~S}_{3} \mathrm{H}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
|  | $\mathrm{NO}^{+}$ | 126 | 100 | $1.9 \mathrm{E}-09$ | $\mathrm{C}_{2} \mathrm{H}_{6} \mathrm{~S}_{3}^{+*}$ |
| Hydrogen sulfide | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 35 | 100 | $1.6 \mathrm{E}-09$ | $\mathrm{H}_{3} \mathrm{~S}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 53 |  | 1.6 E-09 | $\mathrm{H}_{3} \mathrm{~S}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
| Methyl mercaptan | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 49 | 100 | 1.8 E-09 | $\mathrm{CH}_{4} \mathrm{~S} \cdot \mathrm{H}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 67 |  | $1.8 \mathrm{E}-09$ | $\mathrm{CH}_{4} \mathrm{~S} \cdot \mathrm{H}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
| Esters |  |  |  |  |  |
| Ethyl acetate | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 89 | 100 | $2.9 \mathrm{E}-09$ | $\mathrm{CH}_{3} \mathrm{COOC}_{2} \mathrm{H}_{5} \cdot \mathrm{H}^{+*}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 107 |  | $2.9 \mathrm{E}-09$ | $\mathrm{CH}_{3} \mathrm{COOC}_{2} \mathrm{H}_{5} \cdot \mathrm{H}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
|  | $\mathrm{O}_{2}^{+}$ | 31 | 20 | $2.4 \mathrm{E}-09$ | $\mathrm{CH}_{3} \mathrm{O}^{+}$ |
|  | $\mathrm{O}_{2}^{+}$ | 61 | 40 | $2.4 \mathrm{E}-09$ | $\mathrm{C}_{2} \mathrm{H}_{5} \mathrm{O}_{2}^{+}$ |
| Amines |  |  |  |  |  |
| Ammonia | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 18 | 100 | 2.6 E-09 | $\mathrm{NH}_{4}^{+}$ |
|  | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 36 |  | 2.6 E-09 | $\mathrm{NH}_{4}^{+} \cdot \mathrm{H}_{2} \mathrm{O}$ |
|  | $\mathrm{O}_{2}^{+}$ | 17 | 100 | $2.6 \mathrm{E}-09$ | $\mathrm{NH}_{3}^{+*}$ |
| Dimethyl amine | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 46 | 100 | $2.1 \mathrm{E}-09$ | $\left(\mathrm{CH}_{3}\right)_{2} \mathrm{~N} \cdot \mathrm{H}^{+*}$ |
| Piperidine | $\mathrm{H}_{3} \mathrm{O}^{+}$ | 86 | 90 | $3.4 \mathrm{E}-09$ | $\mathrm{C}_{5} \mathrm{H}_{12} \mathrm{~N}^{+*}$ |
| Trimethyl amine | $\mathrm{NO}^{+}$ | 59 | 100 | 1.6 E-09 | $\left(\mathrm{CH}_{3}\right)_{3} \mathrm{~N}^{+*}$ |

Table B.10: Volatile organic compounds (VOCs) quantified in packaged salmon samples with SIFT-MS: product ions, mass to charge ratios ( $\mathrm{m} / \mathrm{z}$ ), branching ratios (b) and reaction rate coefficients (k). Product ions denoted with * were selected for quantifying the respective VOC.

|  | L4: $60 / 5 / 35\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ |  |  |  | M4: $60 / 21 / 19\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ |  |  |  | $\text { H4: } 60 / 40 / 0\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}(\text { Session } 1)$ <br> Session 1 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| VOC | 1 | 5 | 7 | 11 | 1 | 5 | 9 | 11 | 1 | 5 | 7 | 11 |
| 2,3-butanediol | 3.364 | 3.346 | 18.252 | 56.706 | 6.878 | 2.776 | 37.537 | 60.387 | 3.364 | 3.346 | 18.252 | 56.706 |
| 2,3-butanedione | 30.827 | 25.241 | 27.975 | 35.713 | 8.058 | 9.340 | 30.651 | 46.667 | 30.827 | 25.241 | 27.975 | 35.713 |
| 3-methyl-1-butanol | 69.969 | 38.739 | 74.312 | 108.830 | 78.425 | 36.895 | 102.129 | 234.800 | 69.969 | 38.739 | 74.312 | 108.830 |
| 3-methylbutanal | 11.077 | 13.685 | 22.287 | 37.672 | 11.515 | 6.325 | 38.273 | 41.292 | 11.077 | 13.685 | 22.287 | 37.672 |
| 3-methylbutanoic acid | 2.706 | 1.576 | 2.411 | 1.785 | 3.800 | 3.371 | 3.778 | 4.776 | 2.706 | 1.576 | 2.411 | 1.785 |
| Acetic acid | 19.388 | 12.849 | 49.566 | 141.172 | 28.330 | 28.822 | 61.852 | 79.713 | 19.388 | 12.849 | 49.566 | 141.172 |
| Acetoin | 2.876 | 9.199 | 208.459 | 140.520 | 36.278 | 21.813 | 441.317 | 550.187 | 2.876 | 9.199 | 208.459 | 140.520 |
| Acetone | 33.458 | 79.570 | 74.324 | 103.152 | 220.917 | 39.446 | 84.347 | 94.515 | 33.458 | 79.570 | 74.324 | 103.152 |
| Ammonia | 32.098 | 44.324 | 35.209 | 134.844 | 45.046 | 46.719 | 36.622 | 74.282 | 32.098 | 44.324 | 35.209 | 134.844 |
| Butanone | 9.874 | 6.547 | 6.979 | 9.931 | 10.283 | 10.667 | 7.900 | 10.101 | 9.874 | 6.547 | 6.979 | 9.931 |
| Carbon disulfide | 110.238 | 57.162 | 182.141 | 113.236 | 79.248 | 22.806 | 57.775 | 69.204 | 110.238 | 57.162 | 182.141 | 113.236 |
| Dimethyl amine | 5.258 | 6.511 | 12.332 | 11.327 | 8.094 | 11.188 | 18.799 | 26.877 | 5.258 | 6.511 | 12.332 | 11.327 |
| Dimethyl disulfide | 0.675 | 0.599 | 1.043 | 1.000 | 0.239 | 0.838 | 0.945 | 1.584 | 0.675 | 0.599 | 1.043 | 1.000 |
| Dimethyl sulfide | 272.343 | 249.033 | 418.534 | 487.090 | 322.210 | 310.615 | 435.056 | 454.743 | 272.343 | 249.033 | 418.534 | 487.090 |
| Dimethyl trisulfide | 7.983 | 5.064 | 6.154 | 3.170 | 1.315 | 0.923 | 2.629 | 3.510 | 7.983 | 5.064 | 6.154 | 3.170 |
| Ethanol | 780.155 | 530.175 | 2563.089 | 6622.378 | 177.903 | 381.374 | 4103.657 | 8793.104 | 780.155 | 530.175 | 2563.089 | 6622.378 |
| Ethyl acetate | 8.726 | 17.981 | 326.924 | 250.536 | 44.736 | 33.089 | 603.213 | 760.385 | 8.726 | 17.981 | 326.924 | 250.536 |
| Ethyl benzene | 5.986 | 1.703 | 4.189 | 30.100 | 5.341 | 7.145 | 11.189 | 16.679 | 5.986 | 1.703 | 4.189 | 30.100 |
| Hydrogen sulfide | 2.253 | 4.445 | 96.559 | 278.736 | 1.665 | 8.387 | 356.215 | 416.202 | 2.253 | 4.445 | 96.559 | 278.736 |
| Isobutyl alcohol | 6.851 | 2.727 | 7.308 | 13.374 | 22.323 | 1.541 | 9.511 | 16.432 | 6.851 | 2.727 | 7.308 | 13.374 |
| Methyl mercaptan | 9.821 | 5.152 | 16.568 | 31.769 | 4.683 | 8.976 | 22.949 | 45.037 | 9.821 | 5.152 | 16.568 | 31.769 |
| Piperidine | 1.667 | 1.005 | 2.043 | 1.056 | 0.964 | 0.293 | 0.753 | 0.390 | 1.667 | 1.005 | 2.043 | 1.056 |
| Propyl benzene | 10.927 | 2.982 | 6.417 | 14.685 | 9.280 | 6.238 | 10.377 | 11.409 | 10.927 | 2.982 | 6.417 | 14.685 |
| Styrene | 9.807 | 2.006 | 7.139 | 7.275 | 11.683 | 7.741 | 5.802 | 12.101 | 9.807 | 2.006 | 7.139 | 7.275 |
| Trimethyl amine | 12.749 | 9.738 | 9.870 | 11.359 | 12.370 | 6.603 | 7.913 | 15.246 | 12.749 | 9.738 | 9.870 | 11.359 |

Table B.11: Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of salmon samples used for ranking tests.

| VOC | $\begin{gathered} \text { AN4: } 0 / 0 / 100\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C} \\ \text { Session } 1 \end{gathered}$ |  |  |  | ANH4: $60 / 0 / 40\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$Session 1 |  |  |  | $\begin{aligned} & \hline \text { A4: air } 4^{\circ} \mathrm{C} \\ & \text { Session } 1 \end{aligned}$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | 5 | 9 | 13 | 1 | 5 | 9 | 11 | 1 | 5 | 9 | 11 |
| 2,3-butanediol | 1.817 | 5.791 | 28.804 | 13.462 | 3.027 | 2.516 | 7.831 | 8.429 | 1.807 | 5.898 | 44.987 | 49.511 |
| 2,3-butanedione | 22.211 | 18.109 | 19.722 | 14.926 | 47.125 | 25.119 | 29.958 | 28.938 | 37.922 | 24.955 | 33.170 | 38.096 |
| 3-methyl-1-butanol | 140.511 | 150.848 | 140.696 | 163.813 | 73.264 | 48.142 | 67.866 | 67.110 | 74.189 | 153.845 | 276.768 | 310.116 |
| 3 -methylbutanal | 14.273 | 14.614 | 18.537 | 15.274 | 13.193 | 12.485 | 17.501 | 18.032 | 14.148 | 15.931 | 35.417 | 44.608 |
| 3 -methylbutanoic acid | 1.760 | 0.811 | 1.179 | 1.860 | 2.072 | 2.039 | 1.967 | 2.217 | 0.896 | 0.984 | 2.403 | 4.251 |
| Acetic acid | 13.426 | 16.110 | 70.561 | 38.361 | 10.953 | 14.506 | 46.138 | 45.036 | 11.526 | 11.615 | 46.921 | 45.513 |
| Acetoin | 5.351 | 18.909 | 16.040 | 24.730 | 5.016 | 6.441 | 15.452 | 22.000 | 4.237 | 105.367 | 148.291 | 281.229 |
| Acetone | 35.226 | 42.492 | 41.051 | 59.848 | 43.304 | 59.082 | 56.899 | 70.754 | 38.000 | 35.750 | 144.281 | 190.539 |
| Ammonia | 34.435 | 16.639 | 19.939 | 24.681 | 53.073 | 31.045 | 27.635 | 82.296 | 37.145 | 17.063 | 23.278 | 47.011 |
| Butanone | 11.834 | 6.882 | 11.186 | 21.382 | 14.369 | 11.416 | 49.146 | 446.962 | 11.406 | 8.169 | 19.957 | 26.239 |
| Carbon disulfide | 107.465 | 41.967 | 65.294 | 33.771 | 55.629 | 86.260 | 97.423 | 101.837 | 86.296 | 40.866 | 74.235 | 48.940 |
| Dimethyl amine | 3.142 | 8.225 | 10.513 | 12.201 | 9.027 | 6.602 | 15.131 | 13.349 | 6.300 | 9.027 | 8.706 | 12.962 |
| Dimethyl disulfide | 0.252 | 0.267 | 1.189 | 1.679 | 0.353 | 0.473 | 1.496 | 2.624 | 0.183 | 0.311 | 0.903 | 0.630 |
| Dimethyl sulfide | 25.360 | 92.670 | 122.744 | 176.318 | 334.808 | 286.664 | 571.423 | 473.208 | 21.362 | 96.708 | 135.912 | 230.967 |
| Dimethyl trisulfide | 4.608 | 4.595 | 3.928 | 2.566 | 7.086 | 5.167 | 6.868 | 3.355 | 7.283 | 3.239 | 4.998 | 4.047 |
| Ethanol | 345.210 | 3916.576 | 19568.783 | 26360.550 | 518.051 | 683.710 | 19785.308 | 24222.245 | 303.668 | 1895.423 | 7362.198 | 6445.701 |
| Ethyl acetate | 10.617 | 36.421 | 33.361 | 48.235 | 8.187 | 13.401 | 31.784 | 35.666 | 9.303 | 174.918 | 258.240 | 501.393 |
| Ethyl benzene | 1.809 | 1.803 | 3.842 | 5.279 | 4.525 | 2.344 | 2.931 | 5.037 | 2.059 | 1.962 | 11.501 | 12.339 |
| Hydrogen sulfide | 20.273 | 241.807 | 1959.792 | 5509.738 | 9.841 | 86.838 | 4860.913 | 9593.681 | 0.841 | 14.425 | 895.568 | 1571.189 |
| Isobutyl alcohol | 4.129 | 5.007 | 14.715 | 20.639 | 3.625 | 3.673 | 11.136 | 23.611 | 4.994 | 2.824 | 20.683 | 23.984 |
| Methyl mercaptan | 5.089 | 23.624 | 113.674 | 152.460 | 4.236 | 7.548 | 128.001 | 238.482 | 4.615 | 14.055 | 54.775 | 46.275 |
| Piperidine | 1.867 | 2.304 | 1.832 | 1.171 | 2.523 | 1.951 | 2.021 | 2.235 | 3.782 | 1.993 | 1.688 | 2.407 |
| Propyl benzene | 8.319 | 2.041 | 3.706 | 9.197 | 15.770 | 15.920 | 12.110 | 13.335 | 7.371 | 3.028 | 4.700 | 12.772 |
| Styrene | 4.213 | 4.344 | 4.109 | 5.296 | 7.013 | 5.610 | 6.676 | 10.389 | 3.881 | 6.628 | 5.163 | 5.387 |
| Trimethyl amine | 11.603 | 7.517 | 17.206 | 14.899 | 14.454 | 13.759 | 31.543 | 29.048 | 9.287 | 10.274 | 41.669 | 34.351 |

Table B.11: (Continued) Measured concentrations (ppb) of VOCs detected with SIFTMS on each day in the headspace of salmon samples used for ranking tests.

|  | $\begin{gathered} \mathrm{H} 4: 60 / 40 / 0\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C} \\ \text { Session } 2 \end{gathered}$ |  |  |  | $\begin{gathered} \text { AN4: } 0 / 0 / 100\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C} \\ \text { Session } 2 \end{gathered}$ |  |  |  | ANH4: $60 / 0 / 40\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$Session 2 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| VOC | 1 | 3 | 5 | 7 | 3 | 7 | 9 | 11 | 1 | 3 | 5 | 7 |
| 2,3-butanediol | 2.349 | 2.966 | 2.056 | 3.464 | 1.885 | 15.334 | 27.244 | 15.070 | 1.790 | 2.374 | 2.516 | 3.554 |
| 2,3-butanedione | 11.744 | 33.509 | 51.312 | 31.452 | 22.699 | 20.416 | 6.057 | 19.102 | 7.271 | 32.963 | 25.119 | 7.715 |
| 3-methyl-1-butanol | 27.036 | 72.192 | 34.698 | 33.667 | 222.343 | 108.395 | 34.452 | 299.025 | 20.481 | 63.772 | 48.142 | 26.163 |
| 3-methylbutanal | 7.651 | 13.264 | 10.004 | 23.150 | 17.394 | 16.391 | 11.390 | 19.965 | 4.844 | 12.683 | 12.485 | 10.033 |
| 3-methylbutanoic acid | 2.000 | 1.579 | 3.187 | 2.482 | 2.456 | 1.235 | 1.227 | 2.127 | 2.025 | 1.434 | 2.039 | 2.290 |
| Acetic acid | 14.329 | 16.350 | 15.393 | 13.815 | 17.356 | 48.074 | 75.567 | 43.852 | 14.747 | 14.309 | 14.506 | 14.267 |
| Acetoin | 4.778 | 4.005 | 5.521 | 22.631 | 4.482 | 18.125 | 16.692 | 26.575 | 4.241 | 3.867 | 6.441 | 8.990 |
| Acetone | 44.442 | 50.998 | 47.431 | 54.742 | 32.462 | 42.402 | 30.267 | 60.285 | 40.536 | 51.842 | 59.082 | 48.069 |
| Ammonia | 54.757 | 35.011 | 31.729 | 29.155 | 25.210 | 19.652 | 19.115 | 47.286 | 49.647 | 33.505 | 31.045 | 29.515 |
| Butanone | 13.029 | 11.804 | 10.992 | 13.189 | 11.129 | 7.956 | 15.868 | 11.750 | 15.107 | 10.616 | 11.416 | 9.796 |
| Carbon disulfide | 65.889 | 82.630 | 77.726 | 55.576 | 79.594 | 57.613 | 45.324 | 75.762 | 148.104 | 126.300 | 86.260 | 42.201 |
| Dimethyl amine | 8.473 | 5.274 | 7.608 | 6.986 | 4.644 | 9.953 | 10.868 | 13.466 | 9.153 | 4.857 | 6.602 | 7.220 |
| Dimethyl disulfide | 0.296 | 0.426 | 0.381 | 0.291 | 0.228 | 0.325 | 0.546 | 1.541 | 0.372 | 0.411 | 0.473 | 0.282 |
| Dimethyl sulfide | 305.239 | 284.246 | 261.150 | 298.855 | 27.806 | 88.890 | 121.937 | 181.689 | 326.274 | 279.532 | 286.664 | 290.155 |
| Dimethyl trisulfide | 4.182 | 6.056 | 3.533 | 4.141 | 5.000 | 3.932 | 1.096 | 5.376 | 1.412 | 6.626 | 5.167 | 2.329 |
| Ethanol | 542.226 | 383.931 | 487.583 | 738.949 | 291.600 | 5628.446 | 13016.948 | 28536.568 | 549.600 | 660.944 | 683.710 | 715.748 |
| Ethyl acetate | 6.248 | 8.089 | 9.595 | 36.570 | 13.253 | 30.856 | 35.162 | 55.092 | 8.865 | 6.380 | 13.401 | 18.249 |
| Ethyl benzene | 1.851 | 2.765 | 1.471 | 2.550 | 2.954 | 2.289 | 2.499 | 4.569 | 2.639 | 2.999 | 2.344 | 1.602 |
| Hydrogen sulfide | 1.173 | 1.624 | 5.963 | 3.172 | 78.390 | 589.125 | 1504.663 | 4043.214 | 11.793 | 21.971 | 86.838 | 120.094 |
| Isobutyl alcohol | 2.374 | 2.927 | 2.167 | 2.357 | 5.431 | 4.190 | 5.769 | 25.430 | 3.304 | 3.415 | 3.673 | 4.354 |
| Methyl mercaptan | 5.315 | 5.404 | 5.167 | 7.750 | 5.295 | 35.045 | 78.982 | 189.582 | 4.813 | 4.851 | 7.548 | 7.110 |
| Piperidine | 0.778 | 2.421 | 1.117 | 1.086 | 2.512 | 1.755 | 0.905 | 1.837 | 0.606 | 2.867 | 1.951 | 0.961 |
| Propyl benzene | 15.293 | 7.894 | 13.838 | 15.010 | 7.072 | 4.391 | 3.366 | 15.029 | 13.754 | 8.354 | 15.920 | 16.859 |
| Styrene | 4.215 | 4.783 | 4.809 | 4.997 | 8.731 | 4.912 | 2.616 | 4.777 | 5.688 | 5.290 | 5.610 | 5.254 |
| Trimethyl amine | 13.494 | 12.726 | 16.004 | 19.739 | 9.301 | 8.001 | 17.418 | 16.682 | 15.805 | 15.113 | 13.759 | 11.944 |

Table B.11: (Continued) Measured concentrations (ppb) of VOCs detected with SIFTMS on each day in the headspace of salmon samples used for ranking tests.

| VOC | A4: air $4^{\circ} \mathrm{C}$ (Session 2) Session 2 |  |  |  | H4: 60/40/0 $\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}($ Session 3) |  |  |  | $\begin{gathered} \text { AN4: } 0 / 0 / 100\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}(\text { Session 3) } \\ \text { Session } 3 \end{gathered}$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | 3 | 5 | 7 | 1 | 5 | 9 | 11 | 1 | 5 | 9 | 11 |
| 2,3-butanediol | 1.388 | 1.708 | 3.913 | 9.74 | 4.701 | 5.849 | 7.326 | 8.132 | 1.821 | 4.385 | 7.58 | 8.2 |
| 2,3-butanedione | 12.892 | 12.310 | 13.128 | 39.667 | 13.597 | 17.209 | 21.639 | 23.726 | 36.050 | 24.263 | 15.081 | 26.195 |
| 3-methyl-1-butanol | 63.914 | 201.768 | 124.718 | 285.163 | 76.633 | 96.793 | 103.921 | 97.417 | 61.808 | 38.248 | 109.305 | 126.419 |
| 3 -methylbutanal | 10.963 | 12.976 | 11.195 | 58.148 | 9.099 | 17.365 | 16.052 | 20.820 | 15.690 | 16.935 | 24.563 | 27.965 |
| 3 -methylbutanoic acid | 1.682 | 1.649 | 1.042 | 2.296 | 2.839 | 3.719 | 2.960 | 4.786 | 2.518 | 1.230 | 1.743 | 1.615 |
| Acetic acid | 12.072 | 11.814 | 12.130 | 36.203 | 13.944 | 18.995 | 33.602 | 30.146 | 13.416 | 17.204 | 17.909 | 40.743 |
| Acetoin | 3.883 | 6.949 | 44.063 | 858.603 | 5.509 | 9.550 | 19.732 | 25.666 | 3.618 | 14.589 | 17.885 | 50.286 |
| Acetone | 30.089 | 31.298 | 25.941 | 78.311 | 36.504 | 56.025 | 53.218 | 163.163 | 32.282 | 38.507 | 71.562 | 112.704 |
| Ammonia | 36.398 | 27.416 | 16.921 | 21.645 | 47.877 | 54.489 | 60.961 | 58.157 | 21.811 | 36.061 | 32.619 | 96.918 |
| Butanone | 9.066 | 7.534 | 6.663 | 11.300 | 7.368 | 13.667 | 12.684 | 13.124 | 8.543 | 3.203 | 20.892 | 481.741 |
| Carbon disulfide | 113.118 | 41.633 | 63.370 | 116.868 | 211.452 | 57.460 | 110.498 | 64.871 | 60.241 | 35.125 | 40.978 | 23.908 |
| Dimethyl amine | 4.473 | 4.534 | 4.669 | 13.712 | 6.212 | 8.428 | 10.653 | 9.513 | 6.487 | 6.578 | 15.671 | 20.521 |
| Dimethyl disulfide | 0.493 | 0.357 | 0.214 | 0.779 | 0.391 | 0.656 | 1.404 | 0.950 | 0.345 | 0.150 | 1.382 | 2.172 |
| Dimethyl sulfide | 21.345 | 27.731 | 43.599 | 250.533 | 162.846 | 196.044 | 173.016 | 181.558 | 30.199 | 68.881 | 207.026 | 325.892 |
| Dimethyl trisulfide | 4.432 | 3.863 | 3.057 | 3.595 | 4.295 | 3.726 | 4.598 | 3.592 | 4.728 | 3.072 | 3.670 | 4.055 |
| Ethanol | 259.913 | 180.659 | 695.140 | 4021.798 | 199.664 | 202.291 | 262.983 | 398.029 | 347.009 | 1784.364 | 18913.973 | 29384.565 |
| Ethyl acetate | 10.340 | 16.610 | 85.439 | 1430.190 | 7.697 | 18.813 | 34.346 | 33.828 | 10.476 | 27.872 | 38.901 | 87.462 |
| Ethyl benzene | 2.281 | 3.217 | 2.667 | 11.695 | 2.439 | 7.370 | 6.747 | 5.913 | 1.833 | 1.843 | 5.696 | 7.107 |
| Hydrogen sulfide | 0.725 | 1.614 | 8.217 | 5.594 | 1.462 | 3.736 | 6.428 | 32.201 | 11.624 | 246.254 | 2454.361 | 9646.661 |
| Isobutyl alcohol | 1.491 | 4.232 | 3.084 | 15.132 | 2.075 | 2.873 | 2.938 | 5.234 | 1.648 | 1.559 | 14.554 | 31.959 |
| Methyl mercaptan | 3.097 | 4.238 | 7.577 | 21.488 | 4.973 | 4.265 | 5.078 | 7.989 | 4.903 | 11.027 | 111.668 | 174.984 |
| Piperidine | 1.292 | 1.134 | 1.173 | 1.420 | 1.146 | 1.225 | 1.669 | 1.235 | 1.882 | 1.622 | 2.390 | 1.596 |
| Propyl benzene | 8.506 | 6.304 | 2.795 | 10.487 | 12.498 | 10.651 | 11.722 | 8.053 | 9.052 | 8.696 | 9.240 | 8.956 |
| Styrene | 6.479 | 8.793 | 4.951 | 8.592 | 4.171 | 9.156 | 6.666 | 10.584 | 4.309 | 3.023 | 5.913 | 6.803 |
| Trimethyl amine | 7.388 | 8.001 | 8.061 | 16.765 | 9.027 | 7.810 | 33.926 | 24.970 | 9.813 | 3.104 | 10.074 | 18.848 |

Table B.11: (Continued) Measured concentrations (ppb) of VOCs detected with SIFTMS on each day in the headspace of salmon samples used for ranking tests.

|  | $\begin{gathered} \text { ANH4: } 60 / 0 / 40\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C} \\ \text { Session } 3 \end{gathered}$ |  |  |  | $\begin{gathered} \text { A4: air } 4^{\circ} \mathrm{C} \\ \text { Session } 3 \end{gathered}$ |  |  |  | $\begin{gathered} \mathrm{H} 4: 60 / 40 / 0\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C} \\ \text { Session } 4 \end{gathered}$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| VOC | 1 | 5 | 9 | 11 | 1 | 5 | 9 | 11 | 3 | 5 | 7 | 9 |
| 2,3-butanediol | 5.604 | 7.128 | 9.184 | 25.203 | 1.378 | 11.496 | 41.887 | 48.737 | 1.645 | 4.902 | 7.645 | 6.045 |
| 2,3-butanedione | 27.424 | 18.231 | 25.257 | 24.308 | 16.647 | 35.659 | 50.617 | 27.991 | 9.666 | 9.153 | 14.341 | 30.024 |
| 3-methyl-1-butanol | 72.597 | 96.125 | 53.316 | 91.242 | 45.328 | 69.454 | 369.024 | 314.628 | 46.979 | 95.754 | 134.391 | 103.153 |
| 3 -methylbutanal | 10.335 | 15.997 | 12.952 | 20.503 | 15.380 | 18.723 | 100.458 | 57.569 | 6.672 | 5.578 | 11.354 | 23.819 |
| 3-methylbutanoic acid | 3.218 | 3.608 | 4.699 | 3.872 | 2.247 | 1.492 | 2.219 | 2.363 | 1.128 | 3.391 | 4.861 | 4.882 |
| Acetic acid | 11.430 | 25.425 | 31.813 | 67.132 | 11.059 | 18.577 | 46.599 | 49.634 | 7.134 | 20.584 | 35.801 | 29.862 |
| Acetoin | 6.180 | 9.927 | 27.612 | 67.129 | 3.016 | 15.597 | 969.687 | 775.287 | 3.380 | 10.078 | 10.147 | 22.766 |
| Acetone | 34.499 | 45.686 | 46.411 | 125.314 | 29.483 | 37.921 | 129.165 | 222.891 | 28.043 | 34.691 | 70.651 | 84.229 |
| Ammonia | 49.084 | 60.591 | 57.200 | 57.581 | 25.303 | 35.939 | 35.498 | 97.064 | 29.881 | 62.121 | 76.010 | 64.789 |
| Butanone | 6.643 | 10.738 | 6.363 | 8.817 | 6.550 | 3.669 | 7.214 | 27.587 | 8.327 | 15.290 | 15.205 | 13.786 |
| Carbon disulfide | 162.569 | 90.912 | 71.571 | 78.287 | 16.981 | 50.993 | 76.027 | 49.947 | 118.414 | 70.972 | 70.875 | 137.815 |
| Dimethyl amine | 6.324 | 7.344 | 11.049 | 21.386 | 4.619 | 4.780 | 12.753 | 12.706 | 2.344 | 6.536 | 9.074 | 12.007 |
| Dimethyl disulfide | 0.181 | 0.392 | 0.823 | 0.752 | 0.253 | 0.224 | 1.017 | 0.775 | 0.197 | 0.562 | 0.713 | 0.656 |
| Dimethyl sulfide | 189.641 | 197.212 | 265.138 | 374.264 | 31.597 | 51.853 | 220.940 | 190.662 | 71.741 | 194.732 | 161.486 | 165.235 |
| Dimethyl trisulfide | 6.770 | 4.940 | 6.028 | 6.908 | 4.215 | 4.610 | 3.173 | 3.149 | 3.142 | 2.821 | 4.976 | 4.113 |
| Ethanol | 151.298 | 93.358 | 832.351 | 9338.630 | 215.578 | 447.855 | 2872.446 | 3860.979 | 197.693 | 195.918 | 299.256 | 355.888 |
| Ethyl acetate | 9.644 | 21.172 | 39.776 | 104.242 | 7.093 | 27.285 | 1647.988 | 1314.015 | 5.315 | 13.324 | 19.226 | 33.738 |
| Ethyl benzene | 1.526 | 6.220 | 6.058 | 7.743 | 1.789 | 2.004 | 22.521 | 20.069 | 1.419 | 3.983 | 6.385 | 6.639 |
| Hydrogen sulfide | 1.727 | 62.447 | 184.483 | 1534.412 | 1.814 | 19.746 | 16.155 | 52.089 | 0.912 | 1.598 | 2.954 | 1.796 |
| Isobutyl alcohol | 2.602 | 4.362 | 2.834 | 6.278 | 2.044 | 1.234 | 31.773 | 38.986 | 4.966 | 1.622 | 2.576 | 3.933 |
| Methyl mercaptan | 4.106 | 4.434 | 9.877 | 48.160 | 4.611 | 5.119 | 14.801 | 23.259 | 1.837 | 4.785 | 7.498 | 5.042 |
| Piperidine | 2.029 | 1.080 | 1.242 | 1.027 | 1.619 | 1.336 | 1.402 | 1.356 | 0.947 | 0.558 | 1.403 | 1.620 |
| Propyl benzene | 12.247 | 11.646 | 10.328 | 7.257 | 7.487 | 6.850 | 15.055 | 13.880 | 8.339 | 9.062 | 14.249 | 13.378 |
| Styrene | 3.270 | 10.576 | 9.716 | 7.613 | 4.408 | 4.149 | 4.890 | 4.572 | 3.657 | 7.094 | 9.928 | 11.657 |
| Trimethyl amine | 7.008 | 6.369 | 36.254 | 20.941 | 5.396 | 5.407 | 9.588 | 17.224 | 6.174 | 9.666 | 15.289 | 17.014 |

Table B.11: (Continued) Measured concentrations (ppb) of VOCs detected with SIFTMS on each day in the headspace of salmon samples used for ranking tests.

| VOC | AN4: $0 / 0 / 100\left(\underset{\text { Session } 4}{\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right)} 4^{\circ} \mathrm{C}(\right.$ Session 4$)$ |  |  |  | ANH4: $60 / 0 / 40\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}($ Session 4$)$Session 4 |  |  |  | $\begin{gathered} \text { A4: air } 4^{\circ} \mathrm{C}(\text { Session } 4) \\ \text { Session } 4 \end{gathered}$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | 3 | 5 | 7 | 1 | 3 | 5 | 7 | 1 | 3 | 5 | 7 |
| 2,3-butanediol | 3.007 | 2.688 | 9.618 | 14.316 | 4.117 | 2.277 | 5.981 | 4.979 | 1.710 | 1.908 | 4.655 | 26.135 |
| 2,3-butanedione | 35.879 | 27.275 | 10.909 | 29.885 | 28.502 | 17.261 | 20.662 | 7.636 | 37.736 | 28.532 | 34.659 | 45.595 |
| 3-methyl-1-butanol | 66.046 | 76.073 | 81.874 | 63.681 | 92.043 | 71.846 | 39.460 | 41.512 | 109.606 | 51.334 | 87.211 | 128.824 |
| 3 -methylbutanal | 17.891 | 17.144 | 16.018 | 24.535 | 11.549 | 12.262 | 14.412 | 11.365 | 18.296 | 16.836 | 23.837 | 66.039 |
| 3 -methylbutanoic acid | 2.052 | 1.973 | 1.229 | 1.410 | 2.506 | 0.898 | 4.843 | 3.811 | 2.078 | 1.982 | 1.011 | 1.960 |
| Acetic acid | 11.012 | 17.028 | 24.239 | 31.292 | 12.639 | 8.002 | 29.198 | 28.933 | 9.255 | 9.924 | 14.686 | 29.086 |
| Acetoin | 5.467 | 11.320 | 28.608 | 13.688 | 4.842 | 3.393 | 14.738 | 37.209 | 3.917 | 7.464 | 33.516 | 326.493 |
| Acetone | 32.525 | 56.043 | 46.018 | 59.195 | 34.157 | 42.453 | 45.978 | 82.737 | 28.911 | 40.988 | 33.623 | 51.685 |
| Ammonia | 24.465 | 33.025 | 36.687 | 28.779 | 48.241 | 29.983 | 60.689 | 77.747 | 25.256 | 31.342 | 35.231 | 27.525 |
| Butanone | 5.353 | 4.884 | 12.112 | 4.977 | 5.567 | 5.685 | 11.167 | 8.612 | 6.664 | 4.365 | 2.916 | 3.916 |
| Carbon disulfide | 40.882 | 31.103 | 26.817 | 20.093 | 156.954 | 94.252 | 119.545 | 74.733 | 30.266 | 35.729 | 20.932 | 55.162 |
| Dimethyl amine | 4.684 | 6.333 | 10.726 | 7.550 | 6.317 | 2.305 | 8.419 | 13.828 | 3.365 | 6.081 | 5.765 | 12.144 |
| Dimethyl disulfide | 0.531 | 0.501 | 0.328 | 0.843 | 0.445 | 0.258 | 0.603 | 0.909 | 0.508 | 0.469 | 0.297 | 0.287 |
| Dimethyl sulfide | 31.899 | 39.112 | 136.683 | 126.575 | 188.443 | 72.432 | 186.614 | 191.872 | 27.670 | 46.811 | 49.745 | 170.669 |
| Dimethyl trisulfide | 6.793 | 4.628 | 3.507 | 4.117 | 5.580 | 4.085 | 6.959 | 2.320 | 6.030 | 4.211 | 5.952 | 4.163 |
| Ethanol | 231.608 | 318.473 | 5556.635 | 11585.522 | 124.727 | 85.454 | 136.217 | 340.900 | 258.835 | 205.423 | 540.527 | 2521.270 |
| Ethyl acetate | 11.009 | 19.017 | 53.906 | 25.687 | 8.182 | 6.004 | 20.460 | 54.960 | 9.033 | 14.612 | 62.953 | 581.266 |
| Ethyl benzene | 1.012 | 3.317 | 3.202 | 2.352 | 2.126 | 2.263 | 4.935 | 7.564 | 3.071 | 1.319 | 3.059 | 7.327 |
| Hydrogen sulfide | 34.306 | 71.645 | 503.032 | 1850.067 | 1.453 | 2.620 | 80.402 | 21.862 | 0.898 | 2.575 | 9.303 | 4.176 |
| Isobutyl alcohol | 1.828 | 3.363 | 3.946 | 6.069 | 3.021 | 7.305 | 3.266 | 3.223 | 2.468 | 1.300 | 2.018 | 7.495 |
| Methyl mercaptan | 3.793 | 4.411 | 33.718 | 65.865 | 4.462 | 1.022 | 4.170 | 6.840 | 4.430 | 2.907 | 4.667 | 15.123 |
| Piperidine | 2.314 | 1.791 | 1.611 | 2.412 | 2.264 | 1.181 | 1.506 | 1.027 | 2.968 | 1.848 | 2.566 | 1.916 |
| Propyl benzene | 7.357 | 9.530 | 8.631 | 8.009 | 12.758 | 11.499 | 9.508 | 14.249 | 8.792 | 7.168 | 7.132 | 7.526 |
| Styrene | 4.042 | 5.491 | 5.470 | 3.619 | 5.280 | 6.462 | 6.087 | 7.649 | 7.034 | 3.177 | 3.681 | 3.341 |
| Trimethyl amine | 5.913 | 7.256 | 13.791 | 5.499 | 7.655 | 3.940 | 5.606 | 6.487 | 7.503 | 6.554 | 5.586 | 5.637 |

Table B.11: (Continued) Measured concentrations (ppb) of VOCs detected with SIFTMS on each day in the headspace of salmon samples used for ranking tests.

| VOC | L4: $60 / 5 / 35\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ |  |  |  | M4: $60 / 21 / 19\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ |  |  |  | H4: 60/40/0 ( $\left.\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ (Session 1) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | 5 | 7 | 11 | 1 | 5 | 9 | 11 | 1 | 5 | 9 | 11 |
| 2,3-butanediol | 3.032 | 4.218 | 15.511 | 72.092 | 4.924 | 2.867 | 13.451 | 28.483 | 2.349 | 2.056 | 4.388 | 11.100 |
| 2,3-butanedione | 32.166 | 15.886 | 19.760 | 27.022 | 9.192 | 8.373 | 54.598 | 64.128 | 11.744 | 51.312 | 74.286 | 240.448 |
| 3-methyl-1-butanol | 50.208 | 101.395 | 162.593 | 251.699 | 15.095 | 22.373 | 212.388 | 255.106 | 27.036 | 34.698 | 55.808 | 96.237 |
| 3-methylbutanal | 12.812 | 15.632 | 39.587 | 29.162 | 6.122 | 6.389 | 79.809 | 82.683 | 7.651 | 10.004 | 90.261 | 399.463 |
| 3 -methylbutanoic acid | 1.846 | 1.124 | 2.246 | 3.136 | 3.737 | 4.645 | 3.900 | 6.402 | 2.000 | 3.187 | 2.301 | 3.254 |
| Acetic acid | 16.280 | 10.056 | 34.374 | 172.506 | 30.053 | 20.648 | 41.402 | 61.015 | 14.329 | 15.393 | 21.313 | 33.882 |
| Acetoin | 2.537 | 24.183 | 288.919 | 74.388 | 38.001 | 10.407 | 734.099 | 1047.554 | 4.778 | 5.521 | 75.744 | 391.280 |
| Acetone | 36.012 | 43.144 | 85.297 | 60.966 | 102.434 | 41.400 | 97.428 | 125.389 | 44.442 | 47.431 | 52.299 | 101.242 |
| Ammonia | 30.584 | 46.795 | 36.841 | 132.367 | 45.797 | 48.774 | 34.183 | 77.119 | 54.757 | 31.729 | 29.389 | 81.548 |
| Butanone | 7.377 | 5.575 | 9.441 | 10.062 | 11.202 | 9.508 | 13.359 | 40.689 | 13.029 | 10.992 | 14.282 | 27.328 |
| Carbon disulfide | 105.587 | 50.554 | 101.074 | 97.181 | 76.967 | 62.931 | 112.423 | 96.591 | 65.889 | 77.726 | 103.205 | 306.991 |
| Dimethyl amine | 4.920 | 5.866 | 14.982 | 12.116 | 10.284 | 8.640 | 23.641 | 31.269 | 8.473 | 7.608 | 9.285 | 11.905 |
| Dimethyl disulfide | 0.365 | 0.269 | 1.099 | 2.301 | 0.465 | 0.830 | 1.015 | 1.127 | 0.296 | 0.381 | 0.429 | 0.401 |
| Dimethyl sulfide | 247.117 | 272.234 | 625.191 | 531.339 | 336.511 | 266.040 | 526.795 | 485.127 | 305.239 | 261.150 | 318.674 | 360.202 |
| Dimethyl trisulfide | 8.886 | 5.471 | 3.475 | 3.539 | 3.198 | 1.744 | 2.114 | 1.936 | 4.182 | 3.533 | 3.160 | 1.388 |
| Ethanol | 306.789 | 426.557 | 7463.026 | 14930.766 | 211.003 | 170.704 | 5843.617 | 6537.605 | 542.226 | 487.583 | 727.628 | 2099.779 |
| Ethyl acetate | 9.471 | 43.834 | 447.630 | 188.034 | 41.767 | 19.183 | 898.853 | 1371.872 | 6.248 | 9.595 | 119.207 | 593.967 |
| Ethyl benzene | 2.837 | 5.753 | 11.867 | 37.247 | 5.809 | 4.540 | 17.545 | 23.660 | 1.851 | 1.471 | 2.872 | 8.346 |
| Hydrogen sulfide | 1.706 | 7.319 | 329.006 | 5771.282 | 2.451 | 4.207 | 573.261 | 233.193 | 1.173 | 5.963 | 12.946 | 25.477 |
| Isobutyl alcohol | 8.115 | 6.276 | 14.592 | 29.057 | 6.015 | 2.020 | 21.140 | 24.362 | 2.374 | 2.167 | 2.983 | 7.191 |
| Methyl mercaptan | 6.273 | 5.857 | 39.553 | 72.967 | 7.352 | 6.244 | 32.342 | 37.311 | 5.315 | 5.167 | 8.134 | 14.327 |
| Piperidine | 2.251 | 1.010 | 0.749 | 1.589 | 0.629 | 1.054 | 0.651 | 0.759 | 0.778 | 1.117 | 1.072 | 2.181 |
| Propyl benzene | 8.957 | 2.329 | 8.514 | 13.198 | 10.373 | 7.524 | 10.358 | 10.993 | 15.293 | 13.838 | 11.300 | 13.942 |
| Styrene | 6.419 | 10.586 | 12.446 | 11.282 | 7.494 | 6.860 | 7.881 | 8.766 | 4.215 | 4.809 | 5.795 | 5.260 |
| Trimethyl amine | 6.978 | 7.090 | 12.515 | 10.757 | 10.299 | 8.066 | 8.018 | 15.517 | 13.494 | 16.004 | 17.423 | 16.367 |

Table B.12: Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of salmon samples used for labelling tests.

| VOC | AN4: 0/0/100 ( $\left.\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ (Session 1) |  |  |  | ANH4: 60/0/40 ( $\left.\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ (Session 1) |  |  |  | A4: air $4^{\circ} \mathrm{C}$ (Session 1) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | 5 | 9 | 13 | 1 | 5 | 9 | 11 | 1 | 5 | 9 | 11 |
| 2,3-butanediol | 2.006 | 5.699 | 27.244 | 12.676 | 2.466 | 3.089 | 6.611 | 6.805 | 1.388 | 3.913 | 57.365 | 75.847 |
| 2,3-butanedione | 6.802 | 19.634 | 6.057 | 17.549 | 40.531 | 26.619 | 20.728 | 31.615 | 12.892 | 13.128 | 31.467 | 37.179 |
| 3-methyl-1-butanol | 20.145 | 133.477 | 34.452 | 119.522 | 53.347 | 65.580 | 62.792 | 55.501 | 63.914 | 124.718 | 337.350 | 704.592 |
| 3 -methylbutanal | 4.911 | 12.573 | 11.390 | 12.579 | 10.342 | 11.841 | 20.701 | 29.850 | 10.963 | 11.195 | 51.694 | 80.381 |
| 3 -methylbutanoic acid | 2.446 | 0.923 | 1.227 | 1.427 | 2.108 | 1.949 | 1.747 | 2.947 | 1.682 | 1.042 | 2.213 | 2.573 |
| Acetic acid | 12.550 | 20.437 | 75.567 | 49.374 | 15.172 | 13.978 | 43.404 | 37.868 | 12.072 | 12.130 | 73.520 | 77.951 |
| Acetoin | 5.804 | 19.144 | 16.692 | 37.417 | 4.276 | 5.498 | 20.667 | 23.416 | 3.883 | 44.063 | 476.361 | 464.441 |
| Acetone | 23.944 | 44.589 | 30.267 | 65.153 | 41.525 | 49.630 | 43.870 | 69.688 | 30.089 | 25.941 | 156.868 | 109.053 |
| Ammonia | 37.965 | 16.352 | 19.115 | 24.988 | 53.202 | 29.014 | 30.767 | 82.614 | 36.398 | 16.921 | 20.265 | 49.183 |
| Butanone | 14.664 | 9.432 | 15.868 | 205.463 | 23.834 | 9.527 | 12.669 | 59.992 | 9.066 | 6.663 | 10.179 | 16.857 |
| Carbon disulfide | 180.327 | 53.936 | 45.324 | 39.662 | 89.498 | 44.663 | 81.386 | 74.015 | 113.118 | 63.370 | 79.004 | 108.411 |
| Dimethyl amine | 5.289 | 8.665 | 10.868 | 12.309 | 8.698 | 5.708 | 19.016 | 21.165 | 4.473 | 4.669 | 10.408 | 18.232 |
| Dimethyl disulfide | 0.395 | 0.283 | 0.546 | 2.099 | 0.436 | 0.678 | 0.953 | 1.631 | 0.493 | 0.214 | 1.215 | 0.709 |
| Dimethyl sulfide | 23.285 | 86.216 | 121.937 | 174.570 | 313.277 | 250.217 | 536.621 | 593.665 | 21.345 | 43.599 | 193.235 | 283.752 |
| Dimethyl trisulfide | 2.465 | 3.707 | 1.096 | 3.686 | 5.785 | 4.565 | 3.617 | 6.023 | 4.432 | 3.057 | 3.694 | 5.055 |
| Ethanol | 447.789 | 4396.259 | 13016.948 | 28474.965 | 2241.201 | 490.517 | 12004.680 | 14238.473 | 259.913 | 695.140 | 4094.356 | 5519.912 |
| Ethyl acetate | 9.573 | 31.865 | 35.162 | 61.177 | 7.484 | 9.343 | 36.192 | 38.753 | 10.340 | 85.439 | 805.008 | 834.677 |
| Ethyl benzene | 1.716 | 1.875 | 2.499 | 5.578 | 2.633 | 2.190 | 3.144 | 2.122 | 2.281 | 2.667 | 15.682 | 22.128 |
| Hydrogen sulfide | 100.909 | 261.057 | 1504.663 | 16638.469 | 17.862 | 68.528 | 867.874 | 1631.121 | 0.725 | 8.217 | 31.842 | 68.885 |
| Isobutyl alcohol | 1.405 | 3.791 | 5.769 | 26.530 | 3.625 | 3.391 | 6.730 | 8.465 | 1.491 | 3.084 | 30.031 | 43.663 |
| Methyl mercaptan | 6.849 | 23.223 | 78.982 | 185.738 | 13.299 | 4.812 | 54.423 | 68.909 | 3.097 | 7.577 | 22.997 | 33.497 |
| Piperidine | 0.866 | 1.774 | 0.905 | 1.825 | 2.147 | 1.957 | 0.961 | 1.507 | 1.292 | 1.173 | 1.633 | 2.238 |
| Propyl benzene | 8.536 | 2.314 | 3.366 | 8.947 | 14.794 | 17.435 | 12.138 | 13.759 | 8.506 | 2.795 | 6.288 | 14.812 |
| Styrene | 3.833 | 6.792 | 2.616 | 8.864 | 5.145 | 4.659 | 5.681 | 6.489 | 6.479 | 4.951 | 7.505 | 14.087 |
| Trimethyl amine | 10.950 | 8.159 | 17.418 | 26.511 | 51.411 | 10.994 | 16.279 | 29.133 | 7.388 | 8.061 | 23.409 | 21.399 |

Table B.12: (Continued) Measured concentrations (ppb) of VOCs detected with SIFTMS on each day in the headspace of salmon samples used for labelling tests.

|  | H4: $60 / 40 / 0\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ (Session 2) |  |  |  | AN4: 0/0/100 ( $\left.\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ (Session 2) |  |  |  | ANH4: 60/0/40 ( $\left.\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ (Session 2) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| VOC | 1 | 3 | 5 | 7 | 3 | 7 | 9 | 11 | 1 | 3 | 5 | 7 |
| 2,3-butanediol | 2.231 | 1.965 | 2.440 | 2.353 | 2.262 | 19.862 | 24.890 | 29.164 | 2.466 | 1.965 | 2.564 | 4.616 |
| 2,3-butanedione | 24.581 | 25.530 | 37.115 | 21.259 | 22.469 | 17.991 | 17.920 | 14.975 | 40.531 | 25.530 | 25.742 | 29.875 |
| 3-methyl-1-butanol | 42.391 | 54.029 | 74.707 | 34.577 | 139.112 | 107.896 | 95.595 | 190.173 | 53.347 | 54.029 | 74.780 | 65.047 |
| 3-methylbutanal | 8.710 | 12.733 | 17.325 | 19.431 | 16.950 | 15.740 | 15.571 | 18.218 | 10.342 | 12.733 | 10.812 | 14.770 |
| 3-methylbutanoic acid | 1.925 | 1.223 | 2.196 | 4.725 | 1.995 | 1.368 | 1.095 | 1.632 | 2.108 | 1.223 | 2.059 | 3.213 |
| Acetic acid | 14.812 | 14.429 | 12.155 | 15.485 | 15.038 | 46.124 | 67.639 | 73.276 | 15.172 | 14.429 | 18.303 | 23.756 |
| Acetoin | 5.277 | 4.331 | 8.778 | 11.722 | 4.071 | 18.737 | 12.410 | 29.241 | 4.276 | 4.331 | 9.974 | 40.430 |
| Acetone | 38.732 | 54.721 | 45.282 | 55.877 | 32.570 | 37.637 | 45.976 | 59.132 | 41.525 | 54.721 | 45.793 | 46.748 |
| Ammonia | 56.869 | 34.378 | 30.183 | 27.295 | 24.972 | 18.721 | 18.992 | 48.909 | 53.202 | 34.378 | 29.307 | 28.132 |
| Butanone | 14.896 | 19.364 | 14.317 | 11.109 | 7.453 | 10.949 | 7.677 | 10.125 | 23.834 | 19.364 | 10.774 | 10.134 |
| Carbon disulfide | 120.623 | 57.111 | 64.532 | 55.329 | 51.235 | 30.390 | 63.161 | 43.200 | 89.498 | 57.111 | 40.240 | 53.742 |
| Dimethyl amine | 9.490 | 5.097 | 6.916 | 6.080 | 4.338 | 9.219 | 7.348 | 13.170 | 8.698 | 5.097 | 6.429 | 12.861 |
| Dimethyl disulfide | 0.679 | 0.398 | 0.356 | 0.632 | 0.394 | 0.638 | 0.825 | 1.903 | 0.436 | 0.398 | 0.443 | 0.269 |
| Dimethyl sulfide | 300.423 | 216.367 | 265.162 | 320.113 | 31.390 | 134.334 | 98.812 | 180.333 | 313.277 | 216.367 | 251.993 | 386.462 |
| Dimethyl trisulfide | 3.172 | 4.357 | 3.984 | 3.209 | 6.802 | 4.062 | 3.478 | 4.231 | 5.785 | 4.357 | 5.195 | 6.080 |
| Ethanol | 858.689 | 1266.100 | 722.966 | 419.298 | 178.358 | 8758.607 | 11289.767 | 25970.632 | 2241.201 | 1266.100 | 597.803 | 2187.953 |
| Ethyl acetate | 9.932 | 9.162 | 16.916 | 19.380 | 7.710 | 35.391 | 26.875 | 64.010 | 7.484 | 9.162 | 16.984 | 65.562 |
| Ethyl benzene | 2.394 | 2.543 | 2.341 | 1.397 | 2.949 | 2.769 | 2.881 | 5.458 | 2.633 | 2.543 | 3.317 | 3.043 |
| Hydrogen sulfide | 1.254 | 16.910 | 1.091 | 1.247 | 96.694 | 851.306 | 1164.356 | 5991.871 | 17.862 | 16.910 | 78.828 | 147.370 |
| Isobutyl alcohol | 2.368 | 3.834 | 3.285 | 2.065 | 2.373 | 4.206 | 7.103 | 19.849 | 3.625 | 3.834 | 4.082 | 3.314 |
| Methyl mercaptan | 6.747 | 7.484 | 6.320 | 7.148 | 3.496 | 51.545 | 62.427 | 183.439 | 13.299 | 7.484 | 4.758 | 12.937 |
| Piperidine | 1.675 | 1.928 | 1.884 | 1.183 | 3.101 | 1.729 | 2.357 | 2.280 | 2.147 | 1.928 | 2.001 | 2.108 |
| Propyl benzene | 13.258 | 8.137 | 13.816 | 11.733 | 5.882 | 4.613 | 4.181 | 12.578 | 14.794 | 8.137 | 19.265 | 21.027 |
| Styrene | 5.294 | 4.203 | 5.924 | 4.765 | 5.887 | 3.330 | 3.523 | 5.893 | 5.145 | 4.203 | 7.237 | 3.605 |
| Trimethyl amine | 20.057 | 29.671 | 18.890 | 12.950 | 7.331 | 12.241 | 11.761 | 17.898 | 51.411 | 29.671 | 13.506 | 11.469 |

Table B.12: (Continued) Measured concentrations (ppb) of VOCs detected with SIFTMS on each day in the headspace of salmon samples used for labelling tests.

| VOC | A4: air $4^{\circ} \mathrm{C}$ (Session 2) |  |  |  | H4: $60 / 40 / 0\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ (Session 3) |  |  |  | AN4: 0/0/100 ( $\left.\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ (Session 3) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | 3 | 5 | 7 | 1 | 5 | 9 | 11 | 1 | 5 | 9 | 11 |
| 2,3-butanediol | 2.529 | 1.865 | 5.898 | 64.161 | 4.676 | 6.531 | 4.979 | 8.970 | 1.494 | 7.047 | 16.457 | 28.672 |
| 2,3-butanedione | 37.251 | 33.004 | 24.955 | 47.515 | 16.387 | 20.309 | 33.937 | 22.454 | 42.495 | 23.943 | 28.999 | 25.354 |
| 3-methyl-1-butanol | 167.384 | 117.628 | 153.845 | 339.983 | 77.904 | 102.297 | 102.537 | 99.365 | 64.545 | 68.551 | 124.953 | 108.024 |
| 3 -methylbutanal | 14.496 | 15.985 | 15.931 | 47.596 | 9.281 | 13.166 | 29.757 | 22.012 | 16.472 | 17.729 | 23.800 | 24.099 |
| 3 -methylbutanoic acid | 1.119 | 1.538 | 0.984 | 2.626 | 1.869 | 4.909 | 5.322 | 3.180 | 2.077 | 0.925 | 1.773 | 2.391 |
| Acetic acid | 13.028 | 12.636 | 11.615 | 68.409 | 14.692 | 22.361 | 25.382 | 25.360 | 13.612 | 19.679 | 31.590 | 51.964 |
| Acetoin | 4.093 | 4.220 | 105.367 | 652.771 | 5.957 | 9.939 | 38.767 | 26.649 | 5.464 | 16.976 | 26.995 | 61.363 |
| Acetone | 34.771 | 28.826 | 35.750 | 80.939 | 44.298 | 32.176 | 58.258 | 196.040 | 38.235 | 43.321 | 74.390 | 90.353 |
| Ammonia | 35.889 | 25.518 | 17.063 | 20.382 | 46.650 | 62.001 | 62.424 | 55.579 | 22.788 | 35.010 | 34.908 | 98.963 |
| Butanone | 11.876 | 12.651 | 8.169 | 13.227 | 7.863 | 13.276 | 9.891 | 10.805 | 7.050 | 3.135 | 486.917 | 4742.194 |
| Carbon disulfide | 73.138 | 23.929 | 40.866 | 54.277 | 133.092 | 77.284 | 48.489 | 108.450 | 29.538 | 18.185 | 43.625 | 24.843 |
| Dimethyl amine | 5.428 | 4.514 | 9.027 | 12.002 | 7.429 | 8.862 | 10.303 | 10.557 | 5.502 | 12.572 | 19.081 | 31.624 |
| Dimethyl disulfide | 0.261 | 0.472 | 0.311 | 1.053 | 0.375 | 0.744 | 0.566 | 1.199 | 0.332 | 0.593 | 1.913 | 3.539 |
| Dimethyl sulfide | 16.025 | 37.820 | 96.708 | 215.855 | 198.523 | 153.236 | 159.040 | 221.691 | 37.433 | 133.640 | 267.233 | 492.838 |
| Dimethyl trisulfide | 6.611 | 6.904 | 3.239 | 5.019 | 4.595 | 6.082 | 5.571 | 4.747 | 7.499 | 3.747 | 3.085 | 5.350 |
| Ethanol | 302.324 | 263.898 | 1895.423 | 3898.454 | 353.314 | 113.499 | 286.208 | 973.549 | 234.497 | 5019.237 | 32290.024 | 36125.899 |
| Ethyl acetate | 10.572 | 10.209 | 174.918 | 1099.198 | 7.539 | 16.160 | 51.573 | 42.150 | 8.960 | 33.908 | 52.303 | 108.562 |
| Ethyl benzene | 2.425 | 2.405 | 1.962 | 15.525 | 2.409 | 4.395 | 8.995 | 5.231 | 1.281 | 2.908 | 5.551 | 5.888 |
| Hydrogen sulfide | 1.205 | 1.213 | 14.425 | 3.544 | 2.171 | 2.574 | 3.611 | 4.504 | 42.435 | 674.332 | 10695.747 | 36990.031 |
| Isobutyl alcohol | 4.564 | 1.284 | 2.824 | 21.638 | 2.088 | 2.606 | 3.801 | 4.689 | 1.604 | 3.309 | 33.886 | 200.709 |
| Methyl mercaptan | 5.251 | 5.233 | 14.055 | 21.503 | 6.378 | 4.268 | 5.902 | 11.792 | 5.253 | 26.785 | 206.287 | 3184.060 |
| Piperidine | 1.833 | 3.116 | 1.993 | 2.159 | 1.578 | 2.528 | 2.230 | 1.847 | 3.067 | 1.845 | 2.400 | 2.526 |
| Propyl benzene | 6.781 | 6.614 | 3.028 | 9.579 | 11.651 | 13.685 | 11.340 | 7.754 | 10.701 | 8.534 | 10.164 | 11.878 |
| Styrene | 11.657 | 5.752 | 6.628 | 7.476 | 3.708 | 5.307 | 10.315 | 8.504 | 2.842 | 5.825 | 7.219 | 28.027 |
| Trimethyl amine | 12.395 | 8.729 | 10.274 | 16.815 | 12.296 | 8.115 | 18.351 | 27.431 | 6.723 | 6.373 | 26.615 | 49.999 |

Table B.12: (Continued) Measured concentrations (ppb) of VOCs detected with SIFTMS on each day in the headspace of salmon samples used for labelling tests.

|  | ANH4: $60 / 0 / 40\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ (Session 3) |  |  |  | A4: air $4^{\circ} \mathrm{C}$ (Session 3) |  |  |  | H4: $60 / 40 / 0\left(\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}($ Session 4) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| VOC | 1 | 5 | 9 | 11 | 1 | 5 | 9 | 11 | 3 | 5 | 7 | 9 |
| 2,3-butanediol | 4.117 | 4.923 | 9.890 | 14.866 | 1.710 | 4.503 | 51.237 | 48.470 | 2.156 | 5.849 | 8.097 | 7.326 |
| 2,3-butanedione | 28.502 | 25.341 | 25.954 | 20.754 | 37.736 | 21.831 | 65.672 | 48.491 | 11.991 | 17.209 | 13.997 | 21.639 |
| 3-methyl-1-butanol | 92.043 | 45.339 | 46.802 | 149.046 | 109.606 | 65.457 | 458.589 | 414.936 | 50.071 | 96.793 | 124.632 | 103.921 |
| 3-methylbutanal | 11.549 | 13.007 | 16.367 | 23.337 | 18.296 | 18.984 | 80.696 | 78.540 | 10.433 | 17.365 | 9.816 | 16.052 |
| 3-methylbutanoic acid | 2.506 | 3.050 | 3.644 | 3.451 | 2.078 | 0.744 | 2.473 | 1.800 | 0.764 | 3.719 | 4.983 | 2.960 |
| Acetic acid | 12.639 | 21.720 | 38.517 | 49.848 | 9.255 | 13.089 | 44.695 | 32.208 | 6.040 | 18.995 | 25.063 | 33.602 |
| Acetoin | 4.842 | 9.445 | 37.779 | 34.505 | 3.917 | 81.332 | 834.416 | 1273.988 | 3.639 | 9.550 | 16.960 | 19.732 |
| Acetone | 34.157 | 44.001 | 52.195 | 163.072 | 28.911 | 35.210 | 128.022 | 190.806 | 48.414 | 56.025 | 64.482 | 53.218 |
| Ammonia | 48.241 | 61.324 | 62.745 | 53.607 | 25.256 | 35.719 | 32.632 | 101.524 | 29.631 | 54.489 | 76.098 | 60.961 |
| Butanone | 5.567 | 11.331 | 9.603 | 7.351 | 6.664 | 3.862 | 8.160 | 16.454 | 7.730 | 13.667 | 14.176 | 12.684 |
| Carbon disulfide | 156.954 | 135.886 | 54.784 | 57.285 | 30.266 | 23.145 | 100.475 | 106.847 | 122.033 | 57.460 | 132.567 | 110.498 |
| Dimethyl amine | 6.317 | 8.464 | 11.350 | 15.845 | 3.365 | 5.326 | 15.818 | 14.644 | 2.695 | 8.428 | 10.057 | 10.653 |
| Dimethyl disulfide | 0.445 | 0.943 | 0.913 | 0.831 | 0.508 | 0.275 | 0.842 | 1.088 | 0.110 | 0.656 | 0.573 | 1.404 |
| Dimethyl sulfide | 188.443 | 191.263 | 223.032 | 304.714 | 27.670 | 72.450 | 263.711 | 203.700 | 76.705 | 196.044 | 126.170 | 173.016 |
| Dimethyl trisulfide | 5.580 | 7.740 | 5.740 | 6.016 | 6.030 | 4.364 | 4.711 | 5.372 | 3.623 | 3.726 | 5.023 | 4.598 |
| Ethanol | 124.727 | 120.524 | 1215.125 | 5915.157 | 258.835 | 877.430 | 4789.076 | 4159.172 | 196.720 | 202.291 | 239.979 | 262.983 |
| Ethyl acetate | 8.182 | 13.331 | 46.636 | 56.609 | 9.033 | 148.152 | 1446.359 | 2109.152 | 6.959 | 18.813 | 26.240 | 34.346 |
| Ethyl benzene | 2.126 | 4.210 | 5.158 | 7.016 | 3.071 | 2.532 | 26.130 | 22.958 | 1.371 | 7.370 | 7.038 | 6.747 |
| Hydrogen sulfide | 1.453 | 48.343 | 196.163 | 491.380 | 0.898 | 6.983 | 19.438 | 27.986 | 1.467 | 3.736 | 2.817 | 6.428 |
| Isobutyl alcohol | 3.021 | 1.648 | 2.127 | 5.913 | 2.468 | 1.731 | 33.644 | 53.823 | 3.530 | 2.873 | 1.936 | 2.938 |
| Methyl mercaptan | 4.462 | 5.575 | 10.505 | 29.063 | 4.430 | 7.308 | 25.178 | 20.089 | 2.174 | 4.265 | 6.833 | 5.078 |
| Piperidine | 2.264 | 1.097 | 1.625 | 1.973 | 2.968 | 1.344 | 1.535 | 2.028 | 0.934 | 1.225 | 1.311 | 1.669 |
| Propyl benzene | 12.758 | 9.633 | 10.798 | 6.809 | 8.792 | 6.852 | 13.159 | 15.468 | 9.401 | 10.651 | 16.957 | 11.722 |
| Styrene | 5.280 | 7.100 | 6.271 | 12.172 | 7.034 | 4.028 | 7.556 | 3.998 | 1.684 | 9.156 | 9.440 | 6.666 |
| Trimethyl amine | 7.655 | 5.955 | 26.530 | 21.921 | 7.503 | 5.175 | 11.301 | 16.075 | 5.687 | 7.810 | 11.868 | 33.926 |

Table B.12: (Continued) Measured concentrations (ppb) of VOCs detected with SIFTMS on each day in the headspace of salmon samples used for labelling tests.

| VOC | AN4: 0/0/100 ( $\left.\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ (Session 4) |  |  |  | ANH4: 60/0/40 ( $\left.\mathrm{CO}_{2} / \mathrm{O}_{2} / \mathrm{N}_{2}\right) 4^{\circ} \mathrm{C}$ (Session 4) |  |  |  | A4: air $4^{\circ} \mathrm{C}$ (Session 4) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | - | 3 | 5 | 7 | , | 3 | 5 | 7 | 1 | 3 | 5 | 7 |
| 2,3-butanediol | 1.821 | 2.490 | 4.385 | 7.176 | 4.988 | 2.117 | 7.128 | 6.410 | 2.503 | 1.817 | 11.496 | 21.845 |
| 2,3-butanedione | 36.050 | 26.751 | 24.263 | 21.539 | 19.623 | 15.300 | 18.231 | 19.651 | 45.851 | 33.604 | 35.659 | 43.893 |
| 3-methyl-1-butanol | 61.808 | 47.892 | 38.248 | 125.459 | 51.220 | 43.511 | 96.125 | 92.201 | 103.975 | 71.154 | 69.454 | 89.894 |
| 3 -methylbutanal | 15.690 | 18.194 | 16.935 | 20.859 | 9.125 | 9.743 | 15.997 | 15.661 | 16.104 | 19.474 | 18.723 | 49.503 |
| 3 -methylbutanoic acid | 2.518 | 1.696 | 1.230 | 1.881 | 2.429 | 0.935 | 3.608 | 8.208 | 2.152 | 1.523 | 1.492 | 1.775 |
| Acetic acid | 13.416 | 13.744 | 17.204 | 23.192 | 15.338 | 7.503 | 25.425 | 32.609 | 10.909 | 12.859 | 18.577 | 23.402 |
| Acetoin | 3.618 | 9.112 | 14.589 | 16.784 | 6.203 | 2.719 | 9.927 | 15.744 | 5.600 | 4.763 | 15.597 | 386.188 |
| Acetone | 32.282 | 58.056 | 38.507 | 47.112 | 36.596 | 33.115 | 45.686 | 63.738 | 32.708 | 44.577 | 37.921 | 58.628 |
| Ammonia | 21.811 | 28.595 | 36.061 | 29.928 | 51.450 | 28.812 | 60.591 | 75.974 | 23.822 | 31.186 | 35.939 | 28.415 |
| Butanone | 8.543 | 5.370 | 3.203 | 10.355 | 5.969 | 5.605 | 10.738 | 11.240 | 8.880 | 3.245 | 3.669 | 3.653 |
| Carbon disulfide | 60.241 | 29.916 | 35.125 | 30.580 | 156.419 | 92.197 | 90.912 | 82.621 | 64.133 | 31.964 | 50.993 | 45.824 |
| Dimethyl amine | 6.487 | 7.070 | 6.578 | 11.050 | 5.965 | 2.426 | 7.344 | 10.530 | 3.822 | 5.932 | 4.780 | 9.166 |
| Dimethyl disulfide | 0.345 | 0.026 | 0.150 | 0.417 | 0.199 | 0.065 | 0.392 | 0.973 | 0.500 | 0.287 | 0.224 | 0.350 |
| Dimethyl sulfide | 30.199 | 38.574 | 68.881 | 181.415 | 197.105 | 80.217 | 197.212 | 191.905 | 36.343 | 35.804 | 51.853 | 117.666 |
| Dimethyl trisulfide | 4.728 | 4.753 | 3.072 | 4.794 | 3.464 | 3.071 | 4.940 | 5.553 | 7.549 | 4.519 | 4.610 | 5.057 |
| Ethanol | 347.009 | 148.724 | 1784.364 | 14581.110 | 198.064 | 124.955 | 93.358 | 152.952 | 279.218 | 172.534 | 447.855 | 1772.026 |
| Ethyl acetate | 10.476 | 17.353 | 27.872 | 31.811 | 9.237 | 5.227 | 21.172 | 32.833 | 11.172 | 12.181 | 27.285 | 697.756 |
| Ethyl benzene | 1.833 | 1.406 | 1.843 | 4.792 | 2.353 | 2.062 | 6.220 | 5.862 | 3.774 | 1.423 | 2.004 | 5.660 |
| Hydrogen sulfide | 11.624 | 20.002 | 246.254 | 1628.826 | 6.239 | 0.890 | 62.447 | 11.798 | 0.972 | 4.518 | 19.746 | 2.308 |
| Isobutyl alcohol | 1.648 | 2.168 | 1.559 | 8.468 | 2.569 | 4.334 | 4.362 | 4.584 | 1.846 | 1.031 | 1.234 | 2.929 |
| Methyl mercaptan | 4.903 | 4.365 | 11.027 | 98.861 | 5.083 | 1.440 | 4.434 | 5.825 | 5.300 | 4.404 | 5.119 | 11.116 |
| Piperidine | 1.882 | 2.089 | 1.622 | 1.553 | 1.490 | 1.174 | 1.080 | 2.013 | 3.806 | 2.294 | 1.336 | 1.804 |
| Propyl benzene | 9.052 | 8.847 | 8.696 | 9.036 | 12.341 | 10.448 | 11.646 | 15.156 | 8.328 | 8.526 | 6.850 | 8.547 |
| Styrene | 4.309 | 4.236 | 3.023 | 6.104 | 3.998 | 3.594 | 10.576 | 11.971 | 5.625 | 4.343 | 4.149 | 4.639 |
| Trimethyl amine | 9.813 | 3.261 | 3.104 | 6.001 | 8.482 | 6.044 | 6.369 | 6.572 | 7.712 | 5.289 | 5.407 | 6.507 |

Table B.13: (Continued) Measured concentrations (ppb) of VOCs detected with SIFTMS on each day in the headspace of salmon samples used for labelling tests.

| VOC | Monday |  |  |  | Tuesday |  |  |  | Wednesday |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{A}_{3}$ | $\mathrm{B}_{4}$ | $\mathrm{C}_{5}$ | $\mathrm{D}_{6}$ | $\mathrm{A}_{1}$ | $\mathrm{B}_{2}$ | $\mathrm{C}_{3}$ | $\mathrm{D}_{4}$ | $\mathrm{A}_{4}$ | $\mathrm{B}_{5}$ | $\mathrm{C}_{6}$ | $\mathrm{D}_{7}$ |
| 2,3-butanedione | 6.29 | 29.41 | 38.92 | 28.11 | 4.39 | 29.93 | 13.99 | 32.45 | 7.05 | 36.88 | 23.87 | 27.70 |
| 2,3-butanediol | 2.05 | 1.53 | 8.83 | 2.93 | 1.47 | 1.81 | 1.70 | 2.41 | 2.25 | 4.22 | 4.18 | 10.13 |
| 3-methyl-1-butanol | 45.56 | 40.02 | 92.38 | 25.86 | 43.04 | 163.46 | 34.94 | 66.95 | 59.93 | 52.38 | 25.92 | 125.71 |
| 3-methylbutanal | 10.63 | 15.10 | 22.75 | 16.51 | 7.72 | 19.80 | 16.02 | 19.04 | 13.63 | 19.85 | 23.24 | 28.91 |
| 3 -methylbutanoic acid | 0.53 | 0.42 | 0.69 | 0.42 | 0.35 | 0.55 | 0.45 | 0.44 | 0.29 | 0.50 | 0.43 | 0.45 |
| Acetic acid | 8.11 | 7.09 | 15.37 | 12.86 | 9.18 | 8.08 | 9.16 | 9.69 | 9.69 | 12.46 | 12.97 | 15.04 |
| Acetoin | 2.58 | 2.35 | 12.49 | 8.30 | 1.97 | 3.16 | 4.16 | 5.62 | 7.57 | 8.53 | 19.14 | 7.22 |
| Acetone | 32.74 | 25.64 | 39.43 | 30.26 | 23.19 | 29.13 | 33.19 | 27.90 | 30.58 | 31.39 | 28.71 | 36.37 |
| Ammonia | 7.33 | 8.44 | 9.58 | 10.49 | 8.73 | 9.32 | 12.32 | 10.64 | 7.35 | 10.97 | 10.78 | 7.49 |
| Butanone | 7.22 | 5.90 | 8.51 | 5.99 | 8.90 | 4.99 | 7.84 | 5.81 | 7.87 | 6.63 | 7.74 | 7.33 |
| Carbon disulfide | 138.48 | 124.04 | 181.20 | 84.26 | 69.03 | 141.33 | 79.50 | 78.66 | 104.50 | 133.57 | 49.79 | 144.91 |
| Dimethyl amine | 1.008 | 0.712 | 1.831 | 0.608 | 1.150 | 1.048 | 0.765 | 0.699 | 1.305 | 1.097 | 1.257 | 1.357 |
| Dimethyl disulfide | 0.22 | 0.21 | 1.72 | 0.33 | 0.21 | 0.19 | 0.23 | 0.26 | 0.30 | 0.80 | 0.46 | 3.25 |
| Dimethyl sulfide | 54.97 | 32.86 | 216.86 | 67.96 | 28.77 | 23.35 | 33.17 | 42.12 | 62.19 | 118.01 | 68.96 | 231.30 |
| Dimethyl trisulfide | 0.84 | 3.27 | 3.64 | 3.02 | 0.30 | 3.11 | 2.28 | 3.01 | 0.64 | 3.08 | 2.05 | 2.44 |
| Ethanol | 144.16 | 97.05 | 4912.39 | 456.26 | 107.61 | 62.14 | 111.62 | 220.15 | 357.28 | 1628.48 | 988.91 | 7631.94 |
| Ethyl acetate | 9.05 | 6.43 | 35.42 | 20.35 | 5.93 | 9.32 | 9.25 | 14.26 | 21.21 | 22.25 | 43.68 | 32.68 |
| Ethyl benzene | 3.93 | 2.87 | 5.39 | 3.14 | 3.72 | 3.13 | 3.13 | 2.98 | 4.74 | 3.63 | 3.22 | 4.31 |
| Hydrogen sulfide | 92.00 | 47.27 | 482.84 | 69.18 | 5.04 | 9.80 | 61.55 | 43.87 | 85.87 | 155.98 | 101.04 | 1451.76 |
| Isobutyl alcohol | 21.11 | 22.98 | 40.75 | 21.39 | 15.92 | 14.56 | 19.77 | 28.17 | 22.60 | 24.89 | 22.03 | 39.67 |
| Methyl mercaptan | 1.07 | 0.94 | 35.43 | 3.41 | 0.76 | 0.67 | 0.62 | 1.42 | 1.95 | 9.40 | 5.70 | 59.77 |
| Piperidine | 0.82 | 1.65 | 1.82 | 1.19 | 0.61 | 1.98 | 1.44 | 1.40 | 0.68 | 1.28 | 1.43 | 2.38 |
| Propyl benzene | 0.95 | 0.99 | 1.65 | 1.27 | 0.62 | 0.74 | 0.62 | 0.97 | 0.90 | 0.99 | 1.05 | 1.41 |
| Styrene | 4.37 | 3.23 | 8.30 | 2.46 | 4.18 | 5.01 | 3.18 | 3.70 | 3.44 | 4.28 | 3.17 | 6.80 |
| Trimethyl amine | 2.25 | 3.71 | 10.64 | 7.78 | 2.31 | 2.13 | 3.82 | 3.71 | 5.64 | 13.01 | 9.84 | 8.19 |

Table B.14: Measured concentrations (ppb) of VOCs detected with SIFT-MS on each day in the headspace of salmon samples used in experiment AN4* (scoring and ranking with ties).

|  | Thursday |  |  |  | Friday |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{A}_{2}$ | $\mathrm{B}_{3}$ | $\mathrm{C}_{4}$ | $\mathrm{D}_{5}$ | $\mathrm{A}_{5}$ | $\mathrm{B}_{6}$ | $\mathrm{C}_{7}$ | $\mathrm{D}_{8}$ |
| 2,3-butanedione | 5.19 | 26.83 | 10.66 | 35.00 | 38.23 | 24.53 | 14.93 | 28.17 |
| 2,3-butanediol | 1.28 | 1.49 | 1.28 | 3.69 | 2.96 | 3.10 | 16.45 | 12.38 |
| 3-methyl-1-butanol | 12.18 | 35.62 | 25.94 | 45.87 | 52.80 | 53.93 | 53.09 | 40.69 |
| 3 -methylbutanal | 8.50 | 13.85 | 16.68 | 17.91 | 17.28 | 21.92 | 29.08 | 20.49 |
| 3 -methylbutanoic acid | 0.40 | 0.59 | 0.46 | 0.45 | 0.60 | 0.55 | 0.51 | 0.33 |
| Acetic acid | 7.52 | 7.47 | 6.65 | 9.77 | 13.74 | 12.64 | 27.28 | 24.60 |
| Acetoin | 1.08 | 3.83 | 2.70 | 14.69 | 15.58 | 6.32 | 18.31 | 11.50 |
| Acetone | 23.59 | 26.38 | 25.44 | 32.27 | 39.10 | 27.78 | 37.47 | 32.52 |
| Ammonia | 7.87 | 11.48 | 8.83 | 9.60 | 10.05 | 9.14 | 7.19 | 6.84 |
| Butanone | 7.15 | 5.72 | 7.18 | 7.01 | 7.15 | 6.42 | 7.91 | 5.98 |
| Carbon disulfide | 54.93 | 98.73 | 75.50 | 138.42 | 77.18 | 65.74 | 69.84 | 56.05 |
| Dimethyl amine | 1.01 | 0.64 | 0.82 | 0.97 | 0.99 | 1.32 | 2.06 | 1.71 |
| Dimethyl disulfide | 0.11 | 0.15 | 0.09 | 0.57 | 0.34 | 0.35 | 2.38 | 2.14 |
| Dimethyl sulfide | 29.70 | 29.77 | 29.28 | 92.77 | 77.45 | 58.18 | 314.99 | 180.43 |
| Dimethyl trisulfide | 0.13 | 3.01 | 1.41 | 2.96 | 4.35 | 2.22 | 1.40 | 2.32 |
| Ethanol | 94.19 | 91.10 | 168.02 | 1072.36 | 537.23 | 485.03 | 6886.99 | 5410.47 |
| Ethyl acetate | 4.66 | 9.72 | 7.78 | 36.92 | 37.11 | 17.64 | 50.42 | 38.67 |
| Ethyl benzene | 2.36 | 2.29 | 2.09 | 3.32 | 3.68 | 3.41 | 2.93 | 2.48 |
| Hydrogen sulfide | 29.62 | 50.55 | 73.91 | 98.10 | 94.88 | 61.06 | 389.25 | 551.13 |
| Isobutyl alcohol | 11.81 | 19.41 | 15.43 | 27.62 | 20.74 | 19.84 | 36.85 | 28.60 |
| Methyl mercaptan | 1.03 | 0.83 | 1.21 | 7.60 | 3.17 | 3.33 | 51.19 | 46.18 |
| Piperidine | 0.31 | 1.39 | 1.00 | 1.48 | 1.38 | 1.07 | 1.44 | 1.36 |
| Propyl benzene | 0.76 | 1.01 | 0.76 | 1.66 | 1.36 | 1.03 | 1.33 | 1.37 |
| Styrene | 1.43 | 2.94 | 2.21 | 3.83 | 3.62 | 3.85 | 5.27 | 2.53 |
| Trimethyl amine | 1.53 | 2.70 | 3.90 | 11.58 | 13.08 | 6.17 | 7.84 | 7.91 |

Table B.14: (Continued)Measured concentrations (ppb) of VOCs detected with SIFTMS on each day in the headspace of salmon samples used in experiment AN4* (scoring and ranking with ties).

## Curriculum Vitae

## Personalia

| Name | Marc Sader |
| :--- | :--- |
| Date of birth | $06 / 10 / 1987$ |
| Place of birth | Beirut, Lebanon |
| Nationality | Lebanese |
| E-mail | Marc.Sader@UGent.be |

## Education

## University

$\diamond$ 2011-2013
Masters of Science in Analytical Instruments, Measurement and Sensor Technology, Coburg University, Coburg, Germany.
$\diamond 2012$
Masters of Science in Measuring and Testing Technologies and Instruments, University of Shanghai for Science and Technology, Shanghai, China.
$\diamond 2005-2010$
Bachelor of Engineering in Electrical Engineering, Lebanese American University, Byblos, Lebanon.

## Primary \& Secondary school

$\diamond$ 2002-2005:
International College, Beirut, Lebanon
$\diamond$ 1993-2002:
International College, Ain Aar, Lebanon

## Employment

$\diamond 2018$ - Current

Patent attorney trainee at NLO (Nederlandsch Octrooibureau)
$\diamond 2014-2017$
Full-time researcher at the Research Unit Knowledge-Based Systems, Department of Data Analysis and Mathematical Modelling, Faculty of Bioscience Engineering, Ghent University

## Scientific output

## Publications in international journals (ISI-papers)

$\diamond$ Pérez-Fernández R., Sader, M., and De Baets, B. (2018) "Joint consensus evaluation of multiple objects on an ordinal scale: an approach driven by monotonicity," Information Fusion, 42, 64-74.
$\diamond$ Kuuliala, L., Al Hage, Y., Ioannidis, A.-G., Sader, M., Kerckhof, F.-M., Vanderroost, M., Boon, N., De Baets, B., De Meulenaer, B., Ragaert, P., and Devlieghere, F. (2018) "Microbiological, chemical and sensory spoilage analysis of raw Atlantic cod (Gadus morhua) stored under modified atmospheres," Food Microbiology, 70, 232-244.
$\diamond$ Sader, M., Pérez-Fernández R., Kuuliala L., Devlieghere, F., and De Baets, B. (2018) "A combined scoring and ranking approach for determining overall food quality," International Journal of Approximate Reasoning, 100, 161-176.
$\diamond$ Sader, M., Verwaeren J., Pérez-Fernández R., and De Baets, B. (2019) "Integrating expert and novice evaluations for augmenting ordinal regression models," Information Fusion, 51, 1-9.
$\diamond$ Kuuliala L., Sader, M., Solimeo A., Pérez-Fernández R., Vanderroost M., De Baets B., De Meulenaer B., Ragaert P. and Devlieghere F. (2019) "Characterizing the spoilage of raw Atlantic salmon (Salmo salar) stored under modified atmospheres by multivariate statistics and augmented ordinal regression," International Journal of Food Microbiology, 303, 46-57.
$\diamond$ Sader, M., Pérez-Fernández R., Kuuliala L., Devlieghere, F., and De Baets, B. (2019) "The constrained median: a way to incorporate side information in the assessment of food samples," Applied Soft Computing, under review.

## Conference proceedings

$\diamond$ Sader, M., Verwaeren J., Ioannidis, A.G., Vanderroost, M., Devlieghere, F., and De Baets, B. (2015) "Predicting consumer acceptance of packaged meat using $\ell_{1}$-regularized ordinal regression," in Communications in Agricultural and Applied Biological Sciences, 80(1),123-127. Presented at the 20th National symposium of Applied Biological Sciences (NSABS 2015), Louvain-la-Neuve, Belgium.
$\diamond$ Sader, M., Pérez-Fernández R., and De Baets, B. (2018) "Combining absolute and relative information in studies on food quality," in Medina J. et al. (Eds.), Information Processing and Management of Uncertainty in KnowledgeBased Systems. Theory and Foundations. IPMU 2018. Communications in Computer and Information Science, 854, 379-388. Springer, Cham. Presented at the 17th International Conference IPMU, Cádiz, Spain.

## Conference abstracts

$\diamond$ Ioannidis, A.-G., Sader, M., Vanderroost, M., Devlieghere, F., and De Meulenaer, B. (2015) "Spoilage characterization of chicken breast fillets stored under different packaging and temperature conditions, using SIFTMS," in Applied Biological Sciences, 20th National Symposium, Abstracts, 80(2), 47. Poster presentation at the 20th National symposium of Applied Biological Sciences (NSABS 2015), Louvain-la-Neuve, Belgium.
$\diamond$ Kuuliala, L., Ioannidis, A.-G., Sader, M., Vanderroost, M., De Meulenaer, B., and Devlieghere, F. (2015) "Production of volatile organic compounds during spoilage of raw Atlantic cod fillets in modified atmosphere packaging." Poster presentation at the 20th National symposium of Applied Biological Sciences (NSABS 2015), Louvain-la-Neuve, Belgium.
$\diamond$ Kuuliala, L., Ioannidis, A.-G., Sader, M., Vanderroost, M., De Meulenaer, B., and Devlieghere, F. (2015) "Microbiological, chemical and sensory spoilage analysis of raw Atlantic cod (Gadus morhua): towards an integrated optical sensor for monitoring food quality." Presented at the Innovations in Food Packaging, Shelf Life and Food Safety Conference, Erding, Munich, Germany.
$\diamond$ Sader, M., Verwaeren J., and De Baets, B. (2016) "Combining expert and amateur opinions in sensory analysis of food." Poster presentation at the 21st

National symposium of Applied Biological Sciences (NSABS 2016), Antwerp, Belgium.
$\diamond$ De Baets, B., Pérez-Fernández R., and Sader, M. (2018) "Improving the quality of the assessment of food samples by combining absolute and relative information," in Book of Abstracts of the 32nd Annual Conference of the Belgian Operational Research Society (ORBEL32), pp. 157-158. Presented at ORBEL32 in Liège, Belgium.
$\diamond$ De Baets, B., Pérez-Fernández R., and Sader, M. (2018) "Joint aggregation of rankings and ordinal labels: A case study in food science," in Book of Abstracts of the International Symposium on Aggregation and Structures (ISAS 2018), pp. 55-56. Presented at ISAS 2018 in Valladolid, Spain.


[^0]:    1 We use the singular pronoun they and its derivative forms, such as them, their, etc., as an epicene (i.e., gender-neutral) singular pronoun. Although the use of singular pronouns in formal English has been the target of criticism since the late 19-th century, there is an increasing trend towards an epicene language 17 .
    ${ }^{2}$ This problem differs from a consensus reaching process, in which panellists discuss and modify their personal evaluations in order to reach a consensus evaluation [18, 19,20 .

[^1]:    1 The use of a MAP cannot reduce the initial contamination of a sample. Thus, it is not a replacement of good hygiene and handling practices throughout the food chain, which will assure that a MAP's spoilage inhibiting factors can be fully taken advantage of.

[^2]:    ${ }^{1}$ Note that a dot product $(\mathbf{x} \cdot \mathbf{x})$ is computationally equivalent to multiplying $\mathbf{x}$ (a $1 \times n$ vector) with $\mathbf{x}^{\top}$ (an $n \times 1$ vector).

[^3]:    2 The case where $\lambda=0$ is not considered because the problem would boil down to minimizing the (non-regularized) empirical risk.

[^4]:    ${ }^{1}$ Packaging was done in 950 ml trays ( $\mathrm{PP} / \mathrm{EVOH} / \mathrm{PP}$ ), and a top film ( $\mathrm{PA} / \mathrm{EVOH} /-$ $\mathrm{PA} / \mathrm{PP}$ ) with a thickness of $65 \mu \mathrm{~m}$ that allowed for O 2 , N 2 and CO 2 transmission rates $<5,1$ and $23 \mathrm{~cm}^{3} /\left(\mathrm{m}^{2} \cdot 24 \mathrm{~h} \cdot\right.$ bar $)$, respectively, at $23^{\circ} \mathrm{C}$ and $50 \% \mathrm{RH}$ and 1 atm .

[^5]:    2 Packaging was done in high barrier film bags ( O 2 transmission rate $<2.7 \mathrm{~cm}^{3} /\left(\mathrm{m}^{2} \cdot 24 \mathrm{~h} \cdot\right.$ bar) at $23^{\circ} \mathrm{C}$ and $\left.0 \% \mathrm{RH}\right)$.
    ${ }^{3}$ Plastic cups and lids with a diameter of 67 mm were purchased from AVA in Temse, Belgium.

[^6]:    ${ }^{4}$ FAO Major Fishing Areas. Atlantic, Northeast (Major Fishing Area 27).

[^7]:    5 Packaging was done with tray sealer MECA 900 (DecaTechnic, Herentals, Belgium) using multilayer packaging trays ( $\mathrm{PP} / \mathrm{EVOH} / \mathrm{PP}$ ) that allowed for an O 2 transmission rate of $0.035 \mathrm{~cm}^{3} /($ tray $\cdot 24 \mathrm{~h})$ at $23^{\circ} \mathrm{C}$ and $50 \% \mathrm{RH}$ and top film $(\mathrm{PA} / \mathrm{EVOH} / \mathrm{PA} / \mathrm{PP})$ that allowed for an O2 transmission rate of $6.57 \mathrm{~cm}^{3} /\left(\mathrm{m}^{2} \cdot 24 \mathrm{~h} \cdot \mathrm{~atm}\right)$ at $23^{\circ} \mathrm{C}, 50 \% \mathrm{RH}$ and 1 atm .
    ${ }^{6}$ Packaging was done in high barrier film bags ( O 2 transmission rate $<2.7 \mathrm{~cm}^{3} /\left(\mathrm{m}^{2} \cdot 24 \mathrm{~h} \cdot\right.$ bar) at $23^{\circ} \mathrm{C}$ and $\left.0 \% \mathrm{RH}\right)$.

[^8]:    7 Plastic cups and lids with a diameter of 67 mm were purchased from AVA in Temse, Belgium.

[^9]:    8 FAO Major Fishing Areas. Atlantic, Northeast (Major Fishing Area 27).
    9 Packaging was done with a tray sealer (MECA 900, DecaTechnic, Herentals, Belgium) using multilayer packaging trays ( $\mathrm{PP} / \mathrm{EVOH} / \mathrm{PP}$ ) that allowed for an O 2 transmission rate of $0.03 \mathrm{~cm}^{3} /($ tray $\cdot 24 \mathrm{~h})$ at $23^{\circ} \mathrm{C}$ and $50 \% \mathrm{RH}$ and top film (PA/EVOH/PA/PP) that allowed for an O 2 transmission rate of $6.57 \mathrm{~cm}^{3} /\left(\mathrm{m}^{2} \cdot 24 \mathrm{~h} \cdot \mathrm{~atm}\right)$ at $23^{\circ} \mathrm{C}, 50 \% \mathrm{RH}$ and 1 atm .

[^10]:    Ranking
    ${ }^{10}$ Packaging was done in high barrier film bags ( O 2 transmission rate $<2.7 \mathrm{~cm}^{3} /\left(\mathrm{m}^{2} \cdot 24 \mathrm{~h} \cdot\right.$ bar) at $23^{\circ} \mathrm{C}$ and $0 \% \mathrm{RH}$ ).
    ${ }^{11}$ Plastic cups and lids with a diameter of 67 mm were purchased from AVA in Temse, Belgium.

[^11]:    ${ }^{12}$ FAO Major Fishing Areas. Atlantic, Northeast (Major Fishing Area 27).
    ${ }^{13}$ Packaging was done with tray sealer MECA 900 (DecaTechnic, Herentals, Belgium), using 610 ml trays (PP/EVOH/PP) that allowed for an O2 transmission rate of $0.001 \mathrm{~cm}^{3} /($ tray $\cdot 24 \mathrm{~h})$ at $23^{\circ} \mathrm{C}$ and $50 \% \mathrm{RH})$ and top film ( $\mathrm{PA} / \mathrm{EVOH} / \mathrm{PA} / \mathrm{PP}$ ) that allowed for an O 2 transmission rate of

[^12]:    $6.57 \mathrm{~cm}^{3} /\left(\mathrm{m}^{2} \cdot 24 \mathrm{~h} \cdot \mathrm{~atm}\right)$ at $23^{\circ} \mathrm{C}, 50 \% \mathrm{RH}$ and 1 atm .
    ${ }^{14}$ Packaging was done in high barrier film bags ( O 2 transmission rate $<2.7 \mathrm{~cm}^{3} /\left(\mathrm{m}^{2} \cdot 24 \mathrm{~h} \cdot\right.$ bar) at $23^{\circ} \mathrm{C}$ and $0 \% \mathrm{RH}$ ).

[^13]:    ${ }^{15}$ Plastic cups and lids with a diameter of 67 mm were purchased from AVA in Temse, Belgium.

[^14]:    1 The reader may note that these functions are not known by the name of monometric in other fields of application, although most of the times the considered functions turn out to satisfy the three axioms of a monometric.

[^15]:    2 The monometric $M$ can be understood as the ground distance function for the Earth Mover's Distance. Intuitively, the Earth Mover's Distance is a distance between probability distributions that describes the minimum amount of 'soil' that needs to be moved in order to transform the first distribution into the second one [122].

[^16]:    4 The search for a closest marginally monotone matrix of labels can be addressed as a particular case of the optimization problem discussed in this section applied separately to each singleton subset of the set of samples.
    5 An order needs to be assigned to the elements in both sets, for instance, the lexicographical order.

[^17]:    6 We recall that, for a singleton, both the marginal majority and the majority consensus states are equivalent.
    ${ }^{7}$ It suffices to change four labels $L_{5}$ and two labels $L_{7}$ into six labels $L_{6}$, leading to the vector of frequencies ( $1,2,1,0,0,6,0)$.
    8 We recall that, for a singleton, both the marginal monotonicity and the monotonicity consensus states are equivalent.
    ${ }^{9}$ It suffices to change one label $L_{2}$ into a label $L_{3}$, one label $L_{3}$ into a label $L_{4}$ and one label $L_{7}$ into a label $L_{6}$, leading to the vector of frequencies $(1,1,1,1,4,1,1)$.

[^18]:    ${ }^{1}$ Note that the subscript ' $U$ ' (in $n_{U}$ ) refers to untrained panellists as opposed to the subscript ' $T$ ' (in $n_{T}$ ) which refers to trained panellists.

[^19]:    2 The Borda count is similar to the arithmetic mean, where the position of a sample in the aggregated ranking is the average of the individual positions of the considered sample.

[^20]:    3 This is based on the Kendall distance function between rankings, which will be described in detail in the following subsection.
    ${ }^{4}$ Later on in this dissertation, this will be called a votrix, which will be defined more formally.

[^21]:    5 The Kemeny distance and the Kendall distance will be described in detail in the following subsection.

[^22]:    1 We use the verb 'to integrate' to specify that the first part is united with the second part on the basis of equality.

[^23]:    ${ }^{2}$ We use the verb 'to incorporate' to specify that the first part is assimilated with the second part to form a whole.

[^24]:    3 Note that the use of superscripts in this example denotes a first and second ranking. This is not to be confused with the use of subscripts which denote the ranking with ties provided by a panellist.

[^25]:    ${ }^{4}$ We write the word 'distance' in between quotation marks since we are comparing objects of a different nature, and, thus, we are lacking the semantics associated with the mathematical formalization of a distance (metric).
    ${ }^{5}$ Generally, both terms to be combined in a convex combination may take values with significantly different orders of magnitude. Hence, to normalize $C_{\alpha}(\mathbf{s})$, we divide the distances associated with the vectors of scores and the rankings by their respective upper bounds $B_{T}$ and $B_{U}$.

[^26]:    6 This is not to be confused with the notation used in Section 7.4.2 where the same problem setting is considered, however, the scores are not represented in increasing order.

[^27]:    7 The situation where samples are tied is similar to that where samples are in the same cluster. For simplicity, the special case where $\epsilon=0$ is considered.

[^28]:    ${ }^{8}$ In general, for $n$ samples and $k$ scores, the number of possible vectors of scores is $k^{n}$.

[^29]:    9 Values of the cophenetic correlation coefficient closer to 1 indicate a higher clustering accuracy.

[^30]:    1 The case where $\alpha=1$ is not considered because the problem would boil down to minimising only the regularization parameter.

