

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

**METODOLOGIAS PARA SELEÇÃO DE VARIÁVEIS EXPLICATIVAS  
E DETECÇÃO DE INCONFORMIDADES DE PREDIÇÃO  
APLICADAS À ESPECTROSCOPIA POR FLUORESCÊNCIA**

TESE DE DOUTORADO

*Lucas Ranzan*

**Porto Alegre**

**2021**



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

**METODOLOGIAS PARA SELEÇÃO DE VARIÁVEIS  
EXPLICATIVAS E DETECÇÃO DE  
INCONFORMIDADES DE PREDIÇÃO APLICADAS À  
ESPECTROSCOPIA POR FLUORESCÊNCIA**

*Lucas Ranzan*

Tese de Doutorado apresentada como requisito parcial para obtenção do título de Doutor no Programa de Pós-Graduação em Engenharia Química da UFRGS

*Área de concentração:* Pesquisa e Desenvolvimento de Processos

*Linha de Pesquisa:* Projeto, Simulação, Modelagem, Controle e Otimização de Processos Químicos e Bioprocessos.

**Orientadores:**

**Prof. Dr. Jorge Otávio Trierweiler**

**Prof<sup>a</sup>.Dr<sup>a</sup>. Luciane Ferreira Trierweiler**

**Porto Alegre**

**2021**

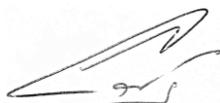




UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

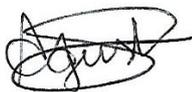
A Comissão Examinadora, abaixo assinada, aprova a Proposta de Pesquisa *Metodologias para Seleção de Variáveis Explicativas e Detecção de Inconformidades de Predição Aplicadas à Espectroscopia por Fluorescência*, elaborada por Lucas Ranzan, como requisito parcial para obtenção do Grau de Doutor em Engenharia.

Comissão Examinadora:



---

Prof. Dr. Adilson Ben da Costa



---

Prof. Dra. Caroline Borges Agustini



---

Prof. Dr. Michel J. Anzanello





# Resumo

A capacidade de prever eventos futuros a partir de conhecimentos históricos é a base para a modelagem preditiva. Criar um modelo capaz de quantificar variáveis de interesse, classificar ocorrências ou prever comportamentos, acompanham a evolução dos algoritmos modernos de aprendizado de máquina. Na indústria de transformação, muitas das informações mais relevantes para o controle de processos ainda são adquiridas unicamente através de técnicas laboratoriais, que são custosas, destrutivas e morosas (como, por exemplo, concentração molecular de espécies de interesse, pureza de fármacos, lubrificidade de óleos, teor de proteína em alimentos, etc.). Um possível caminho para automação destes sistemas é o estudo de novos sensores capazes de capturar uma informação auxiliar de fácil obtenção, que possa ser transformada matematicamente nas saídas de interesse. Surge então a aspiração por estudos que combinam a escolha de sensores adequados com metodologias capazes de extrair de maneira eficiente a informação útil contida nestes dados. Neste trabalho são apresentadas metodologias baseadas em diferentes estratégias para seleção de variáveis explicativas e otimização de modelos empíricos. Ainda, é proposta uma metodologia para qualificação de inconformidades em novas leituras utilizando redes neurais. É apresentada a metodologia AnTSbe, um algoritmo híbrido baseado nas meta-heurísticas Colônia de Formigas (ACO) e Busca Tabu (TS), desenvolvido para otimizar a seleção de variáveis de entrada em problemas combinatórios complexos. A hibridização das meta-heurísticas visa evitar a estagnação precoce e a ciclagem de subgrupos, comuns nessas metodologias. O algoritmo também introduz o uso da expansão polinomial e combinatória das variáveis de entrada, em um esforço para incrementar o poder preditivo dos modelos. Como estudo de caso, espectroscopia por fluorescência é utilizada para prever concentração de enxofre em diesel combustível. Os modelos preditivos ajustados foram superiores a outras técnicas descritas na literatura, com erros absolutos percentuais médios de predição menores que 4%. As adaptações propostas se mostraram eficientes, quando comparadas a pesquisas prévias com a mesma base de dados. Uma adaptação é proposta ao algoritmo AnTSbe, focada para dados de fluorescência, com o conceito de Delta Pair. Uma nova camada de otimização é introduzida no algoritmo a fim de selecionar um par Excitação/Emissão que serve como regulador do meio, tendo sua intensidade de fluorescência decrescida de todos os pontos do espectro. Neste estudo, são acompanhados três processos distintos de envelhecimento de cachaça, com o intuito de prever a concentração de fenólicos na bebida ao longo do tempo, com base em dados fluorescência. A adaptação Delta Pair se mostrou especialmente funcional quando combinada com expansão de bases e para predição de cachaças envelhecidas comerciais, que não participaram da etapa de calibração dos modelos. A seguir, matrizes excitação – emissão de fluorescência captadas *in situ* em fermentações com *S. cerevisiae* foram utilizadas para calibrar uma rede neural convolucional residual, como intuito de prever glicose, etanol e biomassa no meio biológico. Em paralelo, foi desenvolvida uma metodologia baseada em redes neurais do tipo *autoencoder* (AE), capazes de corretamente reconstruir os espectros originais. A metodologia utiliza o erro de reconstrução da rede AE treinada para triagem não supervisionada de novos espectros, conseguindo identificar espectros com inconformidades, e qualificar a confiança que se pode atribuir a um novo dado, baseado na magnitude deste erro. Por fim, a metodologia AnTSbe é utilizada para prever impurezas

nas correntes de uma unidade de separação de propano/propeno, expandindo o uso da metodologia para casos da indústria petroquímica com base em dados simulados de processo (e não de fluorescência). A metodologia se mostrou capaz de corretamente prever os perfis de concentração das três colunas de separação do processo com erros absolutos percentuais médios inferiores a 5%, com foco especial para quantificação dos contaminantes em cada corrente, que precisam ser mantidos sob controle para garantir a lucratividade da operação. Os artigos desenvolvidos demonstram, inclusive na ordem apresentada, o sucesso das metodologias propostas em aprofundar a seleção de variáveis significativas e otimização de modelos empíricos preditivos. A sucessão dos casos estudados parte do desenvolvimento do algoritmo estocástico base, segue para a busca de um reforço na capacidade de generalização dos modelos otimizados baseados em espectroscopia por fluorescência, apresenta uma técnica para qualificação de novas amostras e conclui com o uso dos algoritmos desenvolvidos em um caso industrial.



## Abstract

The ability to predict future events from historical observations is the basis for predictive modeling. Creating a model capable of quantifying variables of interest, classifying occurrences or predicting behavior, follows the evolution of modern machine learning algorithms. In the manufacturing industry, much of the most relevant information for process control is still acquired only through laboratory techniques, which are costly, destructive and time-consuming (such as, for example, molecular concentration of species, purity of drugs, lubricity of oils, protein content in food, etc.). A possible way to automate these systems is the study of new sensors capable of capturing auxiliary information of easy application, which can be mathematically transformed in the outputs of interest. This is the aspiration for studies that combine the choice of skilled sensors with methodologies capable of efficiently extracting the useful information contained in the data. In this work we propose methodologies based on different machine learning methods for the optimization of empirical models. AnTSbe methodology is presented, a hybrid algorithm based on Ant Colony (ACO) and Tabu Search (TS) metaheuristics, developed to optimize the selection of input variables in complex combinatorial problems. The hybridization of metaheuristics aims to avoid early stagnation and cycling of subgroups, common in these methodologies. The algorithm also introduces the use of polynomial and combinatorial expansion of the input variables, in an effort to increase the predictive power of the models. As a case study, fluorescence spectroscopy is used to predict sulfur concentration in diesel fuel. The adjusted predictive models were superior to other techniques from literature, with mean absolute percentage errors of prediction smaller than 4%. The proposed adaptations were efficient, when compared to previous researches with the same database. An adaptation is proposed to the AnTSbe algorithm, focused on fluorescence data, with the concept of DeltaPair. A new optimization layer is introduced in the algorithm in order to select an Excitation/Emission pair that serves as a medium regulator, having its fluorescence intensity decreased from all other points in the spectrum. In this study, three distinct cachaça aging processes are followed, in order to predict the concentration of phenolics in the spirit over time, based on fluorescence data. The DeltaPair adaptation is especially functional when combined with base expansion and for the prediction of aged commercial cachaças, which does not participate in the calibration stage of the models. Following, fluorescence excitation - emission matrices, collected *in situ* in fermentations with *S. cerevisiae*, were used to calibrate a residual convolutional neural network, in order to predict glucose, ethanol and biomass in the biological environment. In parallel, a methodology based on autoencoder-type neural networks (AE) was developed, capable of correctly reconstructing the original spectra. The methodology uses the trained AE reconstruction error for unsupervised screening of new spectra, managing to identify abnormal spectra, and to qualify the confidence that can be attributed to a new data, based on the magnitude of this error. Despite the focus on fluorescence spectroscopy data, most of the methodologies were designed to be of general use, whatever the data source, with little or no modification. Finally, the AnTSbe methodology is used to predict impurities in

the streams of a propane/propylene splitter unit, expanding the use of the methodology for cases in the petrochemical industry based on simulated process data (and not fluorescence). The methodology proved to be capable of correctly predicting the concentration profiles of the three process' separation columns with mean absolute percentage errors below 5%, with a special focus on quantifying the contaminants in each stream, which need to be kept under control to ensure profitability of the operation. The articles developed demonstrate, in the order presented, the success of the proposed methodologies in deepening the selection of significant variables and the optimization of predictive empirical models. The succession of the studied cases starts from the development of the base stochastic algorithm, goes on to seek a reinforcement in the generalizability of the optimized models based on fluorescence spectroscopy, presents a technique for qualifying new samples and concludes with the use of the algorithms developed in an industrial case.



*“Tanto a estrada quanto a história têm sido longas, você não concordaria? A viagem tem sido longa e o custo tem sido alto...mas nunca uma coisa grande foi alcançada com facilidade. Uma longa história, como uma Torre alta, tem que ser construída pedra por pedra”.*

*(Stephen King – A Torre Negra)*





## Agradecimentos

Início estes agradecimentos dirigindo minhas palavras as pessoas que, sem dúvida nenhuma, foram as responsáveis por eu poder estar aqui escrevendo este texto: minha família. Mãe e pai, obrigado por me darem todas as condições de suporte, amor e carinho para seguir minhas escolhas e meu próprio caminho. Mano, muito obrigado por estar sempre do meu lado, principalmente nas horas mais difíceis, quando a esperança e a força de vontade falhavam. Vocês sempre foram e sempre serão meus maiores modelos, não só por me mostrarem, cada um do seu jeito, a importância da educação, mas por serem a representação física e palpável do que eu considero o amor incondicional de uma família.

Agradeço imensamente aos meus orientadores Jorge Otávio Trierweiler e Luciane Ferreira Trierweiler, sem os quais esta tese não existiria. Vocês não só foram importantes para todo o desenvolvimento conceitual e a aplicação prática das ideias que desenvolvemos neste trabalho, como tem sido a base sólida sobre a qual construí grande parte da minha vida acadêmica. Obrigado por todo suporte ao longo desses vários anos, pelos ensinamentos, por apoiarem as minhas ideias e por todas as oportunidades que me proporcionaram. Agradeço ao Prof. Dr. Bernd Hitzmann e aos colegas que me acolheram em seu grupo de pesquisa em Hohenheim.

Sem dúvida, não posso deixar de agradecer aos amigos e amigas que trilharam ao meu lado essa longa caminhada. Muito obrigado aos meus irmãos e agregados da G.M., a minha vida não teria metade da graça sem vocês nela. Obrigado Karen e Twin, por sempre serem meus portos-seguros. Sei que vocês nunca duvidaram, por nenhum segundo sequer, que eu chegaria até aqui. Obrigado ao Ka-tet Chambourcy, Presuntinho, Carol, Ju, Yana, Judel, Pistolinha e aos amigos que a eng. química me proporcionou. Vocês foram e continuam sendo parte especial de mim. Agradeço a Marina por dividir comigo por muitos anos a carga extenuante de produzir esse trabalho. Obrigado Elisa por estar ao meu lado no fim desta jornada, pelo conforto e por todo apoio nessa nova fase que se inicia na minha vida. Obrigado aos amigos que me acolheram e me ajudaram durante a minha estadia na Alemanha. Nenhuma jornada é construída sozinho, e eu tenho sorte em poder ter dividido essa com vocês. Agradeço aos colegas e amigos que fiz no DEQUI, em especial ao Pedro e aos demais que participaram ativamente como colaboradores das minhas pesquisas, e aos alunos que tive que a honra de orientar.

Por fim, agradeço a todas as pessoas que tiveram parte neste trabalho e na minha vida nestes anos. Muito Obrigado.





## SUMÁRIO

<b>Capítulo 1 – Introdução .....</b>	<b>1</b>
1.1 Motivação.....	1
1.2 Objetivos e Estruturação do Trabalho .....	4
1.3 Produção Científica e Contribuições Paralelas.....	6
1.4 Contribuições .....	7
1.5 Resumo Gráfico .....	8
1.6 Referências.....	8
<b>Capítulo 2 – Revisão Bibliográfica .....</b>	<b>10</b>
2.1 Modelos Empíricos .....	10
2.1.1 Regressões Lineares e Soma dos Quadrados dos Resíduos .....	12
2.1.2 Normas de Regularização .....	13
2.2 Otimização de Modelos e Seleção de Variáveis.....	15
2.2.1 Meta-heurísticas de Otimização .....	17
2.3 Redes Neurais Artificiais.....	20
2.3.1 Redes Neurais Convolucionais.....	23
2.3.2 Normalização em lote - Batch normalization .....	25
2.3.3 Evolução da arquitetura CNN .....	25
2.3.4 Redes Autoencoder - AE .....	26
2.4 Espectroscopia .....	29
2.4.1 Absorção, Emissão e Deslocamento de Stokes .....	31
2.4.2 Matriz Excitação – Emissão e Espectrofluorímetro .....	33
2.4.3 Supressão de Fluorescência - Quenching .....	35
2.5 Referências.....	35
<b>Capítulo 3 – Prediction of Sulfur Content in Diesel Fuel using Fluorescence Spectroscopy and a Hybrid Ant Colony - Tabu Search Algorithm with Polynomial Basis Expansion      43</b>	
3.1 Introduction.....	44
3.2 Methodology .....	45
3.2.1 Preprocessing .....	45
3.2.2 AnTSbe – Ant Colony Optimizer hybridized with Tabu Search and basis expansion ....	47
3.3 Case Study – Quantifying total sulfur content in diesel fuel samples using EEM fluorescence spectroscopy.....	53
3.3.1 Dataset.....	54
3.3.2 AnTSbe Pre-processing and Parameters.....	55
3.4 Results and Discussions.....	57
3.4.1 Diesel S100.....	57
3.4.2 Diesel S10.....	59
3.4.3 Tabu Memory Activations.....	62
3.4.4 Contrast with Previous Works .....	63
3.5 Conclusions.....	63
3.6 Acknowledgment .....	64
3.7 Abbreviations .....	64
3.8 References.....	64
<b>Capítulo 4 – Prediction of Total Phenolic Content in Wood-Aged Cachaças using a Hybrid Ant Colony – Tabu Search algorithm and Fluorescence Spectroscopy with a Reference Spectral Pair .....</b>	<b>69</b>

4.1	Introduction.....	70
4.2	Material and Methods.....	70
4.2.1	Barrels and Cachaça for Aging .....	70
4.2.2	Quantification of Total Phenolic Concentration.....	71
4.2.3	EEM Fluorescence Spectroscopy .....	71
4.2.4	Data Preprocessing .....	72
4.2.5	Chemometric analysis – Adapted AnTSbe algorithm .....	72
4.3	Results and Discussion .....	75
4.3.1	Quantification of Total Phenolic Concentration.....	75
4.3.2	EEM Fluorescence Spectroscopy .....	76
4.3.3	Chemometric analysis – Adapted AnTSbe algorithm .....	78
4.4	Conclusions.....	81
4.5	Acknowledgments.....	82
4.6	References.....	82
4.7	Appendix A .....	83
<b>Capítulo 5 – Avoiding Misleading Predictions in Fluorescence-based Soft Sensors using Autoencoders.....</b>		<b>88</b>
5.1	Introduction.....	89
5.2	Materials and Methods.....	92
5.2.1	Saccharomyces cerevisiae Fermentation Monitoring using EX/EM Fluorescence.....	92
5.2.2	Neural Networks.....	93
5.2.3	Trust Screening and Anomaly Detection .....	96
5.3	Results and Discussions.....	97
5.3.1	R-ResNet .....	97
5.3.2	Autoencoder .....	98
5.4	Conclusions.....	101
5.5	References.....	102
<b>Capítulo 6 – Developing Impurity Soft Sensors for a Propylene/Propane Splitter Unit 104</b>		
6.1	Introduction.....	105
6.2	Materials and Methods .....	106
6.2.1	Propylene/propane Splitter Unit (PPSU) .....	106
6.2.2	Development of soft-sensors for the PPSU .....	108
6.3	Results and Discussions.....	110
6.3.1	Column T-01.....	110
6.3.2	Column T-02.....	113
6.3.3	Column T-03.....	115
6.4	Conclusions.....	119
6.5	References.....	120
<b>Capítulo 7 – Conclusões e Trabalhos Futuros.....</b>		<b>122</b>
7.1	Conclusões.....	122
7.2	Sugestões para Trabalhos Futuros .....	125

## LISTA DE FIGURAS

Figura 1.1. Resumo gráfico, indicando as correlações entre objetivos, capítulos e contribuições do trabalho. ....	8
Figura 2.1. Representação visual de um modelo caixa-preta. ....	10
Figura 2.2. Comparação entre as penalizações Lasso (esquerda) e regressão Ridge (direita). A área em azul representa as regiões de restrição $ \beta_1  +  \beta_2  \leq t$ e $\beta_1^2 + \beta_2^2 \leq t^2$ , respectivamente. As elipses vermelhas representam o contorno da função erro dos mínimos quadrados. Fonte: (HASTIE, TREVOR, TIBSHIRANI, ROBERT, FRIEDMAN, 2009)..	14
Figura 2.3. Simplificação visual da forma matricial da regressão PLS. ....	16
Figura 2.4. Definição de como formigas reais escolhem o menor caminho. (a) As formigas chegam ao ponto de decisão. (b) Randomicamente decidem o caminho a seguir. (c) As formigas que escolheram o caminho mais curto chegam ao ponto oposto em menos tempo. (d) O feromônio acumula no caminho mais curto em uma taxa mais alta. Assim, mais formigas tendem a seguir por ele. As linhas pontilhadas representam uma aproximação proporcional do feromônio depositado pelas formigas. Fonte: (DORIGO; GAMBARELLA, 1997). ....	19
Figura 2.5. Estrutura básica de um neurônio artificial. Fonte: Schimidt <i>et al.</i> (2016). ....	20
Figura 2.6. Diferentes arquiteturas de redes neurais artificiais. Fonte: (VERMA; KUMAR SINGH, 2015). ....	21
Figura 2.7. Funções de ativação comuns em redes neurais artificiais. Fonte: (CHARTE <i>et al.</i> , 2018). ....	22
Figura 2.8. Exemplo de CNN aplicada na classificação de imagens. A imagem original possui 224x224 pixels e três canais ( <i>Red-Blue-Green</i> ). O espaço original é reduzido por camadas <i>max pooling</i> . A terceira dimensão apresentada é referente ao número de <i>kernels</i> (filtros) utilizados na camada convolucional. Fonte: (SIMONYAN; ZISSERMAN, 2015). ....	23
Figura 2.9. Representação de operação convolucional. Adaptado de Yamashita <i>et al.</i> (2018). ....	24
Figura 2.10. Camada <i>max pooling</i> com kernel [2x2], <i>stride 2</i> e sem <i>padding</i> . Adaptado de Yamashita <i>et al.</i> (2018). ....	25
Figura 2.11. Histórico evolutivo das CNN profundas, com destaque para as inovações nas arquiteturas. Fonte: (KHAN <i>et al.</i> , 2020). ....	26
Figura 2.12. Esquema genérico da arquitetura de um AE. Fonte: (BAHI; BATOUCHE, 2018). ....	27
Figura 2.13. Projeção espacial dos dois primeiros componentes principais – PCA (A) e do espaço latente de 2 neurônios da rede AE (B), para o conjunto dígitos escritos a mão, MNIST. Fonte: (HINTON; SALAKHUTDINOV, 2006). ....	27
Figura 2.14. Resumo das escolhas necessárias para o design de um AE. Fonte: (CHARTE <i>et al.</i> , 2018). ....	28
Figura 2.15. Diagrama de Jablonski. $S_0$ representa o estado eletrônico fundamental, $S_1$ , e $T_1$ são os estados excitados singleto e tripleto. $S_2$ é um segundo estado excitado singleto. As linhas horizontais pontilhadas são os vários níveis de energia vibracional dos estados. As setas retas representam os processos envolvendo fótons, e as setas onduladas representam transições não-radioativas. Fonte: (SKOOG; HOLLER; CROUCH, 2009). ....	31
Figura 2.16. Espectro de absorção e emissão de fluorescência do perileno (superior) e da quinina (inferior). O Deslocamento de Stokes pode ser visto claramente em ambas moléculas. O perileno segue a regra da Imagem Espelhada, mas o mesmo não ocorre para a quinina. Fonte: (LAKOWICZ, 2006). ....	33
Figura 2.17. Matrizes excitação – emissão de fluorescência para uma mistura de antraceno e ovaleno (a), apresentada na forma tridimensional, e para 8-	

hidroxibenzopireno (b), apresentada como curvas de contorno. Fonte: (SKOOG; HOLLER; CROUCH, 2009).....	34
Figura 2.18. Representação genérica da estrutura de um espectrofluorímetro. Fonte: (SKOOG; HOLLER; CROUCH, 2009). .....	35
Figure 3.1. Example of pheromone-based input selection with a random trigger of 0.856. ....	50
Figure 3.2. Schematic representation of the AnTSbe algorithm.....	53
Figure 3.3. Average EEM fluorescence spectra for the Diesel S10 and Diesel S100 sample groups.....	55
Figure 3.4. Diesel S100 Measured vs Predicted outputs for the <i>Ridge</i> , <i>LassoLars</i> , and Filtered <i>Ridge</i> global solutions, with model size 5. ....	59
Figure 3.5. Global solutions' ( $Nw = 5$ ) selected fluorescence pairs for Diesel S100 (Black hexagon – <i>LassoLars</i> ; Red cross – <i>Ridge</i> ; Yellow star – Filtered <i>Ridge</i> ). .....	59
Figure 3.6. Diesel S10 Measured vs. Predicted outputs for the Filtered <i>Ridge</i> global solution, with model size 4.....	61
Figure 3.7. Global solution' ( $Nw = 4$ ) selected fluorescence pairs for Diesel S10.....	61
Figure 3.8. Individually normalized final pheromone trail of optimizations with Diesel S100 and model sizes 3, 4, and 5. ....	62
Figure 4.1. Schematic representation of the original AnTSbe algorithm (A) and the proposed adapted AnTSbe (B).....	73
Figure 4.2. The phenolic concentration of aging cachaças by time. $CA_x$ indicates the aging process.....	75
Figure 4.3. EEM fluorescence spectra of cachaças with increasing phenolic concentration. Above each subplot, the $CA_x$ refers to the aging process, followed by the sample's total phenolic content.....	77
Figure 4.4. Mean EEM fluorescence intensity vs total phenolic concentration of each sample. ....	78
Figure 4.5. Predicted vs. Measured outputs for the AD arrangements, using all available samples and model size 5.....	79
Figure 4.6. Predicted vs. Measured outputs for the AD arrangements, using only CA2/CA3 and Amburana commercial samples, with model size 5.....	80
Figure 4.7. MD03 and AD03 selected fluorescence Ex/Em pairs and $\Delta P_s$ (model size 5)... ..	81
Figure 5.1. Plain network (left) and shortcut connection (right). Source: (HE et al., 2016a). ....	90
Figure 5.2. Autoencoder: Example schema of the architecture of the neural network. Adapted from Charte <i>et al.</i> (2018). ....	91
Figure 5.3. Absolute change in the relative fluorescence intensity of each Ex/Em pair, for the fermentations.....	92
Figure 5.4. Offline and extrapolated measurements for the outputs of interest in each fermentation.....	93
Figure 5.5. Residual blocks and overall architecture of the network. (a) First Residual Block with Projection Shortcut (RB – $PS_0$ ). (b) Residual Block with Projection Shortcut (RB – PS). (c) Residual Block with Identity Shortcut (RB – IS). (d) Overall Architecture of the network. ....	94
Figure 5.6. Summary of possible parameters of an autoencoder (CHARTE et al., 2018). ..	95

Figure 5.7. R-ResNet predicted and extrapolated measured outputs for P1, P2, and P3 fermentations. For visualization purposes, data points were linearly connected as continuous lines.....	98
Figure 5.8. Autoencoder predictive reconstruction root mean squared errors for P1, P2, and P3 fermentations.....	99
Figure 5.9. Defect test results. The superior graphics present the R-ResNet output predictions for each fermentation, and the inferior graphics the AE reconstruction RMSE. Samples marked with a dot are the ones that received fluorescence intensity errors in pair Ex450/Em530. The percentage of the original intensity added to each sample is described. For visualization purposes, data points were linearly connected as continuous lines.....	100
Figure 5.10. Autoencoder reconstruction RMSE for all samples with a constant 15% added relative fluorescence intensity. ....	101
Figure 6.1. Simplified flowchart of the studied propane-propylene splitter unit. Source: (SCHULTZ, 2015). ....	107
Figure 6.2. Adhesion of the <i>Ridge</i> model to the real outputs $ZC4 + D1$ , divided by samples with mass ratios above 0.01 (left) and bellow 0.01 (right). ....	112
Figure 6.3. Adhesion of the <i>Ridge</i> (left) and the AnTSbe (right) local models to the real $ZC4 + D1$ of Column T-01, for samples with heavy hydrocarbons concentration bellow 0.01 kg/kg. ....	113
Figure 6.4. Real <i>versus</i> the predicted output of $ZC3 - D2$ for Column T-02, for the <i>Ridge</i> model (left) and the AnTSbe model (right). ....	115
Figure 6.5. Real <i>versus</i> the AnTSbe predicted $ZC3 - D3$ for the distillate stream of Column T-03 .....	117
Figure 6.6. Individual percentage errors for T-03 AnTSbe fitted model, for samples with $ZC3 + D3$ lower than 0.0002 (left) and between 0.0002 to 0.01 (right). ....	118
Figure 6.7. Column T-03 AnTSbe model's adhesion to the real outputs $ZC3 + D3$ , for samples with propane concentration between 0.0002 and 0.01 kg/kg .....	118
Figure 6.8. Column T-03 AnTSbe model's adhesion to the real outputs $ZC3 + D3$ , for samples with propane concentration below 0.0002 kg/kg.....	119
Figure A1. EEM fluorescence spectra of 20 out of 21 samples from CA1 aging process, including first and last collected samples. Above each subplot, the phenolic concentration of that sample is shown.84	
Figure A2. EEM fluorescence spectra of 20 out of 41 samples from CA2 aging process, including first and last collected samples. Above each subplot, the phenolic concentration of that sample is shown.....	85
Figure A3. EEM fluorescence spectra of 20 out of 41 samples from CA3 aging process, including first and last collected samples. Above each subplot, the phenolic concentration of that sample is shown.....	86
Figure A4. EEM fluorescence spectra of all commercial samples. Above each subplot, the name and phenolic concentration of that sample is shown. ....	87
Figure A5. Mean fluorescence intensity of samples vs. their total phenolic concentration. CA <sub>x</sub> indication refers to the aging processes. ....	87



## LISTA DE TABELAS

Tabela 2.1. Características e aplicações de técnicas de espectroscopia vibracional. ....	30
Table 3.1. AnTSbe general optimization parameters for diesel fuel local models. ....	56
Table 3.2. Global solutions' metrics and selected fluorescence pairs for Diesel S100 - <i>first optimization</i> . ....	57
Table 3.3. Global solutions' metrics and selected fluorescence pairs for Diesel S100 - <i>filtered optimization</i> and PLS regression (7 LV) metrics for comparison. ....	58
Table 3.4. Global solutions' metrics and selected fluorescence pairs for Diesel S10 - <i>filtered optimization</i> . PLS regression (4 LV) metrics for comparison. ....	60
Table 3.5. Comparison between Global Solutions of AnTSbe and PSCM (RANZAN et al., 2015) methodologies for diesel S100 and S10, with model sizes 5 and 4, respectively. ....	63
Table 4.1. Adapted AnTSbe general optimization parameters. ....	74
Table 4.2. Phenolic concentration of aged commercial cachaças. ....	76
Table 4.3. Metrics, selected fluorescence pairs, and $\Delta P$ for the MD predictive models with model size 5, using all available samples. ....	78
Table 4.4. Metrics, selected fluorescence pairs, and $\Delta P$ for the predictive AD models with model size 5, using only CA2, CA3 and Amburana commercial samples. ....	80
Table 5.1. R-ResNet structure summary. ....	97
Table 5.2. R-ResNet prediction metrics, relative to the extrapolated offline measurements. ....	98
Table 6.1. Description of the variables, nominal values, ranges, and the number of equidistant points used as inputs for each distillation column's neural networks' simulations. Source: (SCHULTZ, 2015). ....	108
Table 6.2. Description of the main output of interest in T-01, the mass ratio of heavies in the distillate stream ( $ZC4 + D1$ ). ....	111
Table 6.3. Description of the 9 available model inputs for Column T-01. ....	111
Table 6.4. Comparative predictive metrics of the whole test subset for the output $ZC4 + D1$ in Column T-01, using expanded input variables. ....	112
Table 6.5. Comparative predictive metrics of the test subset for the output $ZC4 + D1$ , in concentrations lower than 0.01 kg/kg, using expanded input variables. ....	112
Table 6.6. Predictive metrics for other important constituents of the distillate stream in Column T-01 (AnTSbe method). ....	113
Table 6.7. Description of the main output of interest in T-02, the mass ratio of propylene in the distillate stream ( $ZC3 - D2$ ). ....	114
Table 6.8. Description of the 12 available model inputs for Column T-02. ....	114
Table 6.9. Comparative predictive metrics of the test subset for the output $ZC3 - D2$ in Column T-02, using expanded input variables. ....	114
Table 6.10. Description of the main output of interest in T-03, the mass ratio of propane in the distillate stream ( $ZC3 + D3$ ). ....	115
Table 6.11. Description of the 18 available model inputs for Column T-03. ....	116
Table 6.12. Comparative metrics for the prediction of the propane impurity in the distillate stream of Column T-03, $ZC3 + D3$ , for propane concentrations lower than 0.01 kg/kg, for the test subset. ....	117



# Capítulo 1 – Introdução

## 1.1 Motivação

Na sociedade do século XXI, o uso de modelagem preditiva permeia todas as áreas do conhecimento. Criar um modelo com a capacidade de quantificar variáveis de interesse, classificar ocorrências ou até mesmo prever resultados futuros através do estudo de dados são as bases de muitas das evoluções tecnológicas das últimas décadas. Desde detectar câncer (EINIPOUR; CORRESPONDING, 2011; SUN et al., 2019), prever comportamento de mídias sociais (MORO; RITA; VALA, 2016), até regular produções industriais (BELTRAMO; KLOCKE; HITZMANN, 2019), novos algoritmos de *machine learning* são desenvolvidos constantemente para prever informações que seriam muito custosas, invasivas, ou impossíveis de medir diretamente.

Em uma escala global, os sistemas industriais e de análise de comportamento social se tornam cada vez mais competitivos. A chamada Quarta Revolução Industrial se baseia no conceito de retirar o componente humano das tomadas de decisões que possam ser substituídas por inteligência artificial. Além disso, a interconexão de máquinas e sistemas, o desenvolvimento de sensores inteligentes e o rastreamento e armazenamento de dados tornam possíveis a automação de diversos processos que seriam impraticáveis em décadas passadas (LU, 2017). Em grandes áreas econômicas como a indústria do petróleo e de alimentos, os processos avançados de controle possibilitam uma vantagem econômica inegável, e é provável que empresas que não se adaptarem a esta evolução sejam, aos poucos, excluídas do mercado.

No cenário atual, muitas das informações mais relevantes para o controle de processos da indústria de transformação (como concentração de espécies de interesse, ponto de ebulição de solventes, viscosidade, lubrificidade de óleos, teor de proteínas em alimentos, concentração de biomassa em sistemas biológicos, pureza de fármacos, etc.) ainda são adquiridas unicamente por meio de testes laboratoriais, que são custosos, destrutivos, demorados e requerem pessoal e instalações especializadas. Um possível caminho para a automação destes sistemas pode se dar na forma de sensores que capturem informações auxiliares de fácil aplicação, que por sua vez possam ser transformadas em inferidores

virtuais das variáveis de difícil aferição direta. A abordagem de criar um modelo que relacione os dados de entrada com as saídas de interesse, sem a necessária compreensão dos mecanismos físicos envolvidos, é denominada modelagem empírica.

A evolução tecnológica dos sistemas de coleta e armazenamento de dados faz com que diariamente uma quantidade considerável de informação seja capturada (por exemplo, na forma de historiadores de processo e imagens em alta resolução). Porém, tão importante quanto a coleta dos dados, são os mecanismos para extrair a informação útil dos mesmos. Uma grande quantidade de variáveis geralmente está associada a um alto nível de ruído, de colinearidade e de possíveis variáveis irrelevantes ou redundantes (TANG; ALELYANI; LIU, 2014). Para resolver esses problemas, são empregadas técnicas de extração e seleção de variáveis, conseguindo descrever o sistema de maneira eficiente em um espaço reduzido.

Assim, ao ansiar por sistemas de produção mais automatizados e eficientes, surge a necessidade de estudos que combinam a escolha de sensores hábeis com o desenvolvimento de metodologias de seleção de variáveis capazes de otimizar o ajuste de modelos empíricos preditivos para as saídas de interesse.

Neste contexto, a espectroscopia por fluorescência apresenta diversas características que propiciam sua aplicação como sensor útil em sistemas químicos e biológicos. Diversas moléculas de interesse nestes processos são fluoróforos naturais, como proteínas, hidrocarbonetos poliaromáticos, compostos heterocíclicos e alifáticos altamente insaturados. A aquisição dos espectros é rápida e não demanda manipulação obrigatória das amostras, podendo ser utilizada diretamente em linha para acompanhamento do sistema em tempo real. Por se tratar de uma técnica óptica, seu uso é não-invasivo e não-destrutivo, podendo ser interfaceada no sistema através de janelas de quartzo, o que a torna viável para ambientes com risco de explosão/incêndio ou cultivo biológico. De um único espectro pode-se prever uma gama de informações sobre características e constituintes do meio. Em especial, é uma técnica sensível, capaz de identificar espécies orgânicas e inorgânicas em quantidades traço e permite, muitas vezes, determinação qualitativa e quantitativa de substâncias contidas na amostra (SKOOG; HOLLER; CROUCH, 2009).

Em contraponto às vantagens apresentadas, a otimização de modelos empíricos baseados em dados de fluorescência não é de maneira alguma trivial. Dependendo da resolução, as matrizes excitação-emissão de fluorescência (EEM) podem conter entre centenas e milhares de pares excitação/emissão, em uma distribuição altamente colinear e contendo informações que podem não ter relação com a variável ou propriedade de interesse a qual se pretende prever. Por esses motivos, as estratégias de seleção de variáveis propostas precisam ter um bom equilíbrio entre os mecanismos de prospecção e exploração, conseguindo vagar entre a enormidade de combinações possíveis, mas evoluindo em direção aos modelos mais representativos. Neste âmbito, metodologias baseadas em meta-heurísticas, como as do tipo Colônia de Formigas (ACO), são reconhecidas na literatura como eficientes para lidar com problemas complexos de análise combinatória, ao exemplo da resolução de problemas do tipo Caixeiro Viajante, no qual foram primeiramente apresentadas (DORIGO; GAMBARELLA, 1997). Os algoritmos ACO já foram aplicados previamente em nosso grupo de pesquisa para otimizar com sucesso

modelos baseados tanto em dados de espectroscopia por fluorescência, quanto para outras técnicas espectrométricas (RANZAN et al., 2011, 2014, 2015, 2017). Porém, a grande capacidade adaptativa destes algoritmos abre possibilidades para diversas melhorias que buscam resolver desafios inerentes aos métodos estocásticos iterativos, como estagnação precoce, ciclagem de subgrupos e problemas de generalização. Dentre as adaptações sugeridas, a hibridização com outras técnicas de otimização pode ser vista como caminho natural para solução dos problemas supracitados (BHATTACHARYYA, 2018). A proposta de evolução das meta-heurísticas visa, assim, não apenas ajustar melhores modelos preditivos, mas garantir uma maior eficiência dos algoritmos, com resultados mais consistentes e maior aplicabilidade a diferentes casos de estudo.

Em uma outra abordagem ao ajuste de modelos caixa-preta, é inegável que as técnicas de aprendizado de máquina mais utilizadas e pesquisadas na última década se referem ao uso de Redes Neurais Artificiais (ANN). Modelos baseados nas mais diversas configurações de ANN são o estado-da-arte em múltiplos campos, especialmente no que diz respeito ao reconhecimento de padrões (ABIODUN et al., 2018; JES et al., 2019). Majoritariamente aplicadas para tratamento e classificação de imagens, as Redes Neurais Convolucionais (CNN) são, possivelmente, as mais populares e efetivas arquiteturas atuais de aprendizado profundo (*deep learning*) (RUIZ-DEL-SOLAR; LONCOMILLA; SOTO, 2018). Seu uso, porém, ainda é incipiente na construção de modelos quimiométricos. Isso pode ser ligado ao fato de que tais redes necessitam de uma vasta disponibilidade de dados de segunda ordem para seu treinamento, que são amplamente acessíveis no caso de imagens, mas muitas vezes de difícil aquisição no contexto de processos químicos/biológicos. No escopo deste trabalho, é promissor o emprego de CNN no ajuste de modelos preditivos, utilizando como base de dados matrizes excitação – emissão de fluorescência coletadas continuamente *in situ*. O acoplamento de um espectrofluorímetro a um sistema possibilita um influxo contínuo de EEMs, que alimentam a rede neural para qualificação e quantificação de variáveis de interesse do meio, de maneira a reduzir a dependência de análises laboratoriais ou proporcionar informações *online* para ações de controle.

Uma das limitações dos modelos empíricos é sua baixa capacidade de generalização e extrapolação. Diferentemente de modelos baseados em conceitos fenomenológicos, os modelos empíricos são totalmente dependentes dos dados de treinamento e teste, e só se pode confiar em sua capacidade preditiva dentro de um contexto semelhante ao destas amostras, como, por exemplo, dentro de uma mesma faixa de concentração. Assim, é importante um acompanhamento periódico da qualidade das predições do modelo com metodologias *off-line*, como testes laboratoriais. Porém, o tempo morto associado a estas medidas *off-line* é justamente o que se espera superar com o uso dos inferidos virtuais, e é possível que o sistema passe longos períodos sendo monitorado virtualmente antes que se obtenha confirmação externa de normalidade. Neste meio tempo é concebível que o sistema seja alimentado por entradas que divirjam significativamente do conjunto de calibração. Tomar ações baseadas em predições indignas de confiança pode acarretar inúmeros prejuízos ao sistema. Assim, surge a necessidade de uma metodologia para se qualificar de forma não-supervisionada dados de entrada, com a intenção de avaliar se os mesmos se enquadram nas características dos conjuntos previamente utilizados para calibração dos preditores, e se suas predições devem ser confiadas.

Nesta conjectura, as redes neurais do tipo *AutoEncoder* (AE) (CHARTE et al., 2018) apresentam potencial para triagem e qualificação de novos dados de entrada. As redes AE utilizam uma estrutura simétrica para primeiramente transformar as entradas em um novo

espaço, chamado *encodado*. Nele, características dos dados são explicitadas, em uma forma similar ao espaço criado na Análise de Componentes Principais (PCA), mas com a capacidade de captura de informação não linear (ALMOTIRI; ELLEITHY; ELLEITHY, 2017; JES et al., 2019). O espaço encodado é então *decodificado*, buscando reproduzir as entradas originais. Assim, os parâmetros da rede são otimizados para diminuir o erro de reconstrução entre as saídas e as entradas (que almejam serem as mesmas). Uma rede AE ajustada é capaz de quantificar quanto uma nova entrada é semelhante aquelas com as quais ela foi treinada, com base no seu erro de reconstrução. Aplicações de triagem e reconhecimento de dados anormais podem ser vistas como no caso de identificação de fraudes em cartão de crédito (MISRA et al., 2020; ZOU; ZHANG; JIANG, 2019), na detecção de anomalias em dados de satélites (SAKURADA; YAIRI, 2014) e no diagnóstico computacional de diversas patologias (AMARBAYASGALAN; JARGALSAIKHAN; RYU, 2018). É de interesse avaliar a capacidade de uma rede AE treinada com dados de espectroscopia por fluorescência em detectar falhas ou desvios em novas leituras espectrais, e quantificar a confiabilidade nestes dados.

Com o objetivo de aprofundar os conhecimentos em seleção de variáveis e otimização de modelos empíricos preditivos, neste estudo adaptações e metodologias serão propostas, como a hibridização de algoritmos meta-heurísticos e modificações nos mecanismos de exploração e tratamento dos dados de entrada, assim como o desenvolvimento de técnicas baseadas em redes neurais para ajuste de modelos preditivos e reconhecimento de padrões. As metodologias apresentadas serão então avaliadas utilizando, majoritariamente, dados de espectroscopia por fluorescência para o desenvolvimento de técnicas alternativas às convencionais para predição de variáveis de interesse em diferentes estudos de caso.

## 1.2 Objetivos e Estruturação do Trabalho

O objetivo principal deste trabalho trata sobre o desenvolvimento e evolução de metodologias baseadas em meta-heurísticas para seleção de variáveis e otimização de modelos empíricos, incluindo todas as facetas de tratamento de dados necessários para tal. Além disso, serão propostas metodologias baseadas em redes neurais para predição de variáveis de processo e desenvolvimento de uma rotina de triagem capaz de qualificar e detectar inconformidades na base de dados futuros. Unindo as metodologias se encontra o uso de matrizes de dados obtidos através de espectroscopia por fluorescência bidimensional nos estudos de caso.

Os objetivos específicos ( $O_n$ ) são apresentados na sequência:

- $O_1$  Propor novas adaptações para algoritmos de seleção de variáveis baseados em meta-heurísticas.
- $O_2$  Desenvolver uma metodologia capaz de aprimorar a generalização de modelos empíricos baseados em dados de espectroscopia por fluorescência.
- $O_3$  Desenvolver uma metodologia baseada em redes neurais para triagem e identificação de dados anômalos.

- **O<sub>4</sub>** Avaliar a aplicação de redes neurais convolucionais para predição de variáveis de processos com coleta *online* de espectros.
- **O<sub>5</sub>** Avaliar o uso de espectroscopia por fluorescência para predição de variáveis de interesse que atualmente são quantificadas majoritariamente por testes laboratoriais.
- **O<sub>6</sub>** Expandir a aplicação das metodologias desenvolvidas em estudos de caso da indústria petroquímica que não envolvam dados de fluorescência.

No Capítulo 2 será apresentada uma revisão bibliográfica, contemplando a fundamentação teórica dos assuntos mais relevantes contidos neste trabalho e pertinentes ao seu entendimento.

O Capítulo 3 descreve minuciosamente o desenvolvimento do algoritmo de otimização de modelos empíricos AnTSbe. O artigo descreve as etapas iniciais de tratamento e segmentação dos dados, a metodologia proposta de seleção de variáveis baseada em uma hibridização entre meta-heurísticas Colônia de Formigas e Busca Tabu, e introduz o conceito de expansão de variáveis de entrada. A metodologia é aplicada a um estudo de caso, utilizando espectroscopia por fluorescência para prever concentração de enxofre em amostras de diesel combustível.

O Capítulo 4 apresenta um estudo sobre caracterização de perfis e predição de compostos fenólicos totais em processos de envelhecimento de cachaça em barris de amburana. Três processos de envelhecimento foram realizados em nosso departamento, em uma pesquisa que perdurou por mais de três anos. Ao longo de cada envelhecimento, amostras eram retiradas e seu espectro de fluorescência coletado. A concentração total de fenólicos em cada amostra também foi quantificada por métodos tradicionais. A metodologia proposta para predição da concentração de fenólicos é uma adaptação do algoritmo AnTSbe, e tem como principal objetivo a introdução com uma nova camada externa de seleção de variáveis, criando o conceito de par delta, um par excitação/emissão que é utilizado como referência interna em cada espectro. Os resultados deste trabalho corroboram fortemente com as contribuições do Capítulo 3.

O Capítulo 5 trata sobre o desenvolvimento de uma metodologia que utiliza o erro de reconstrução de uma rede neural do tipo autoencoder como ferramenta de triagem para amostras anômalas. No estudo, espectros por fluorescência são coletados *in situ* durante três processos fermentativos utilizando a levedura *Saccharomyces cerevisiae*. Tais dados são alimentados em dois processos paralelos: o primeiro faz uso de redes neurais residuais para ajustar um modelo de predição para etanol, glicose e biomassa do meio. O segundo utiliza os dados para criar uma rede autoencoder capaz de corretamente representar o conjunto de espectros com erros de reconstrução aceitáveis. Para simular novas amostras, alguns pontos do conjunto de dados original são dopados com erros, e o novo erro de reconstrução da rede autoencoder treinada é utilizado para qualificar as amostras anômalas. Este estudo foi iniciado sob a tutela do Prof. Dr. Bernd Hitzmann, em parceria com a Universidade de Hohenheim – Alemanha, durante o período de doutorado-sanduíche do autor.

O Capítulo 6 relata a expansão da aplicação da metodologia AnTSbe para o desenvolvimento de sensores virtuais para predição de contaminantes em uma Unidade de Separação Propano/Propeno, com o uso de dados industriais (e não de fluorescência). Neste caso, a base de dados utilizada para construção dos modelos preditivos parte da

simulação de três colunas de separação em série, inspiradas por uma unidade real de separação em operação. No estudo, a metodologia de seleção de variáveis ANTSbe é comparada com a aplicação direta de técnicas de regularização de modelos empíricos (Regressão *Ridge* e *LassoLars*). As três colunas de separação são tratadas individualmente, e as concentrações dos compostos de interesse são preditas tanto nas correntes de topo quanto de fundo, caracterizando o perfil de concentrações das colunas. Em especial, se foca na predição das impurezas de cada coluna, que devem ser mantidas ínfimas.

Por fim, o Capítulo 7 traz as conclusões e considerações gerais acerca do trabalho.

### 1.3 Produção Científica e Contribuições Paralelas

Os capítulos que compõem esta Tese de doutorado já foram ou serão submetidos à periódicos e congressos científicos. Ademais, durante a realização deste trabalho, o autor participou ativamente de projetos envolvendo análise de dados, que tem correlação direta com os conhecimentos que adquiriu durante a realização das pesquisas aqui expostas. Estão destacados abaixo os artigos completos publicados até a data desta defesa.

- C. Ranzan, L. Ranzan, L.F. Trierweiler, J.O. Trierweiler. *Sulfur Determination in Diesel using 2D Fluorescence Spectroscopy and Linear Models*. IFAC-PapersOnLine. 48 (2015) 415–420. (Congresso).
- L. Ranzan, C. Ranzan, L.F. Trierweiler, J.O. Trierweiler. *Classification of Diesel Fuel Using Two-Dimensional Fluorescence Spectroscopy*. Energy & Fuels. 31 (2017) 8942–8950.
- J.A. Sebben, J. da Silveira Espindola, L. Ranzan, N. Fernandes de Moura, L.F. Trierweiler, J.O. Trierweiler. *Development of a quantitative approach using Raman spectroscopy for carotenoids determination in processed sweet potato*. Food Chem. 245 (2018) 1224–1231.
- P.V.J.L. Santos, L. Ranzan, M. Farenzena, J.O. Trierweiler. *K-Rank: An Evolution of Y-Rank for Multiple Solutions Problem*. Brazilian J. Chem. Eng. 36 (2019) 409–419.
- D.G. Carvalho, L. Ranzan, L.F. Trierweiler, J.O. Trierweiler. *Determination of the Concentration of Total Phenolic Compounds in Aged Cachaça Using Two-Dimensional Fluorescence and Mid-Infrared Spectroscopy*. Food Chem. 329 (2020) 127142.
- Weber, C.T., Ranzan, L., Liesegang, L.L.M., L.F. Trierweiler, J.O. Trierweiler. *A circular economy model for ethanol and alcohol-based hand sanitizer from sweet potato waste in the context of COVID-19*. Brazilian Journal of Operations & Production Management, Vol. 17 (2020), No. 03, e20201025.
- D.G. Carvalho, L. Ranzan, R.A. Jacques, L.F. Trierweiler, J.O. Trierweiler. *Analysis of total phenolic compounds and caffeine in teas using variable selection approach with two-dimensional fluorescence and infrared spectroscopy*. Microchemical Journal 169(2) (2021), 106570.

Na sequência, estão listados os Trabalhos de Conclusão de Curso em que o autor participou como coorientador, vários dos quais aplicam partes dos códigos e metodologias que o autor desenvolveu:

- Zanatta, Felipe Georg (2016). Potencial econômico e aplicação do bio-óleo da pirólise rápida de casca de arroz como pesticida. Orientadores: J. Trierweiler, L. F. Trierweiler. Coorientador: L. Ranzan. Trabalho de Diplomação em Engenharia Química, Departamento de Engenharia Química – UFRGS.
- Boff, Luiz Gustavo Fracalossi (2018). Estudo do potencial da utilização da espectroscopia para classificação de cervejas. Orientadores: J. Trierweiler, L. F. Trierweiler. Coorientador: L. Ranzan. Trabalho de Diplomação em Engenharia Química, Departamento de Engenharia Química – UFRGS.
- Trucolo, Ana Clara Fernandes (2018). Desenvolvimento de um analisador virtual para determinar o teor de umidade de protreínas isoladas de soja. Orientadores: J. Trierweiler, L. F. Trierweiler. Coorientador: L. Ranzan. Trabalho de Diplomação em Engenharia Química, Departamento de Engenharia Química – UFRGS.
- Freitas, Lucas do Amaral (2018). Estudo da viabilidade técnica na utilização de espectroscopia Raman na caracterização online de solos. Orientador: C. Ranzan. Coorientador: L. Ranzan. Trabalho De Diplomação Em Engenharia Agroindustrial - Agroquímica, Escola De Química e Alimentos – FURG.
- Gomes, Alexandre Rodrigues Conill (2019). Aplicação de análise multivariada em dados de agrotóxicos na água para consumo humano do Brasil de 2014 a 2018. Orientador: J. Trierweiler. Coorientadores: L. Ranzan, L. B. Zini. Trabalho de Diplomação em Engenharia Química, Departamento de Engenharia Química – UFRGS.
- Liesegang, Luciano Luís Menz (2020). Produção de álcool gel em biorrefinarias descentralizadas de etanol a partir de batata-doce. Orientadores: J. Trierweiler, L. F. Trierweiler. Coorientador: L. Ranzan. Trabalho de Diplomação em Engenharia Química, Departamento de Engenharia Química – UFRGS.

#### **1.4 Contribuições**

As principais contribuições desta tese podem ser listadas como:

- **C<sub>1</sub>** AnTSbe: metodologia híbrida ACO – TS para seleção de variáveis e ajuste de modelos com expansão de bases. Implementada em software livre Python e que pode ser aplicada em diferentes bases de dados, e não só para fluorescência.
- **C<sub>2</sub>** Adapted AnTSbe: Desenvolvimento da metodologia DeltaPair para aprimorar a generalização de modelos com expansão de base.
- **C<sub>3</sub>** Mapeamento espectral e captura do perfil fenólico em diferentes processos de envelhecimento de cachaças.
- **C<sub>4</sub>** Calibração de redes neurais convolucionais residuais com matrizes excitação – emissão de fluorescência coletados continuamente e aplicação bem-sucedida dos

modelos ajustados para predição de importantes constituintes em sistemas fermentativos.

- **C<sub>5</sub>** Desenvolvimento da metodologia de triagem e identificação de dados anômalos baseado no erro de reconstrução de redes *autoencoder*.
- **C<sub>6</sub>** Desenvolvimento, nos estudos de caso, de novas metodologias quantitativas baseadas em fluorescência e dados industriais, com potencial para ferramenta de projeto para sensores de processo.

## 1.5 Resumo Gráfico

A Figura 1.1 apresenta um resumo gráfico que correlaciona os objetivos com as contribuições específicas de cada um dos capítulos. Assim, é possível ter uma visão geral sobre a estrutura e interligações entre as diferentes etapas do trabalho.

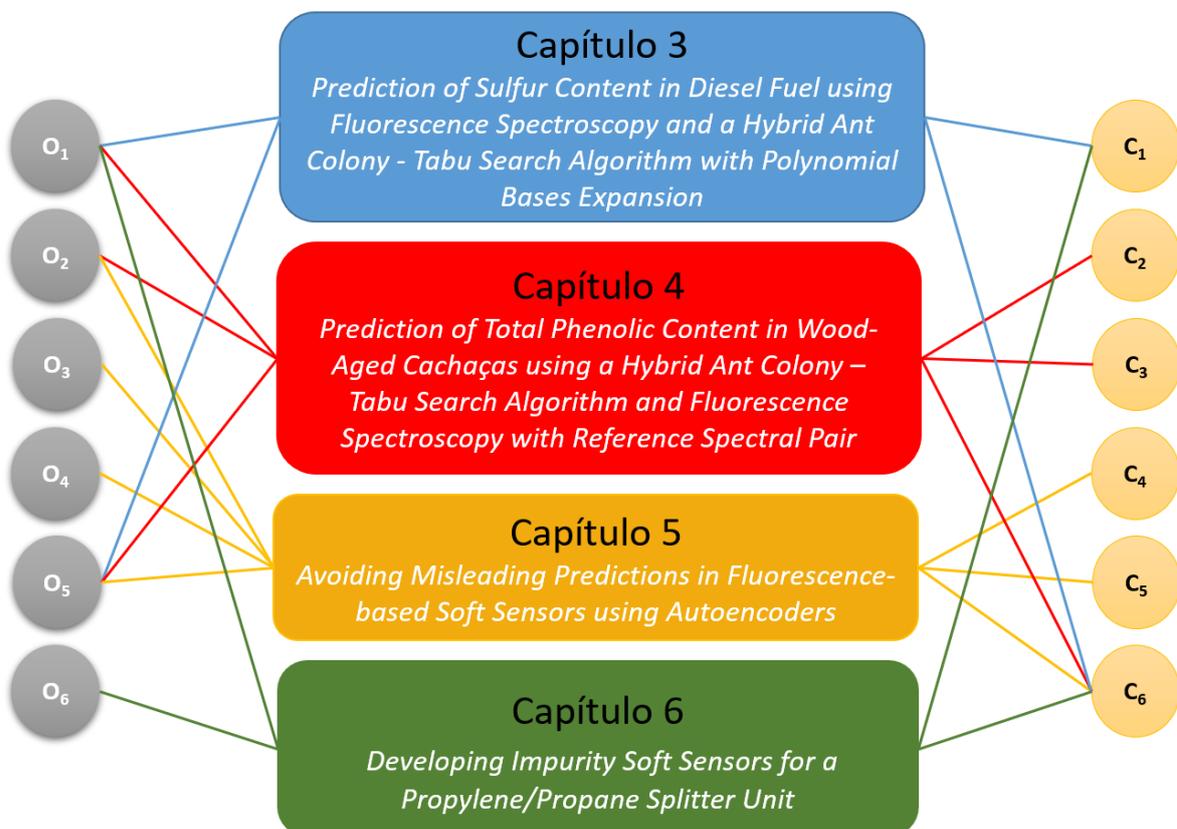


Figura 1.1. Resumo gráfico, indicando as correlações entre objetivos, capítulos e contribuições do trabalho.

## 1.6 Referências

ABIODUN, O. I. et al. **State-of-the-art in artificial neural network applications: A survey** *Heliyon* Elsevier Ltd, , 1 nov. 2018.

ALMOTIRI, J.; ELLEITHY, K.; ELLEITHY, A. Comparison of Autoencoder and Principal Component Analysis Followed by Neural Network for E-Learning Using Handwritten Recognition. **2017 IEEE Long Island Systems, Applications and Technology Conference, LISAT 2017**, n. May, 2017.

AMARBAYASGALAN, T.; JARGALSAIKHAN, B.; RYU, K. Unsupervised Novelty Detection Using Deep Autoencoders with Density Based Clustering. **Applied Sciences**, v. 8, 2018.

BELTRAMO, T.; KLOCKE, M.; HITZMANN, B. Prediction of the biogas production using GA and ACO input features selection method for ANN model. **Information Processing in Agriculture**, v. 6, n. 3, p. 349–356, 2019.

BHATTACHARYYA, S. **Hybrid Metaheuristics**. [s.l.] WORLD SCIENTIFIC, 2018. v. 84

CHARTE, D. et al. A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines. **Information Fusion**, v. 44, n. December 2017, p. 78–96, 2018.

DORIGO, M.; GAMBARDELLA, L. M. Ant colonies for the travelling salesman problem. **Biosystems**, v. 43, n. 2, p. 73–81, 1997.

EINIPOUR, A.; CORRESPONDING. A fuzzy-ACO method for detect breast cancer. **Global Journal of Health Science**, v. 3, 2011.

JES, F. et al. Deep Convolutional Autoencoders vs PCA in a Highly-Unbalanced Parkinson 's Disease Dataset : A DaTSCAN Study. **Springer Nature**, p. 47–56, 2019.

LU, Y. **Industry 4.0: A survey on technologies, applications and open research issues** *Journal of Industrial Information Integration* Elsevier B.V., , 1 jun. 2017.

MISRA, S. et al. **An Autoencoder Based Model for Detecting Fraudulent Credit Card Transaction**. *Procedia Computer Science*. **Anais...Elsevier B.V.**, 1 jan. 2020

MORO, S.; RITA, P.; VALA, B. Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. **Journal of Business Research**, v. 69, n. 9, p. 3341–3351, 2016.

RANZAN, C. et al. **Validação com Dados de Espectroscopia Fluorescente 2D de Modelo Dinâmico para Fermentações Batelada de Saccharomyces cerevisiae** Caxias do Sul, RS, Brasil18 Simpósio Nacional de Bioprocessos, , 2011.

RANZAN, C. et al. Wheat flour characterization using NIR and spectral filter based on Ant Colony Optimization. **Chemometrics and Intelligent Laboratory Systems**, v. 132, n. 0, p. 133–140, 2014.

RANZAN, C. et al. Sulfur Determination in Diesel using 2D Fluorescence Spectroscopy and Linear Models. **IFAC-PapersOnLine**, v. 48, n. 8, p. 415–420, 2015.

RANZAN, L. et al. Classification of Diesel Fuel Using Two-Dimensional Fluorescence Spectroscopy. **Energy & Fuels**, v. 31, n. 9, p. 8942–8950, 2017.

RUIZ-DEL-SOLAR, J.; LONCOMILLA, P.; SOTO, N. A Survey on Deep Learning Methods for Robot Vision. 28 mar. 2018.

SAKURADA, M.; YAIRI, T. Anomaly detection using autoencoders with nonlinear dimensionality reduction. **Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis**, p. 4, 2014.

SKOOG, D. A.; HOLLER, J. F.; CROUCH, S. R. **Princípios de Análise Instrumental**. 6ª edição ed. [s.l.] Bookman, 2009.

SUN, L. et al. A Hybrid Gene Selection Method Based on ReliefF and Ant Colony Optimization Algorithm for Tumor Classification. **Scientific Reports**, v. 9, n. 1, 2019.

TANG, J.; ALELYANI, S.; LIU, H. Feature selection for classification: A review. In: **Data Classification: Algorithms and Applications**. [s.l: s.n.]. p. 37–64.

ZOU, J.; ZHANG, J.; JIANG, P. Credit Card Fraud Detection Using Autoencoder Neural Network. 30 ago. 2019.

## Capítulo 2 – Revisão Bibliográfica

Neste capítulo consta o levantamento bibliográfico acerca dos principais assuntos constituintes desta tese. Por facilidade, o mesmo se encontra dividido em subseções. Primeiramente, serão abordadas as definições de modelos empíricos e suas peculiaridades, como o uso de normas de regularização. Na sequência, serão apresentados conceitos e metodologias aplicadas na seleção de variáveis e otimização de modelos. A seguir, uma visão geral sobre algoritmos do tipo redes neurais, com a definição de tipos específicos de redes, e os parâmetros das mesmas. Por fim, discorreremos sobre os fundamentos da espectroscopia por fluorescência, técnica na qual se baseiam os estudos de caso desta tese.

### 2.1 Modelos Empíricos

Modelos empíricos são aqueles baseados inteiramente em dados experimentais. Em Engenharia, são comumente denominados de modelos caixa-preta, descrevendo um sistema fechado de complexidade potencialmente alta, mas que se limita a medidas das relações entre estímulos de entrada e respostas de saída, sem que seja levada em consideração a estrutura interna (GUIDOTTI et al., 2018). Uma representação visual pode ser vista na Figura 2.1. Os modelos empíricos estabelecem uma relação funcional entre os sistemas de entrada e saída, onde os parâmetros do mesmo não necessitam ter significância fenomenológica (como coeficientes de transferência de massa e calor ou cinética de reação), mas conseguem, de maneira eficiente, representar as tendências de comportamento do processo (ZHANG, 2010).



Figura 2.1. Representação visual de um modelo caixa-preta.

As principais vantagens no uso de modelos empíricos se dão pela sua relativa simplicidade. Com uma base de dados suficientemente grande, os modelos podem correlacionar de maneira acurada entradas e saídas das quais os reais mecanismos

fenomenológicos não são conhecidos ou são muito complexos. Por se tratarem normalmente de relações matemáticas simples, os modelos podem ser facilmente ajustados a novos conjuntos de dados. Considerando a atual performance dos sistemas computacionais, são uma alternativa barata e rápida, tanto para uso quanto para implementação (HEIYANTHUDUWAGE; MOUNOURY; KOVACEVIC, 2011).

Porém, as vantagens apresentadas também carregam desvantagens importantes. Primeiramente, a base de dados experimental tem papel fundamental: a mesma precisa ser não somente vasta, como de qualidade. Isso quer dizer que muitos dados precisam ser coletados, com a maior variabilidade possível no processo, contemplando as variações de comportamento que se espera capturar com o modelo. Esta alta necessidade por dados pode ser proibitiva para sistemas de difícil coleta e/ou experimentação. O modelo ajustado geralmente tem baixo poder de generalização, perdendo potencial preditivo ao se afastar das condições operacionais similares as dos dados de treinamento. Assim, estratégias de retroalimentação e atualização dos modelos são importantes. Ademais, considerando que o modelo não é baseado em relações físicas, as equações e algoritmos finais nem sempre provêm compreensão sobre o processo (GUIDOTTI et al., 2018).

Supondo que nossos dados reais surjam de um modelo estatístico na forma da Equação 2.1:

$$Y = f(X) + \varepsilon \quad (2.1)$$

que relaciona o sistema observado entradas – saídas  $(X,Y)$  através da função  $f$  e do erro randômico  $\varepsilon$ , independente de  $X$  (que captura erros de medição de  $X$  e variáveis não mensuradas que contribuem para  $Y$ ). Assim, o objetivo da modelagem empírica é encontrar uma aproximação útil  $\hat{f}(x)$  para função  $f(x)$ , capaz de representar a relação preditiva entre as observações de entrada e saídas. Para tal, o sistema em estudo é observado e um conjunto de calibração contendo tanto as entradas quanto as saídas na forma  $\mathcal{T} = (x_i, y_i), i = 1, \dots, N$  é agrupado. Os valores de entrada também são alimentados em um sistema artificial (algoritmo de aprendizado), que produz uma saída  $\hat{f}(x_i)$  em resposta as entradas. O algoritmo de aprendizado tem a capacidade de alterar a sua relação entrada/saída para responder as diferenças  $y_i - \hat{f}(x_i)$ , entre as saídas reais e as saídas preditas. Se espera que ao fim do processo de aprendizagem, as saídas reais e artificiais sejam suficientemente parecidas para representar de maneira útil todos os conjuntos de dados de entrada que possam ser encontrados na prática (HASTIE, TREVOR, TIBSHIRANI, ROBERT, FRIEDMAN, 2009).

São diversas as abordagens para definição da função de aproximação  $\hat{f}(x)$ , em sua maioria associando as variáveis de entrada a uma série de parâmetros  $\theta$  que podem ser modificados para se adequar aos dados disponíveis. Entre as mais aplicadas, podemos citar o uso de funções polinomiais (OSTERTAGOVÁA, 2012), de regressões baseadas em transformações do espaço (como as regressões por mínimos quadrados parciais (PLS) (GELADI; KOWALSKI, 1986) e as regressões de componentes principais (PCR) (HEMMATEENEJAD; MIRI; ELYASI, 2012)), e uso de funções não lineares, incluindo modelos do tipo redes neurais (BASHEER; HAJMEER, 2000). Porém, as funções lineares ainda são umas das mais versáteis aproximações utilizadas (YU; YAO, 2017).

### 2.1.1 Regressões Lineares e Soma dos Quadrados dos Resíduos

Considerando um vetor de entradas  $X^T = (x_1, x_2, \dots, x_p)$ , e com a intenção de prever uma saída real  $Y$ , o modelo de regressão linear apresenta a forma da Equação 2.2.

$$f(X) = \beta_0 + \sum_{j=1}^p x_j \beta_j \quad (2.2)$$

Onde  $\beta_j$  são parâmetros desconhecidos e  $p$  é a dimensão do subespaço das variáveis de entrada. Desta forma, as variáveis  $x_j$  podem ser originárias de diversas fontes:

- Variáveis quantitativas;
- Transformações das variáveis quantitativas, como logaritmo, raiz quadrada ou potências;
- Representação polinomial das entradas, como  $x_2 = x_1^2$  e  $x_3 = x_1^3$ ;
- Interação entre as variáveis, como  $x_3 = x_1 \times x_2$ .

Neste caso, independente da fonte de  $x_j$ , o modelo é linear nos parâmetros.

A maneira mais tradicional de estimar os parâmetros  $\beta$  é através do método dos Mínimos Quadrados, onde os coeficientes  $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$  são escolhidos a fim de minimizar a soma dos quadrados dos resíduos (RSS – *Residual Sum of Squares*)

$$RSS(\beta) = \sum_{i=1}^N (y_i - f(x_i))^2$$

$$RSS(\beta) = \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 \quad (2.3)$$

onde  $(x_i, y_i)$  são as observações de entrada e saída do conjunto de calibração.

Para minimizar Eq. 2.3 podemos escrever o RSS como

$$RSS(\beta) = (y - X\beta)^T (y - X\beta) \quad (2.4)$$

Que é uma função quadrática nos  $p + 1$  parâmetros. Derivando em relação a  $\beta$ , obtemos

$$\frac{\partial RSS}{\partial \beta} = -2X^T (y - X\beta) \quad (2.5)$$

$$\frac{\partial^2 RSS}{\partial \beta \partial \beta^T} = 2X^T X$$

Assumindo que  $X$  tem posto de colunas completo, e, assim,  $X^T X$  é positivo definido, a primeira derivada pode ser zerada

$$X^T (y - X\beta) = 0 \quad (2.6)$$

e a solução única é obtida

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2.7)$$

O modelo de predição ajustado é assim representado por

$$\hat{f}(x_i) = X\hat{\beta} = X(X^T X)^{-1} X^T y \quad (2.8)$$

Por se tratar de uma resposta fechada, única e de fácil obtenção (dependendo basicamente de inversão e transposição de matrizes), a abordagem de modelos lineares e ajuste por mínimos quadrados é conveniente desde antes do uso de computadores.

### 2.1.2 Normas de Regularização

Quando variáveis correlacionadas estão presentes em modelos lineares, os coeficientes podem acabar mal determinados e apresentar alta variância. Um coeficiente positivo grande em uma variável pode ser cancelado por um coeficiente negativo similar em uma variável correlacionada. Assim, técnicas de penalização foram desenvolvidas para encolher ou remover os coeficientes dos modelos.

A Regressão Ridge, também conhecida como norma  $l_2$ , reduz os coeficientes da regressão impondo uma penalização quadrática referente ao seu tamanho,

$$\hat{\beta}^{ridge} = \underset{\beta}{argmin} \left\{ \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (2.9)$$

onde  $\lambda \geq 0$  é um parâmetro referente a força da penalização. Com alguma manipulação matemática, a solução para a regressão Ridge pode ser escrita como

$$\hat{\beta}^{ridge} = (X^T X + \lambda I)^{-1} X^T y \quad (2.10)$$

onde  $I$  é a matriz identidade  $p \times p$ . A seleção da penalização quadrática  $\beta^T \beta$  faz com que a solução da regressão Ridge seja novamente uma função linear de  $y$ . Esta solução adiciona uma constante positiva a diagonal  $X^T X$  antes da inversão, fazendo com que o problema se torne não singular, mesmo quando a matriz não possui posto completo. Esta foi a principal característica estudada pelos desenvolvedores da metodologia (HOERL;

KENNARD, 1970). Por sua característica quadrática, a norma  $l_2$  faz com que os coeficientes tendam a zero, mas não podem ser explicitamente zero.

A metodologia Lasso, conhecida como norma  $l_1$ , é definida em Eq 2.11

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (2.11)$$

Neste caso, a penalidade  $l_2$  Ridge  $\sum_1^p \beta_j^2$  é substituída pela penalidade  $l_1$  Lasso  $\sum_1^p |\beta_j|$ . Esta restrição torna as soluções não lineares em  $y_i$ , e, assim, não existe uma expressão de forma fechada como na regressão Ridge. A solução para o Lasso se torna um problema de computação quadrática, porém, existem algoritmos disponíveis que tornam a computação de todo o caminho de soluções com a variação de  $\lambda$  de custo computacional comparável aquele da regressão Ridge, como por exemplo a metodologia LAR (*Least Angle Regression* (EFRON et al., 2004)). Dada a natureza da penalização  $l_1$  os coeficientes das variáveis podem ser reduzidos a exatamente zero.

A Figura 2.2 apresenta uma comparação entre as regularizações Ridge (a direita) e Lasso (a esquerda). A RSS, em vermelho, apresenta contornos elípticos, centrados na estimativa de mínimos quadrados. A áreas em azul representam as regiões das restrições. A solução de ambos os métodos se dá no primeiro ponto de contato entre as regiões e as elipses.

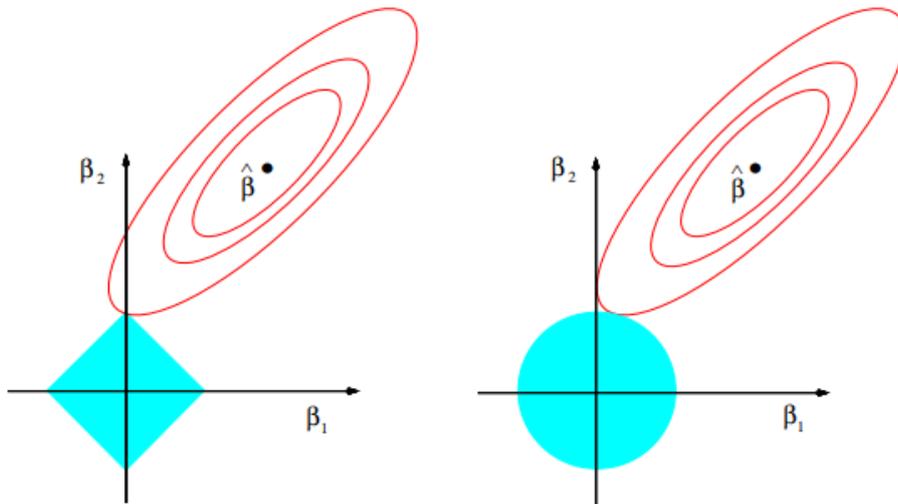


Figura 2.2. Comparação entre as penalizações Lasso (esquerda) e regressão Ridge (direita). A área em azul representa as regiões de restrição  $|\beta_1| + |\beta_2| \leq t$  e  $\beta_1^2 + \beta_2^2 \leq t^2$ , respectivamente. As elipses vermelhas representam o contorno da função erro dos mínimos quadrados. Fonte: (HASTIE, TREVOR, TIBSHIRANI, ROBERT, FRIEDMAN, 2009)

## 2.2 Otimização de Modelos e Seleção de Variáveis

Em muitas aplicações, o número de variáveis de entrada disponíveis em um sistema pode atingir a casa das centenas até os milhares (por exemplo, classificação de imagens (JOHNSON; XIE, 2013; KAVZOGLU, 2017; MOHAPATRA; PATRA; SATPATHY, 2014), dados espectrais (RANZAN et al., 2014, 2017; SEBBEN et al., 2018) e processos industriais (SILVA; SECCHI, 2018; ZHENG et al., 2012)). Todavia, esta abundância de dados muitas vezes não apenas não apresenta vantagens para prever o sistema, como ativamente causam um detrimento dos modelos preditivos. Estes dados podem estar associados a um alto nível de ruído, de colinearidade, e repletos de variáveis redundantes e/ou irrelevantes para a propriedade que se pretende inferir (TANG; ALELYANI; LIU, 2014).

Diversos são os estudos focados no desenvolvimento de técnicas capazes de extrair a informação válida das variáveis de entrada de um sistema, reduzindo o ruído e as entradas sem serventia, mas mantendo todas as relações e padrões importantes do conjunto de dados (BALABIN; SMIRNOV, 2011; BOLÓN-CANEDO; SÁNCHEZ-MAROÑO; ALONSO-BETANZOS, 2013; CHANDRASHEKAR; SAHIN, 2014; PES, 2019; XIAOBO et al., 2010). Este processo de seleção de variáveis é considerado essencial para a otimização de modelos empíricos, sendo muitas vezes a diferença entre modelos falhos de modelos efetivos. Amplamente, estas técnicas podem ser categorizadas em Extração de Características (*Feature Extraction*) e Seleção de Variáveis (*Feature Selection*).

A abordagem de Extração de Características visa projetar as variáveis em um novo espaço com menor dimensionalidade, através da combinação do espaço original das mesmas. O novo espaço tem a intenção de ser informativo, não redundante e facilitar o aprendizado e generalizações a partir dos dados. Muitas vezes, a análise deste novo espaço facilita a interpretação e a busca por padrões nos dados. Os principais exemplos destas técnicas incluem Análise de Componentes Principais (*Principal Component Analysis – PCA*) (CAMACHO; PICÓ; FERRER, 2010; JACKSON, 1991; JOLLIFFE, 1986; TIPPING; BISHOP, 1999), Análise Discriminante Linear (*Linear Discriminant Analysis – LDA*) (SILVA et al., 2016), Regressão por Mínimos Quadrados Parciais (*Partial Least Squares – PLS*) (CHI et al., 2014; FILZMOSER; TODOROV, 2011; GELADI; KOWALSKI, 1986; WOLD; SJÖSTRÖM; ERIKSSON, 2001; XU et al., 2007), e Análise de Fatores Paralelos (PARAFAC) (MURPHY et al., 2013; SILVA et al., 2019).

As técnicas de PCA e regressão PLS são amplamente utilizadas para análise de dados (GODOY; VEGA; MARCHETTI, 2014). Estas técnicas são de fácil aplicação, estando implementadas em praticamente todos os softwares matemáticos modernos. Seu custo computacional também é irrisório, sendo sua aplicação praticamente instantânea. Na PCA, a matriz original dos dados  $\mathbf{X}$  é decomposta como produto de duas outras matrizes,  $\mathbf{T}$  (denominada *scores*) e  $\mathbf{P}$  (denominada *loadings*), somada de uma matriz de residuais  $\mathbf{E}$

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} = \sum_{a=1}^A t_a \mathbf{p}'_a + \mathbf{E} \quad (2.12)$$

O número de colunas da matriz  $\mathbf{T}$  e linhas da matriz  $\mathbf{P}$  correspondem a quantidade componentes principais (PCs), e onde cada  $t_a$  representa um PC. Os PCs são ortogonais entre si, e equivalem ao novo espaço formado a partir da combinação linear do espaço original. De acordo com as características da decomposição, a direção ortogonal de cada

PCs é definida a fim de capturar a máxima variância contida nos dados. Assim, a quantidade de informação independente contida nos dados é acumulada em ordem decrescente em cada PC, até o ponto em que a adição de mais PCs na matriz não mais adiciona informação (variância). Desta forma é realizada a redução do espaço, truncando o número de PCs naquele capaz de capturar suficientemente a variância original dos dados.

A PCA é uma análise não supervisionada, ou seja, considera apenas os dados de entrada sem nenhuma relação com as saídas do processo. Já a regressão PLS consiste de duas correlações externas que unem os resultados das decomposições simultâneas tanto da matriz que contém as variáveis de entrada  $\mathbf{X}$  quanto a matriz contendo as saídas  $\mathbf{Y}$ , e uma correlação interna entre as matrizes de scores de ambas decomposições. O objetivo da modelagem PLS é minimizar a norma de  $\mathbf{F}$  ao mesmo tempo que maximiza a covariância entre  $\mathbf{X}$  e  $\mathbf{Y}$ . Esta relação interna é uma regressão linear múltipla entre as matrizes de scores  $\mathbf{U}$  e  $\mathbf{T}$  em que  $\mathbf{B}$  é uma matriz de coeficiente de regressão  $n \times n$ , determinada por minimização de mínimos quadrados (ROBERTSON et al., 2009). Neste caso, cada nova dimensão do espaço projetado é denominada de variável latente, e elas não necessariamente são ortogonais entre si. Uma simplificação visual da forma matricial da regressão PLS pode ser vista na Figura 2.3. Podemos perceber pela figura que, se realizadas individualmente, as decomposições das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  equivalem a uma PCA.

$$\begin{array}{c}
 \begin{array}{c} n \times m \\ \boxed{X} \end{array} = \begin{array}{c} n \times a \\ \boxed{T} \end{array} \begin{array}{c} a \times m \\ \boxed{P'} \end{array} + \begin{array}{c} n \times m \\ \boxed{E} \end{array} \\
 \begin{array}{c} \text{Correlação interna} \\ \updownarrow \\ U = B \times T \end{array} \\
 \begin{array}{c} \boxed{Y} \\ n \times p \end{array} = \begin{array}{c} \boxed{U} \\ n \times a \end{array} \begin{array}{c} \boxed{Q'} \\ a \times p \end{array} + \begin{array}{c} \boxed{F} \\ n \times p \end{array}
 \end{array}$$

Figura 2.3. Simplificação visual da forma matricial da regressão PLS.

Apesar da grande funcionalidade, as técnicas de Extração apresentam também algumas desvantagens. Primeiramente, o novo espaço criado, apesar de reduzido, ainda é formado pela combinação de todo o espaço original. Sendo assim, se mantém a necessidade de capturar toda a informação original. Se o objetivo da redução de dimensão for reduzir também a necessidade de coleta de dados, isto não acontece com a aplicação direta destas técnicas. Além disso, a combinação faz com que a significância física das variáveis originais seja perdida.

Por outro lado, as abordagens de Seleção de Variáveis visam a seleção de um pequeno subconjunto dos dados originais capaz de reduzir a redundância e maximizar a relevância em relação a função objetivo. Desta forma, as significâncias físicas das variáveis são mantidas e a redução de dimensão acarreta diretamente na redução da necessidade de coleta de dados (TANG; ALELYANI; LIU, 2014). Os métodos de Seleção podem ser categorizados em três grupos (PES, 2019):

- 
- (I) *Filters*. São técnicas que se baseiam nas características dos dados. O processo de seleção é conduzido em uma etapa de pré-processamento, sem interagir diretamente com o algoritmo de aprendizado usado na etapa de construção dos modelos (e.g., Fisher Score (O. DUDA; E. HART; G.STORK, 2019), ReliefF (ROBNIK-ŠIKONJA; KONONENKO, 2003)).
- (II) *Wrapper*. Usam o preditor como um modelo caixa-preta e a performance do preditor como a função objetivo para avaliar o subconjunto de variáveis selecionadas. O componente de seleção de variáveis escolhe um subconjunto que é fornecido ao algoritmo de aprendizado para prever a saída de interesse. A performance do preditor é então devolvida ao componente de seleção de variáveis para ser utilizado na próxima iteração de seleção (TANG; ALELYANI; LIU, 2014). Se trata de uma metodologia iterativa, podendo ser conduzida até o limite onde todas as combinações possíveis de subgrupos sejam avaliadas, se tornando uma Busca Exaustiva. Para muitos casos, porém, o teste de todas as possíveis combinações é infactível. Assim, algoritmos de seleção sequencial (como *Sequential Feature Selection* (SFS) e seus variantes (REUNANEN; GUYON; ELISSEFF, 2003)) e algoritmos de meta-heurísticas (como Algoritmo Genético (GOLDBERG, 1989), Otimização Enxame de Partículas (HU et al., 2019), Otimização Colônia de Formigas (PESSOA et al., 2015), Recozimento Simulado (KIRKPATRICK; GELATT; VECCHI, 1983) e Busca Tabu (MELO, 2008)), que produzem mínimos locais (e não globais), são utilizados.
- (III) *Embedded*. O processo de seleção se baseia na capacidade intrínseca do algoritmo de aprendizado em designar pesos para as variáveis, sem utilizar um mecanismo de busca sistemática entre diversos subconjuntos. O objetivo destes métodos é reduzir o tempo computacional requerido pelos métodos *Wrapper*, aceitando uma troca entre a performance preditiva do modelo final, embutindo a etapa de seleção de variáveis dentro do próprio algoritmo de aprendizado (PES, 2019). Podemos citar como exemplo SVM-RFE (GUYON et al., 2002) e técnicas de regularização baseadas em norma  $L_1$  como regularização Lasso e regularização Rede Elástica (TANG; ALELYANI; LIU, 2014).

### 2.2.1 Meta-heurísticas de Otimização

As meta-heurísticas combinam conhecimento histórico dos resultados anteriores com escolhas aleatórias para guiar o processo de seleção de variáveis. Geralmente são aplicados para resolver problemas de otimização combinatória para os quais não se conhece um algoritmo específico eficiente. Cada meta-heurística aborda de maneira diferente os mecanismos de prospecção e exploração da superfície de resposta e a forma de qualificar e combinar os subconjuntos a cada iteração.

Nos Algoritmos Genéticos (GA), a população de possíveis soluções é tratada como vetores (cromossomos) formados por genes, e atua pela simulação do que seria a reprodução sexual. A cada iteração, os indivíduos resultantes são formados pela combinação das soluções contidas em seus pais. A qualidade dos descendentes é avaliada (aptidão), e indivíduos com baixa aptidão tem menor probabilidade de reprodução. Desta forma, as gerações subseqüentes serão formadas por soluções cada vez melhores (GOLDBERG, 1989).

Os algoritmos de Enxame de Partículas (PSO) são inspirados pelo comportamento de bando (pássaros, peixes e insetos) e foram inicialmente propostos para resolução de problemas contínuos. As soluções em potencial (partículas) são iniciadas aleatoriamente e a melhor partícula é selecionada (qualidade da função objetivo). A cada iteração, as partículas voam através do espaço do problema seguindo na direção da partícula com a melhor solução atual. A direção do movimento a cada nova iteração depende tanto da posição da melhor partícula no enxame, quanto da posição da própria partícula. Regras estocásticas são introduzidas para dar um caráter aleatório ao algoritmo. O PSO se baseia na trajetória das partículas e nos pontos do espaço de busca visitados para qualificar a qualidade da solução, preservando os melhores locais visitados em uma estrutura de memória (KENNEDY; KENNEDY; EBERHART, 1995).

A Busca Tabu (TS), por sua vez, aplica métodos de busca local para otimizar modelos. Uma solução é iniciada (aleatória ou direcionada por algum método) e seus vizinhos imediatos são avaliados (soluções similares a atual, com a troca de pequenos detalhes). Se algum dos vizinhos possuir uma solução melhor que a atual, ela se torna a nova melhor solução do problema, e a solução anterior é adicionada à lista tabu, se tornando um movimento proibido. Para evitar ficar presa rapidamente em mínimos locais, a TS aplica critérios que permitem que a solução se mova em direções de piores respostas que a atual, desde que nenhum movimento melhor esteja disponível. Como a resposta anterior é incluída na lista tabu, o algoritmo evita tomar movimentos repetitivos e retornar à solução anterior, ampliando a região de busca. Regras também podem ser introduzidas para que movimentos (como combinação de certas variáveis) sejam considerados proibidos e assim não considerados durante a escolha dos vizinhos. Os algoritmos de TS geralmente trabalham com três tipos de memórias tabu: de curta, média e longa duração. A memória de curta duração diz respeito a movimentos realizados nas últimas  $t$  iterações, e os mesmos se tornam novamente permitidos com a evolução da otimização. A memória de média duração pode ser utilizada para retornar à otimização a um movimento passado de boa solução, e criar um novo nó na busca, proibindo o caminho anteriormente escolhido. A memória de longa duração pode ser utilizada para recomeçar o algoritmo em um outro ponto da região de busca, aumentando a diversificação. Uma vez atingido o critério de parada do algoritmo, a melhor solução dentre todos os movimentos é apresentada como solução final (GLOVER, 1989, 1990).

A Otimização Colônia de Formigas (*Ant Colony Optimization* – ACO) é um algoritmo multiagente usado para resolver problemas de otimização combinatória NP-complexos (BRUCKER, 1979). Os algoritmos ACO tentam imitar o comportamento de forrageamento de formigas, transformando seu sistema de comunicação em um modelo estatístico. Em uma colônia real, a experiência coletiva das formigas durante a busca por alimento é transmitida através de traçadores de feromônio. A Figura 2.4 apresenta um exemplo desse comportamento, proposto pelos autores da metodologia original (DORIGO; GAMBARDELLA, 1997).

Nos algoritmos ACO, esse sistema é transformado em um vetor memória, e é representado por uma trilha de feromônios. Esta trilha governa a maneira com que os agentes navegam pelo espaço de busca e resumem suas experiências. No caso da seleção de variáveis e otimização de modelos, cada variável de entrada inicia a rotina de otimização

com uma mesma quantidade de feromônio. A cada iteração, cada uma das formigas seleciona um subconjunto das variáveis de entrada. Essa seleção se baseia em um componente randômico e um componente relacionado a quantidade de feromônio nas variáveis. A formiga entrega para o algoritmo de aprendizado o subconjunto, que ajusta um modelo para prever as saídas de interesse. A performance do modelo é quantificada, e cada formiga deposita nas variáveis que selecionou uma quantidade de feromônio condizente: quanto melhor a performance, maior a quantidade de feromônio. Por fim, toda a trilha é evaporada, multiplicando a mesma por um valor entre 0 – 1, para penalizar variáveis não selecionadas. Esta operação garante um aumento da probabilidade, nas sucessivas iterações, da seleção de variáveis de entrada que participaram de modelos de alta qualidade, otimizando o preditor.

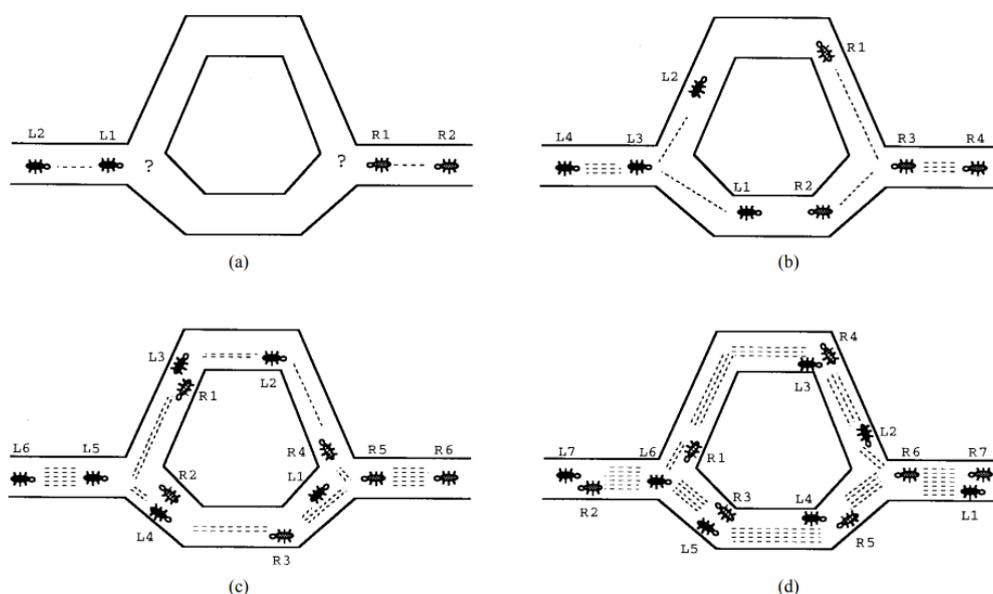


Figura 2.4. Definição de como formigas reais escolhem o menor caminho. (a) As formigas chegam ao ponto de decisão. (b) Randomicamente decidem o caminho a seguir. (c) As formigas que escolheram o caminho mais curto chegam ao ponto oposto em menos tempo. (d) O feromônio acumula no caminho mais curto em uma taxa mais alta. Assim, mais formigas tendem a seguir por ele. As linhas pontilhadas representam uma aproximação proporcional do feromônio depositado pelas formigas. Fonte: (DORIGO; GAMBARDELLA, 1997).

Na literatura, podemos encontrar diversos artigos de revisão que acompanham o desenvolvimento de metodologias baseadas em ACO. Reunindo os trabalhos de Dorigo e Blum (2005), Stützle, López-Ibáñez e Dorigo (2011) e Mohan e Baskaran (2012), são apresentadas mais de uma centena de pesquisas que aplicam ou modificam a metodologia ACO, em diversas áreas de estudo, como: topologia e otimização estrutural em eletrônica; aerodinâmica e dinâmica de fluidos; telecomunicações, bioinformática; finanças; modelagem, simulação e identificação de sistema em química, física e biologia; controle e processamento de sinais e imagens; problemas de roteamento, programação e problemas de produção, logística, transporte e gestão da cadeia de abastecimento.

A metodologia continua ativamente relevante, uma vez que permite muita liberdade para adaptações. Como uma evolução natural, é comum encontrar propostas de hibridização entre ACO e outras meta-heurísticas ou técnicas de otimização, para compensar por desvantagens na metodologia original (como, por exemplo, problemas de

*overfitting*, alta demanda computacional, estagnação precoce e ciclagem de subconjuntos). Algoritmos híbridos ACO já foram propostos para resolver Problemas de Caixeiro Viajante (ZHANG; TANG, 2008), Problema de Roteamento de Veículos (ABDULKADER; GAJPAL; ELMEKKAWY, 2015; BALSEIRO; LOISEAU; RAMONET, 2011; LI; TIAN; LEUNG, 2009; WANG et al., 2020), Problema De Atribuição De Tarefas Para Estações De Trabalho (SERBENCU; MINZU, 2016), Problema De Escalonamento De Projetos Com Restrição De Recursos (MYSZKOWSKI et al., 2015), entre outros. Um aprofundamento e compilação de estudos sobre o uso de meta-heurísticas híbridas pode ser visto nos trabalhos de revisão apresentados por Blum *et al.* (2008), Blum *et al.* (2011) e Bhattacharyya (2018).

### 2.3 Redes Neurais Artificiais

Redes neurais artificiais, como o nome sugere, são modelos matemáticos inspirados na estrutura básica de um cérebro. Este modelo consiste de diversas unidades de processamento (neurônios), operando paralelamente em uma camada, que são conectados aos neurônios de outras camadas por ligações (sinapses), representadas por pesos. Algumas vezes chamados de aproximadores universais, teorizasse que com unidades e informações suficientes os modelos possam aproximar virtualmente qualquer função a qualquer grau de precisão (CYBENKO, 1989; HORNIK; STINCHCOMBE; WHITE, 1989). Uma representação da estrutura de um neurônio artificial pode ser vista na Figura 2.5.

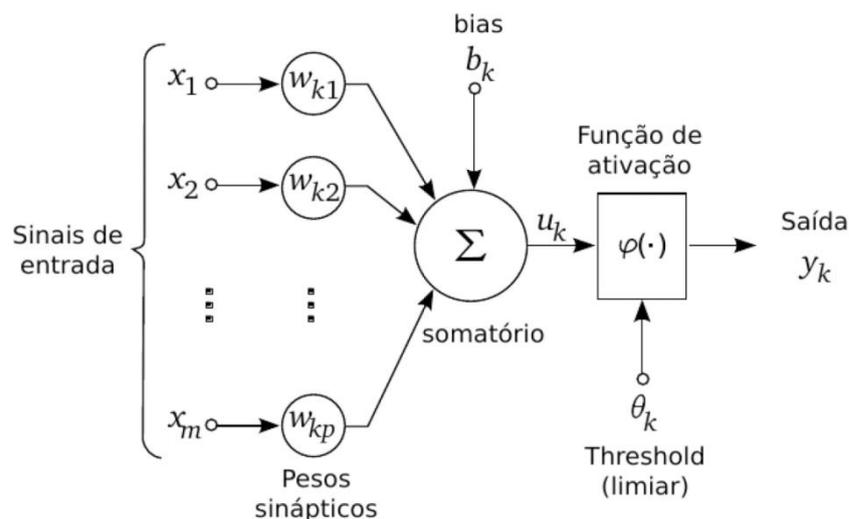


Figura 2.5. Estrutura básica de um neurônio artificial. Fonte: Schimidt *et al.* (2016).

A estrutura de uma rede neural diz respeito à disposição dos neurônios em camadas, e em como estas camadas estão conectadas entre si. As camadas são segmentadas em entradas, ocultas e saídas (sendo as camadas ocultas um número qualquer de camadas entre as entradas e saídas). Uma rede com apenas uma camada oculta é denominada rasa, e o acréscimo de mais camadas ocultas dá origem as chamadas redes profundas (*deep neural network*). O conceito de redes profundas também evoluiu junto com o desenvolvimento de metodologias mais eficientes de treinando das redes e do poderio computacional, sendo hoje possível treinar redes com centenas de camadas ocultas de maneira eficiente em poucos minutos (LEIJNEN; VAN VEEN, 2020).

Os neurônios, por sua vez, podem estar completamente interconectados com todos os neurônios da camada anterior e os da camada seguinte, formando a chamada unidade densa, ou o seu antônimo, chamada de unidade esparsa, onde uma seleção interna da rede poda conexões entre neurônios que não agregam informação útil (pesos próximos ou exatamente zero) (SRINIVAS; SUBRAMANYA; BABU, 2016).

Ainda sobre a conexão entre camadas da rede, a arquitetura mais comum é aquela onde a saída de uma camada é a entrada da camada subsequente. Este tipo de rede se denomina *feedforward*, e a informação é passada adiante de forma sequencial, sem que a função de saída da unidade tenha dependência com a própria saída de nenhuma forma. Em contraponto, redes do tipo recorrentes/*feedback* apresentam conexões em laços ou *loops*, onde a saída de camadas avançadas retorna às camadas anteriores (SCHWENK; BENGIO, 2000). A Figura 2.6 apresenta diferentes arquiteturas de redes neurais, e sua denominação usual.

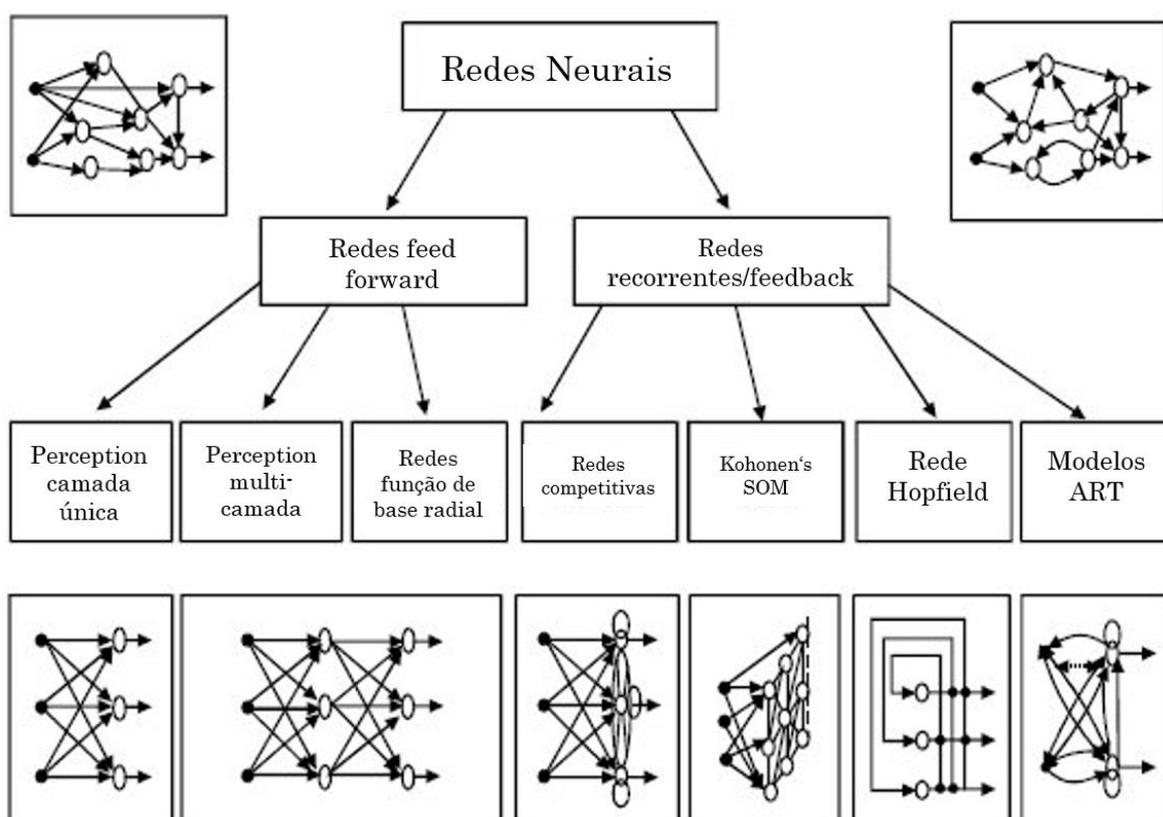


Figura 2.6. Diferentes arquiteturas de redes neurais artificiais. Adaptada de (VERMA; KUMAR SINGH, 2015).

O principal conceito utilizado para o aprendizado das redes é o algoritmo de *backpropagation*, buscando a minimização de uma função objetivo para o ajuste de determinadas saídas. O algoritmo relaciona alterações nos parâmetros de peso e bias da rede às alterações na função objetivo. Essas alterações são calculadas iniciando pela última camada e seguem recursivamente até a primeira. As direções das alterações podem ser calculadas através das derivadas parciais dos erros, buscando a melhor alteração nos parâmetros capaz de minimizar a função objetivo. A magnitude da alteração nos parâmetros a cada iteração é comumente denominada taxa de aprendizado. Durante o processo de aprendizado pode ocorrer a saturação do neurônio, a depender do tipo de função de ativação e das entradas do mesmo. Neste caso, sua saída acaba sendo muito próxima de 0 ou 1, e alterações no peso pouco influenciem os outros neurônios da rede, e

por fim, a função objetivo. A saturação de neurônios (estado em que um neurônio emite valores predominantemente próximos às extremidades assintóticas da função de ativação) causa lentidão no aprendizado, principalmente quando o algoritmo de gradiente descendente é utilizado para otimização dos parâmetros (WIDROW; LEHR, 1990).

Um neurônio representa o somatório das unidades anteriores da rede, acrescido de um bias, sobre o qual é aplicada uma operação matemática, denominada função de ativação, gerando o sinal de saída da unidade. A função de ativação, assim, realiza transformações lineares ou não lineares nos sinais de entrada, permitindo que as mudanças nos pesos da rede sejam reguladas e que a rede consiga aprender a resolução para problemas complexos (CHARTE et al., 2018). Na literatura são sugeridas dezenas de abordagens para funções de ativação, sendo a função sigmoïdal a primeira de grande destaque e as mais modernas variantes de unidades lineares como: ReLU, PReLU, ELU e SELU (ZHANG; LI, 2018). A Figura 2.7 apresenta algumas das mais comuns funções de ativação.

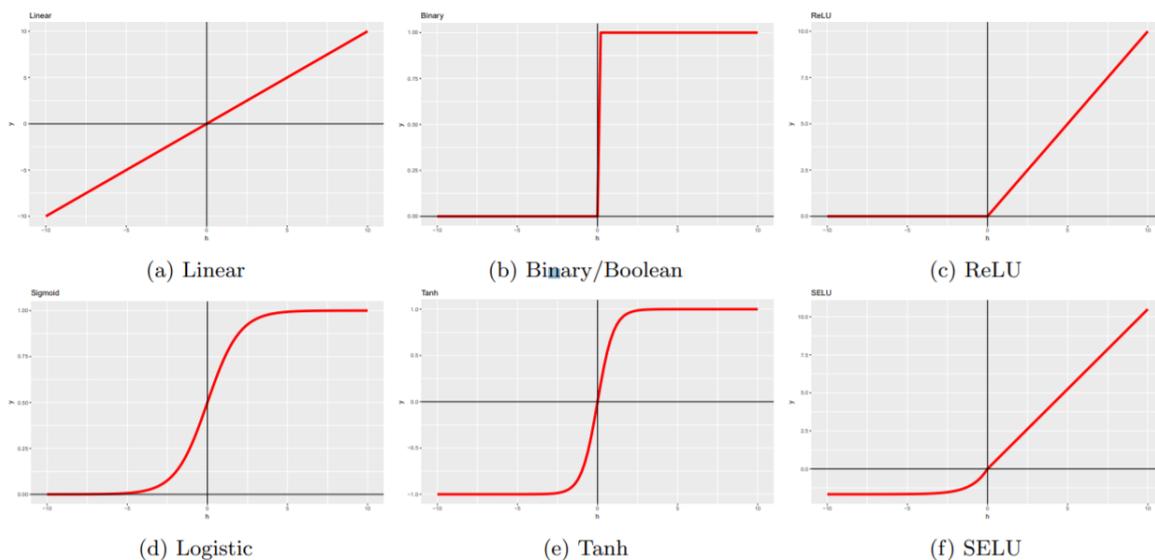


Figura 2.7. Funções de ativação comuns em redes neurais artificiais. Fonte: (CHARTE et al., 2018).

De maneira geral, as centenas de arquiteturas de redes propostas na literatura são combinações entre número de camadas, número de neurônios, e a maneira com que estes neurônios estão conectados entre si. Muitas destas arquiteturas são idealizadas com o objetivo de lidar com uma situação específica, como as redes convolucionais (CNN) e redes autoencoders (AE). As CNN agregam informações espaciais a rede, especialmente úteis para lidar com dados na forma de tensores (múltiplas dimensões, como imagens) (ZEILER; FERGUS, 2014). As redes AE, por sua vez, tem por objetivo treinar uma rede capaz de extrair de maneira não supervisionada padrões dos dados, sendo a rede capaz de codificar o espaço original em um novo espaço, e decodificar o mesmo buscando a reconstrução dos dados originais (CHARTE et al., 2018).

### 2.3.1 Redes Neurais Convolucionais

Redes convolucionais são consideradas por alguns pesquisadores como os melhores algoritmos de aprendizado para tratamento de imagens (CIREŞAN et al., 2012; LIU; DENG; YANG, 2019). Sua principal característica é sua habilidade em explorar correlações temporais e espaciais nos dados. A topologia da rede é dividida em múltiplos estágios compostos da combinação entre camadas convolucionais, unidades de processamento não linear e camadas de subamostragem (*pooling*) (LECUN; KAVUKCUOGLU; FARABET, 2010). As operações convolucionais extraem recursos úteis de dados localmente correlacionados. A saída da camada convolucional é transformada por uma função de ativação não linear, facilitando o aprendizado de diferenças semânticas nos dados. A saída da função de ativação é geralmente seguida de uma operação de subamostragem, que condensa os resultados e torna as entradas menos propensas a distorções geométricas (SCHERER; MÜLLER; BEHNKE, [s.d.]). Após os múltiplos estágios a rede é geralmente conectada a camadas densas de neurônios, para então gerar a saída de interesse. A Figura 2.8 apresenta um exemplo de CNN aplicada na classificação de imagens.

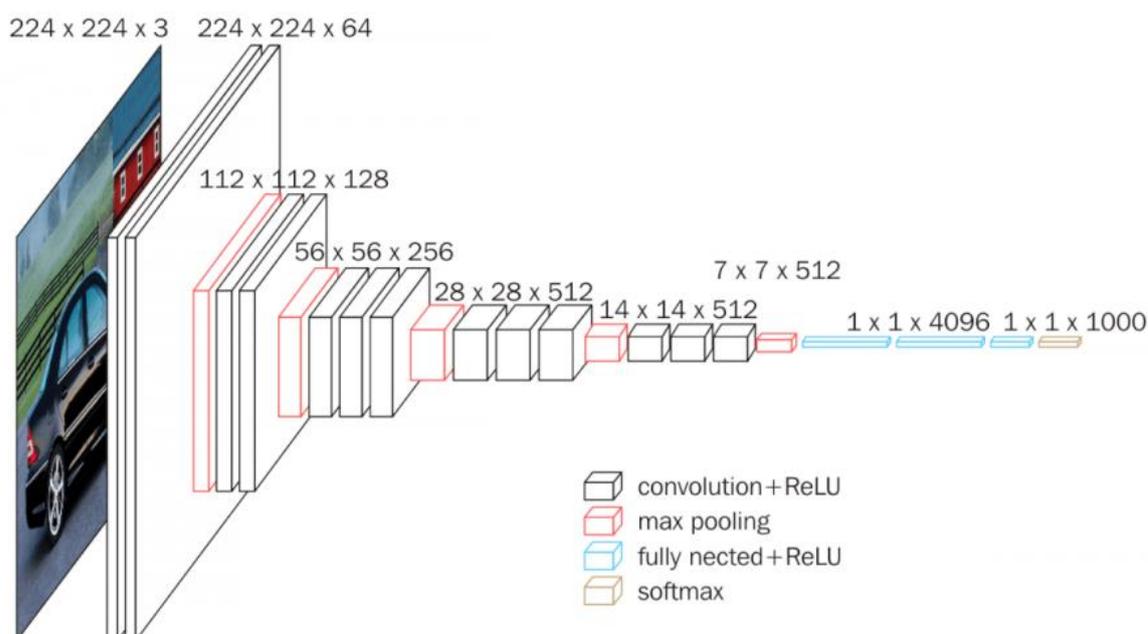


Figura 2.8. Exemplo de CNN aplicada na classificação de imagens. A imagem original possui 224x224 pixels e três canais (*Red-Blue-Green*). O espaço original é reduzido por camadas *max pooling*. A terceira dimensão apresentada é referente ao número de *kernels* (filtros) utilizados na camada convolucional. Fonte: (SIMONYAN; ZISSERMAN, 2015).

A arquitetura de CNN inclui diversos blocos construtivos, explicados a seguir.

#### 2.3.1.1 Camadas Convolucionais

Componente fundamental da arquitetura CNN, especializada na extração de características. Consiste de uma operação linear onde um pequeno número de matrizes, denominadas *kernel*, são aplicados no tensor de entrada. O produto elemento-a-elemento entre cada elemento do *kernel* e o tensor é calculado, em cada localização do tensor, e somado para obter o valor de saída naquela posição do tensor de saída, denominado mapa de recursos. O procedimento é aplicado repetidamente utilizando múltiplos *kernels*, formando um número arbitrário de mapas de recursos, que capturam diferentes características do tensor. A Figura 2.9 representa a operação convolucional.

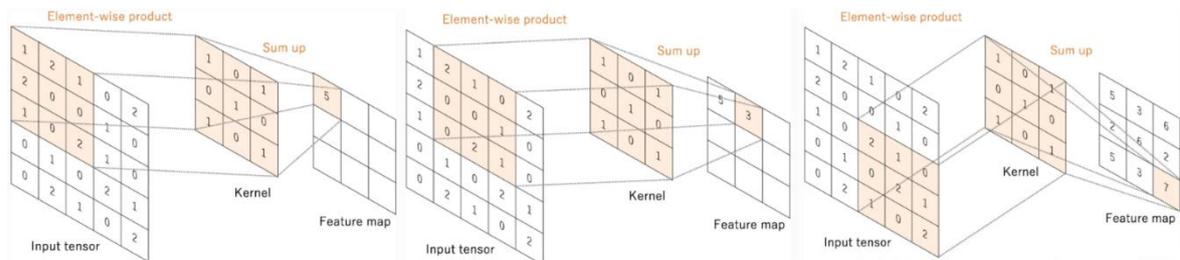


Figura 2.9. Representação de operação convolucional. Adaptado de Yamashita *et al.* (2018).

A distância entre duas posições sucessivas do kernel é denominada *stride*, que também define a operação convolucional. Geralmente, um *stride* de 1 é o mais efetivo, mas valores maiores podem ser usados com o intuito de diminuir o tamanho do mapa de recursos. A operação convolucional reduz as dimensões do mapa de recursos, em comparação ao tensor original. Para evitar esta redução, e assim possibilitar o uso de um maior número de estágios, a técnica conhecida como *padding* é utilizada, consistindo na adição de linhas e colunas contendo zeros a cada lado do tensor, mantendo as dimensões originais após a operação convolucional. Além de manter as dimensões, o uso de *padding* garante que os dados nas pontas do tensor não tenham sua importância menosprezada.

O processo de treinar uma CNN, em relação as camadas convolucionais, versa sobre identificar quais os valores dos kernels que minimizam a função objetivo em questão. Os kernels são os únicos parâmetros aprendidos automaticamente durante o processo de treinamento da camada convolucional. O número de kernels, seu tamanho, o *stride* e o número de linhas/colunas adicionadas com *padding* são hiperparâmetros definidos pelo usuário (YAMASHITA *et al.*, 2018).

### 2.3.1.2 Camadas pooling

As características extraídas pela camada convolucional podem ocorrer em diferentes partes do tensor. Sua localização específica se torna menos importante, desde que mantida a sua posição relativa as outras características extraídas. Assim, camadas *pooling* resumem informação similar nas redondezas da região e expressam como saída apenas a resposta dominante, diminuindo o tamanho do mapa de recursos (LEE; GALLAGHER; TU, 2015). Os principais exemplos desta camada são máximos (apenas o maior valor é mantido) e medios (o valor médio é mantido), mas outras abordagens podem ser utilizadas (HE *et al.*, 2015). Neste caso, nenhum parâmetro é aprendido durante o treinamento da rede, e o tamanho, *stride* e *padding* da camada são hiperparâmetros. A Figura 2.10 representa uma camada de máximo *pooling*, com kernel  $[2 \times 2]$ , *stride* 2 e sem *padding*.

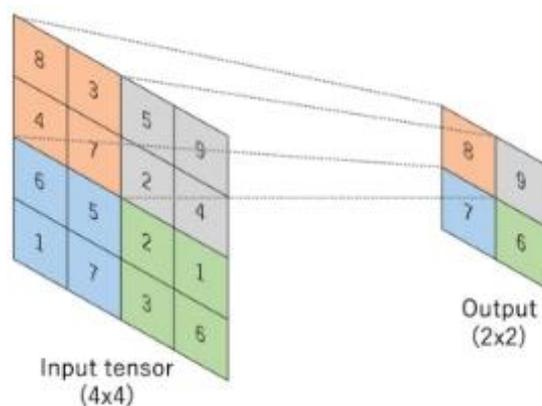


Figura 2.10. Camada *max pooling* com kernel  $[2 \times 2]$ , *stride 2* e sem *padding*. Adaptado de Yamashita *et al.* (2018).

### 2.3.2 Normalização em lote - Batch normalization

Sugerido por Ioffe e Szegedy (2015), a normalização em lote visa resolver o problema do deslocamento de covariância interna entre mapas de recursos. A distribuição de valores no mapa de recursos é unificada levando a média a zero e a variância unitária. O fluxo do gradiente é suavizado, aprimorando a capacidade de generalização da rede.

### 2.3.3 Evolução da arquitetura CNN

O estudo realizado por Khan *et al.* (2020) apresenta uma extensa compilação das diversas evoluções na arquitetura das CNN ao longo dos últimos 40 anos. A Figura 2.11 serve como resumo visual dos principais pontos que marcaram essa evolução, e a nomenclatura das redes mais significativas de cada período.

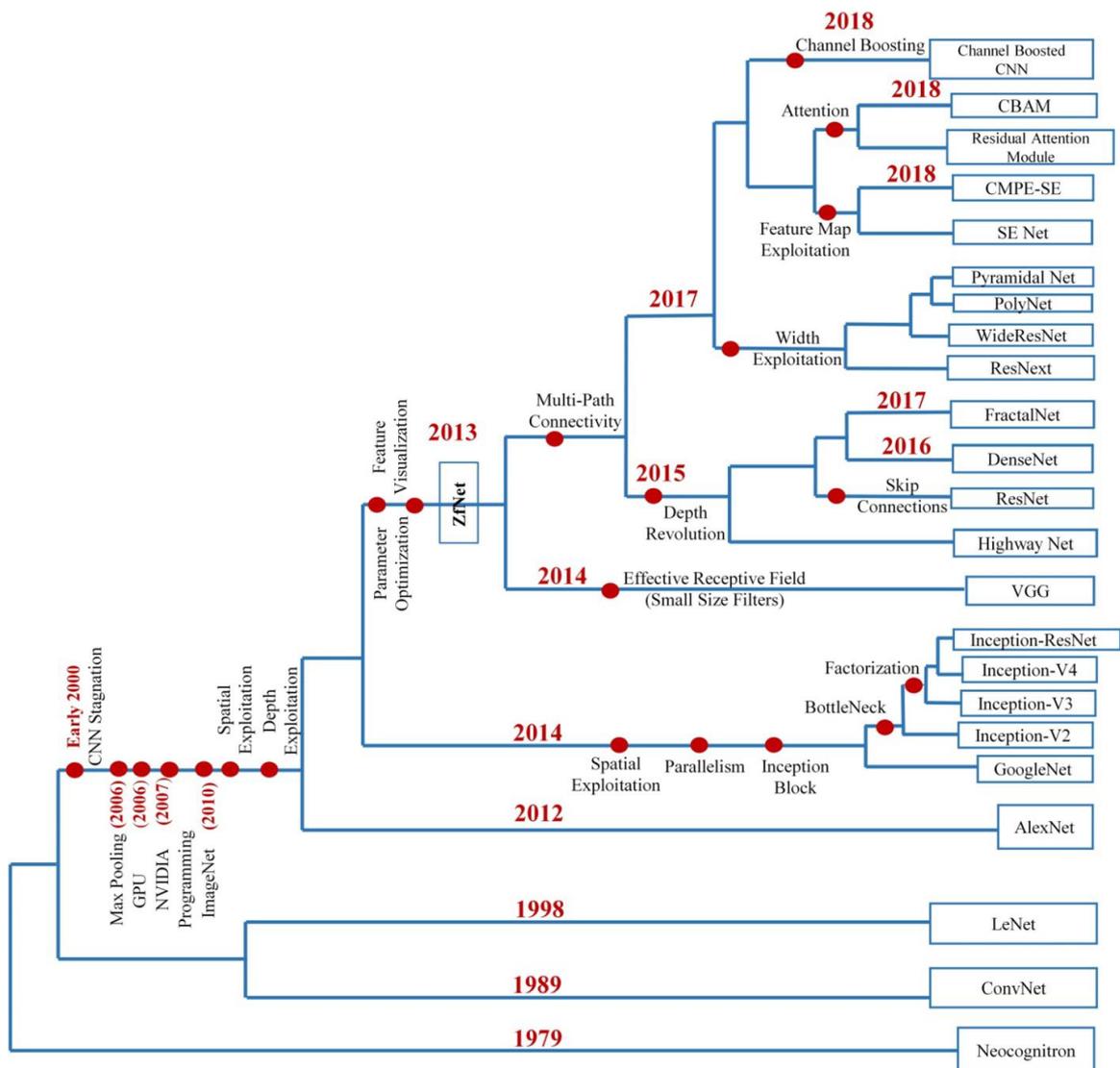


Figura 2.11. Histórico evolutivo das CNN profundas, com destaque para as inovações nas arquiteturas. Fonte: (KHAN et al., 2020).

Apesar de majoritariamente utilizadas para tratamento de imagens, redes convolucionais também podem ser aplicadas em estudos de quimiometria. Rutherford *et al.* (2020) utilizou uma CNN para interpretar matrizes excitação-emissão (EEM) de fluorescência, a fim de identificar a fonte emissora de material particulado proveniente de combustão. Por sua vez, Itakura *et al.* (2018) utilizou uma CNN regressora para estimar maturidade de citros, também com base em EEM de fluorescência. A natureza bidimensional das matrizes espectrais abre a possibilidade para aplicação de classificação ou regressão utilizando CNN, área ainda pouco explorada industrialmente.

### 2.3.4 Redes Autoencoder - AE

Autoencoder são uma técnica de aprendizado não supervisionado que pretende capturar as características do sistema em uma camada interna, que age como extrator de padrões dos dados (WETZEL, 2017). AE podem ser apresentados como uma generalização não linear da análise de componentes principais (PCA), utilizando uma rede neural simétrica para primeiramente codificar os dados de entrada em um novo espaço latente,

e, posteriormente, decodificar este espaço, buscando a reconstrução do espaço original. Restrições são impostas para que a rede não simplesmente se torne uma identidade, repetindo diretamente as entradas nas saídas (CHOLLET, 2015). A Figura 2.12 apresenta um AE genérico do tipo *bottleneck*, com redução da dimensão do espaço original.

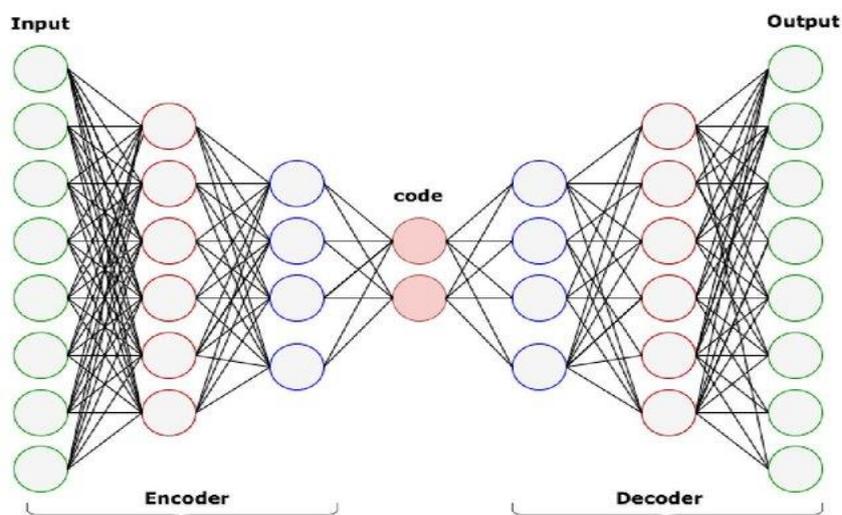


Figura 2.12. Esquema genérico da arquitetura de um AE. Fonte: (BAHI; BATOUCHE, 2018).

Diversos trabalhos na literatura comparam o uso de PCA e AE, geralmente explicitando as vantagens das redes em capturar padrões não lineares nos dados, identificando características no espaço latente que a PCA não consegue capturar (HINTON; SALAKHUTDINOV, 2006; JES et al., 2019; MANNING-DAHAN, 2017). Esta diferença pode ser vista na Figura 2.13, comparando a projeção espacial dos dois primeiros componentes principais – PCA (A) e do espaço latente de 2 neurônios da rede AE (B), para o conjunto dígitos escritos à mão, MNIST. Assim como no caso da PCA, o espaço latente gerado pelo AE pode ser utilizado como base para modelos de regressão (HINTON; SALAKHUTDINOV, 2006).

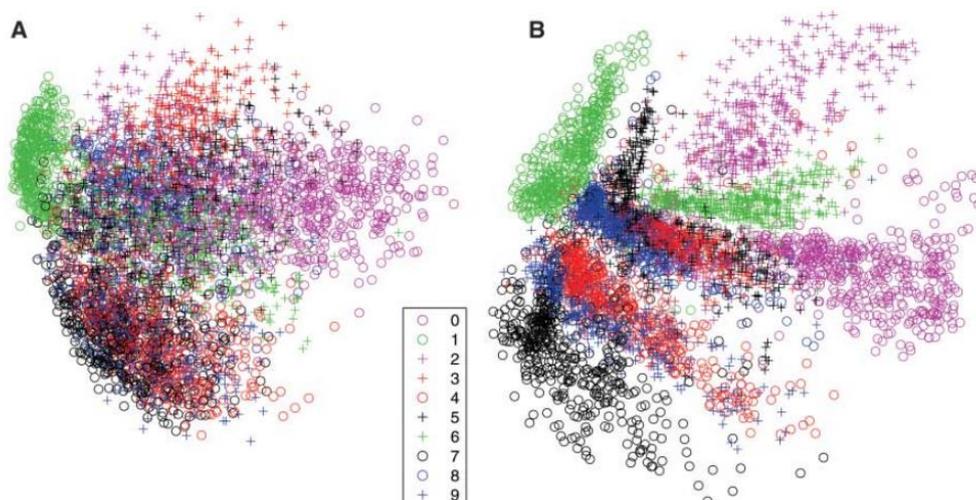


Figura 2.13. Projeção espacial dos dois primeiros componentes principais – PCA (A) e do espaço latente de 2 neurônios da rede AE (B), para o conjunto dígitos escritos a mão, MNIST. Fonte: (HINTON; SALAKHUTDINOV, 2006).

A construção de um AE segue os mesmos princípios básicos de outras redes neurais, e pode ser resumido no esquema apresentado na Figura 2.14.

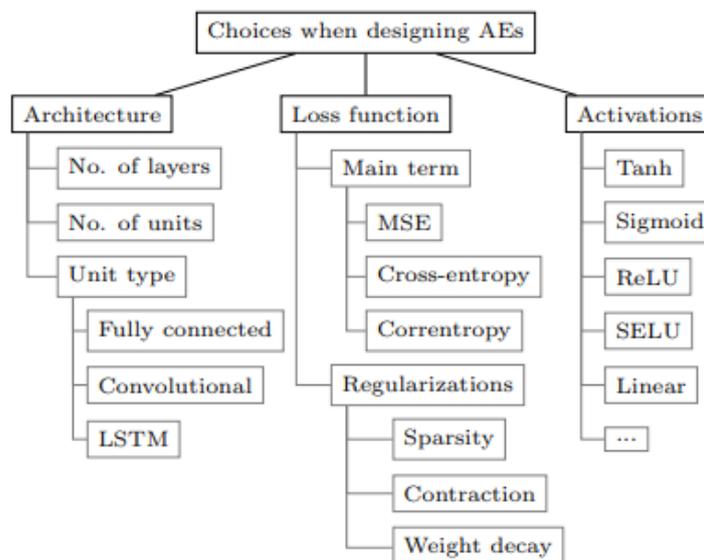


Figura 2.14. Resumo das escolhas necessárias para o design de um AE. Fonte: (CHARTE et al., 2018).

Além da clássica rede *feedforward* apresentada, diversas arquiteturas de redes AE são propostas na literatura para resolver problemas específicos, como:

- *Denosing AE* – DAE (VINCENT et al., 2008): adaptação desenvolvida para criar uma rede que fosse robusta a dados corrompidos. A DAE aprende a gerar padrões robustos das entradas reconstruindo amostras parcialmente destruídas. A estrutura e parâmetros são idênticos a de um AE comum, mas, durante a fase de treinamento da rede, erros estocásticos corrompem as entradas, enquanto o erro de reconstrução a ser minimizado ainda considera a base de dados original. Assim, a rede é treinada para prever os valores falhos, e reconstruir os dados originais de maneira a não depender destas entradas.
- *Convolutional AE* – CAE (MASCI et al., 2011): AE padrões não explicitamente consideram a estrutura bidimensional quando processam os dados. CAE resolvem este problema aplicando camadas convolucionais ao invés de camadas densas. Em alguns casos, CAE são utilizadas como ponto de partida para inicializar CNN.
- *LSTM AE* (SRIVASTAVA; MANSIMOV; SALAKHUTDINOV, 2015): AE básicos não são planejados para modelar dados sequenciais. O uso de unidades LSTM (*Long-Short-Term Memory*) como codificadores e decodificadores da rede resolve este desafio. O codificador comprime a sequência em uma representação de tamanho fixo, e o decodificador treina para extrair a sequência original na ordem inversa. Esta abordagem é especialmente útil para dados sequenciais e grandes, como arquivos de vídeo.
- *Variational AE* – VAE (KINGMA; WELLING, 2014): Este tipo de AE aplica uma abordagem variacional Bayesiana para codificação. Se assume que existe uma variável aleatória latente não observada  $y$ , que por um processo randômico leva

as observações,  $x$ . O objetivo é aproximar a distribuição das variáveis latentes dadas as observações. VAE substituem as funções determinísticas no codificador e decodificador por mapeamento estocástico, e calculam a função objetivo em virtude da função densidade das variáveis randômicas. A principal vantagem deste tipo de rede é que é possível gerar novas amostras a partir da distribuição aprendida, e são muito aplicadas em casos que envolvem a necessidade de gerar novas instâncias (DOSOVITSKIY; BROX, 2016).

As principais aplicações das diversas abordagens de AE são a classificação (transformam os dados de entrada para obter melhores performances em classificadores); compressão de dados (AE treinada com um tipo específico de dados para aprender formas eficientes de compressão); *hashing* (resumem os dados de entrada em vetores binários para acelerar buscas); e visualização (projetam os dados em duas ou três dimensões para representação gráfica).

Uma característica intrínseca das AE é seu potencial para detecção de padrões anormais nos dados. Como a rede é treinada para fielmente reconstruir o padrão aprendido dos dados de calibração, amostras anormais se destacam na base de dados, apresentando erros de reconstrução. Sakurada e Yairi (2014) comparam o potencial de um AE para detectar dados anormais em um sistema artificial e em dados reais de telemetria de aeronaves, comparando com outras abordagens utilizando PCA e kernel PCA. Em ambos os casos, a performance do AE foi superior as demais metodologias. Zou *et al.* (2019) e Misra *et al.* (2020) propõem o uso de AE para detecção de fraudes em cartão de crédito. Amarbayasgalan *et al.* (2018) propõe uma nova metodologia baseada em DAE e clusterização para identificação de dados anômalos em 20 bases de dados de referência. Os autores comparam a metodologia proposta com outras 12 metodologias, baseadas em *Support Vector Machines* e PCA. A metodologia sugerida é superior aos algoritmos estado-da-arte para classificação de 9 das bases avaliadas.

## 2.4 Espectroscopia

A espectroscopia estuda a interação da matéria com radiação eletromagnética. A radiação pode ser absorvida, transmitida ou dispersa pela matéria. Métodos espectroscópicos são aqueles que analisam a quantidade de radiação absorvida ou produzida pelo meio. Tais métodos são classificados de acordo com a região do espectro eletromagnético envolvido e a reação da matéria a esta energia (SETTLE, 1997).

Dentre os métodos óticos, se destacam a espectroscopia por fluorescência e a espectroscopia vibracional (espectroscopia no infravermelho próximo e médio (YADAV; YADAV, 2005) e a espectroscopia Raman (JONES *et al.*, 2019)). A espectroscopia de absorção no infravermelho estuda a interação de radiação nas frequências específicas capazes de alterar o nível vibracional ou rotacional das moléculas. O deslocamento dos átomos em vibração altera o momento dipolar da molécula, e causa estiramento das ligações. Como a energia de cada nível quântico vibracional é específica, a análise da absorção de diversas frequências permite determinar a composição química da amostra. A espectroscopia Raman, por sua vez, analisa o espalhamento inelástico da luz pela matéria. Como este espalhamento depende da estrutura da molécula e de seus níveis de energia vibracional, a espectroscopia Raman consegue detectar o tipo de estrutura, de ligações e outras características físico-químicas do composto. A Tabela 2.1 apresenta algumas características e aplicações da espectroscopia vibracional.

Tabela 2.1. Características e aplicações de técnicas de espectroscopia vibracional.

<b>Técnica</b>	<b>Características</b>	<b>Aplicações</b>
Espectroscopia de infravermelho médio (MIR)	Espectroscopia de absorção Modos fundamentais de vibração (alta intensidade)	Ciência dos materiais (KISCHKAT et al., 2012; PALUSZKIEWICZ et al., 2011), agricultura (ELLEN MACARTHUR FOUNDATION, 2013; SORIANO-DISLA et al., 2014), ciência de alimentos (KAROUI; DE BAERDEMAEKER, 2007; SINELLI et al., 2008), indústria farmacêutica (BRITAIN, 2018)
Espectroscopia de infravermelho próximo (NIR)	Espectroscopia de absorção Sobretons e bandas de combinação (mais fracas)	Ciências agrícolas (WILSON; ZHANG; KOVACS, 2014), pecuária (VALENTI et al., 2013), biotecnologia (YU et al., 2014), combustíveis (CRAMER et al., 2009), química fina (SHINZAWA et al., 2012)
Espectroscopia Raman	Espectroscopia de espalhamento Composição molecular e estrutural	Cristalografia (CAREY, 2014), ciência dos materiais (DRESSELHAUS; JORIO; SAITO, 2010), ciências biológicas (RYGULA et al., 2013), ciência de alimentos (KRIEG, 2014; REID; O'DONNELL; DOWNEY, 2006), indústria farmacêutica (ŠAŠIĆ, 2007)

A luminescência é uma família de processos onde moléculas suscetíveis emitem luz a partir de estados eletrônicos excitados, seja por meio físico (absorção de radiação), mecânico (fricção) ou químico (reação). Quando a geração de luminescência parte da excitação de uma molécula por fótons, o fenômeno se determina fotoluminescência, e é formalmente dividido entre fluorescência e fosforescência, dependendo do tipo da configuração eletrônica do estado excitado e do caminho de emissão de luz. O fenômeno da fluorescência é aquele observado em estados excitados do tipo singleto (o spin do elétron presente no orbital excitado se encontra pareado com o spin do elétron do estado fundamental). No caso da fosforescência, o spin do elétron excitado se encontra desemparelhado, permanecendo paralelo, em um estado denominado tripleto, mais improvável de acontecer (SKOOG; HOLLER; CROUCH, 2009). O Diagrama de Jablonski, apresentado na Figura 2.15, ilustra os processos de mudança de níveis energéticos entre absorção e emissão de luz, e serve como base para o entendimento dos fenômenos de fotoluminescência.

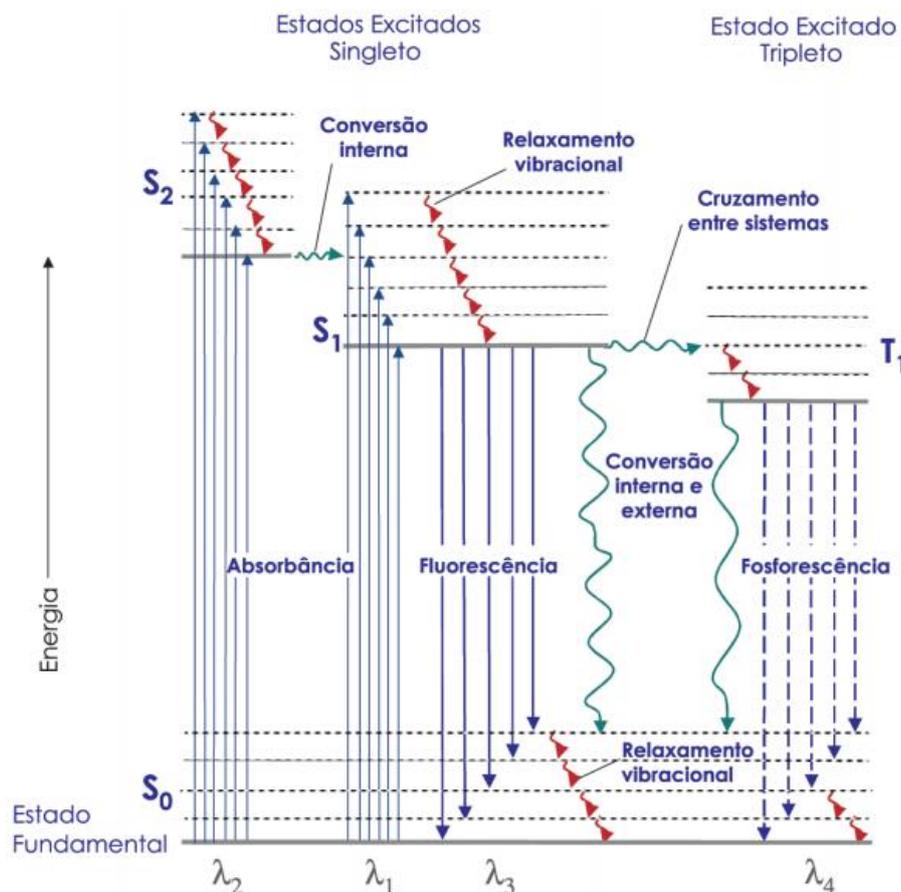


Figura 2.15. Diagrama de Jablonski.  $S_0$  representa o estado eletrônico fundamental,  $S_1$ , e  $T_1$  são os estados excitados singlete e tripleto.  $S_2$  é um segundo estado excitado singlete. As linhas horizontais pontilhadas são os vários níveis de energia vibracional dos estados.

As setas retas representam os processos envolvendo fótons, e as setas onduladas representam transições não-radioativas. Fonte: (SKOOG; HOLLER; CROUCH, 2009).

A análise das relações entre energias presentes no Diagrama de Jablonski explicam grande parte das características do processo de fluorescência. O processo é governado majoritariamente por três eventos: primeiramente, uma molécula suscetível é excitada por fótons. A excitação ocorre do estado fundamental  $S_0$  para estados singletos  $S_1$  e  $S_2$ , em diferentes níveis vibracionais de energia. A energia é geralmente dissipada através de relaxamento vibracional, levando o sistema ao menor nível energético do estado excitado  $S_1$ . Por fim, fótons são emitidos pela molécula, e a mesma retorna ao estado fundamental. Muitas vezes, ela retorna para níveis vibracionais de maior energia no estado fundamental, e rapidamente relaxa para o estado de menor energia.

#### 2.4.1 Absorção, Emissão e Deslocamento de Stokes

Para um típico fluoróforo, a irradiação por um amplo espectro de comprimentos de ondas irá gerar toda uma gama de transições permitidas, que irão povoar os vários níveis de energia vibracional dos estados excitados. Algumas destas transições terão maior probabilidade de acontecer do que outras, e, quando combinadas, constituem o espectro de absorção da molécula. A probabilidade de uma transição ocorrer entre o estado fundamental  $S_0$  para um estado excitado  $S_1$  depende da similaridade entre os níveis de energia vibracional e rotacional dos estados. O comprimento de onda de máxima absorção representa a mais provável separação entre o estado fundamental e um nível vibracional permitido no estado excitado.

Após a absorção de um fóton, diversos processos com diferentes probabilidades ocorrem, sendo os mais comuns deles a conversão interna e o relaxamento vibracional, levando a molécula ao mais baixo nível de energia vibracional do estado excitado  $S_1$ . Neste momento, a molécula pode emitir um fóton, no processo que se conhece como fluorescência. As análises dos diferentes comprimentos de onda emitidos por uma molécula ao ser excitada dão origem ao espectro de emissão. Como é raro que uma molécula emita fótons a partir de estados excitados superiores ao  $S_1$ , ocorrendo rápido relaxamento vibracional para o menor nível de energia do estado excitado, o espectro de emissão de uma molécula é geralmente independente do comprimento de onda de excitação. Este fenômeno pode ser visto na Figura 2.15, acompanhado as setas vermelhas onduladas, e as setas retas de fluorescência. Ao retornar para o estado fundamental, vários níveis de energia vibracional podem ser povoados.

Assim como para absorção, a probabilidade de um elétron em um estado excitado retornar para um nível de energia vibracional específico no estado fundamental é proporcional à sobreposição entre os níveis de energia nos respectivos estados. Desta forma, o espectro de emissão é geralmente uma imagem espelhada do espectro de absorção do estado  $S_0$  para o estado  $S_1$ . De fato, a probabilidade de um elétron retornar a um nível específico de energia vibracional no estado fundamental é similar a probabilidade daquele elétron pertencer aquele nível antes da excitação. Este conceito é conhecido como Regra da Imagem Espelhada.

A Figura 2.16 apresenta os espectros de absorção e de emissão de duas moléculas distintas. Os espectros do perileno (superior) seguem a regra da imagem espelhada. Neste caso, as distâncias entre os níveis de energia vibracional são consideráveis e isso se reflete nos múltiplos picos do espectro. Os espectros da quinina (inferior) não seguem a regra da imagem espelhada. Neste caso, o primeiro pico de absorção (315 nm) é referente a uma transição entre os estados  $S_0$  e  $S_2$ . Como a energia interna é relaxada e a molécula decresce ao menor nível vibracional de  $S_1$  antes de emitir fótons, o espectro de emissão não apresenta este pico, e é uma imagem espelhada das transições  $S_0 - S_1$ . No caso da quinina, as distâncias entre os níveis de energia vibracional são muito próximas e o espectros se apresentam na forma de bandas mais largas.

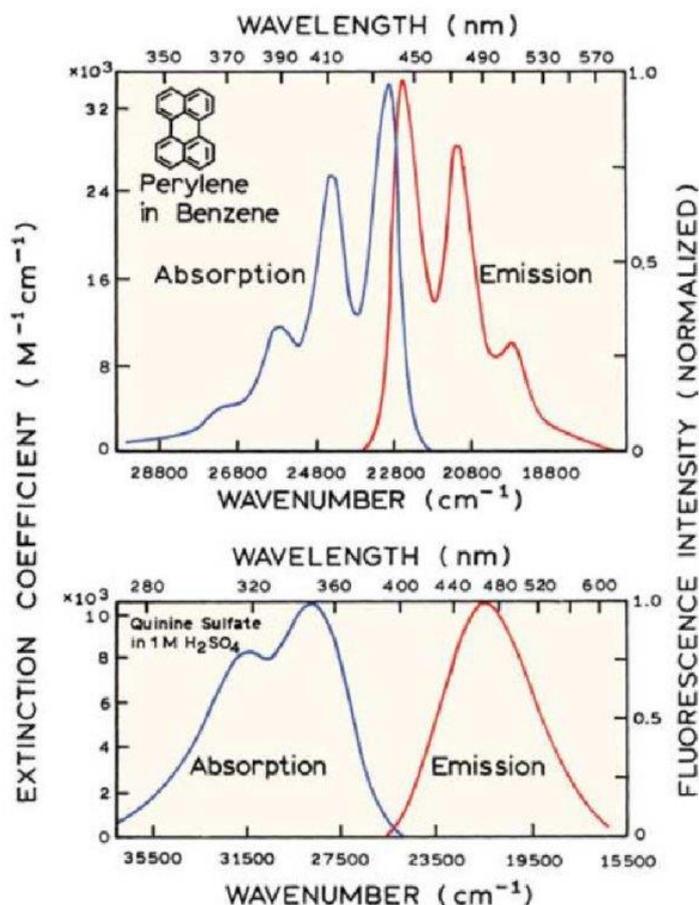


Figura 2.16. Espectro de absorção e emissão de fluorescência do perileno (superior) e da quinina (inferior). O Deslocamento de Stokes pode ser visto claramente em ambas moléculas. O perileno segue a regra da Imagem Espelhada, mas o mesmo não ocorre para a quinina. Fonte: (LAKOWICZ, 2006).

A energia associada às transições de emissão são tipicamente menores que aquelas ligadas a absorção, resultando na emissão de fótons com menor energia (e maiores comprimentos de onda). Este fenômeno é conhecido como Deslocamento de Stokes, e ocorre para virtualmente todos os fluoróforos. A principal causa do efeito é a perda de energia por relaxamento dos elétrons excitados para o menor nível de energia vibracional do estado excitado  $S_1$ . Além disso, a emissão geralmente acontece em transições para níveis de energia vibracional superiores do estado fundamental, e o equilíbrio térmico é novamente alcançado pelo relaxamento vibracional da energia excedente. A existência do Deslocamento de Stokes é essencial para extrema sensibilidade das medidas de fluorescência. O deslocamento dos comprimentos de onda entre absorção e emissão faz com que seja possível o uso de filtros óticos precisos, capazes de bloquear a luz de excitação da fonte emissora, e assim o sinal de emissão pode ser observado em um sistema de baixo ruído.

#### 2.4.2 Matriz Excitação – Emissão e Espectrofluorímetro

Outra forma comum de estudar os efeitos da fluorescência é através da matriz excitação – emissão (*Excitation-Emission Matrix – EEM*), onde os comprimentos de onda de excitação e emissão são varridos em ambos os módulos e apresentados na forma de um mapa tridimensional  $\lambda_{excitação} \times \lambda_{emissão} \times intensidade\ de\ fluorescência$ . Comparado com outros tipos de espectros por fluorescência, os resultados na forma EEM apresentam

uma detecção multidimensional muito mais compreensiva da fluorescência, especialmente para amostras complexas (RUTHERFORD et al., 2020). Um típico EEM pode ser visto na Figura 2.17, tanto na forma tridimensional, quando na forma de curvas de contorno. Uma grande variedade de informações pode ser extraídas das EEM, como intensidade, localização e distribuição dos picos, informação decorrente da decomposição espectral e informação relacionada a energia dos fótons. Dada a natureza dos dados, são geralmente necessárias análises quimiométricas complexas para pré-tratar e extrair informação útil das matrizes, nas quais se destacam as técnicas de PCA/PLS (FAASSEN; HITZMANN, 2015; RANZAN et al., 2017), PARAFAC (BAHRAM et al., 2006; GUIMET et al., 2004; RÍOS-REINA et al., 2017; SILVA et al., 2019), meta-heurísticas de seleção de variáveis (ASSAWAJARUWAN; REINALTER; HITZMANN, 2017; RANZAN et al., 2015) e redes neurais convolucionais (ITAKURA et al., 2018; RUTHERFORD et al., 2020).

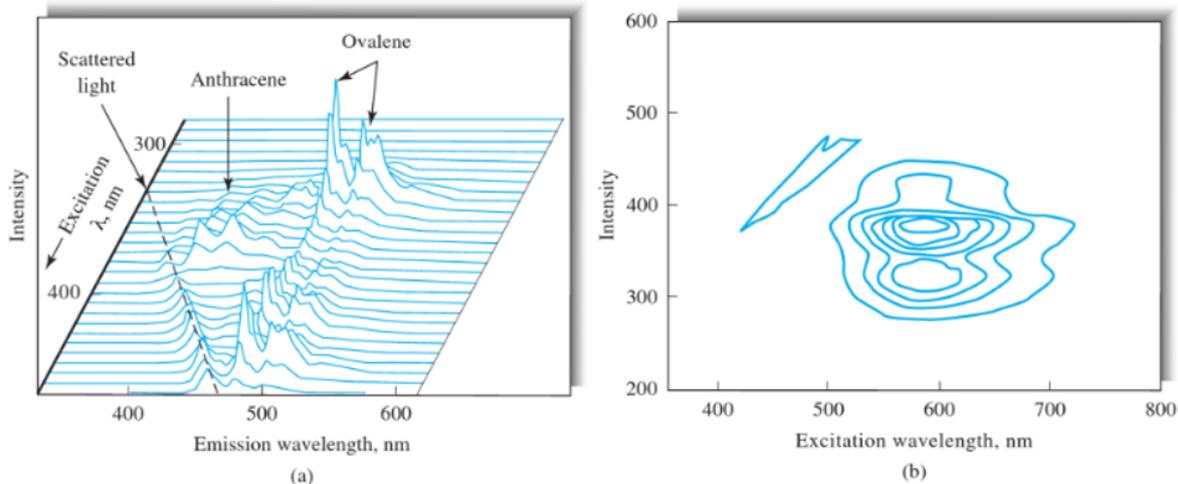


Figura 2.17. Matrizes excitação – emissão de fluorescência para uma mistura de antraceno e ovaleno (a), apresentada na forma tridimensional, e para 8-hidroxibenzopireno (b), apresentada como curvas de contorno. Fonte: (SKOOG; HOLLER; CROUCH, 2009).

O equipamento utilizado para obtenção de EEM de fluorescência se denomina espectrofluorímetro. Sua construção permite a obtenção tanto de espectros de excitação (comprimento de onda de emissão constante), quanto espectros de emissão (comprimento de onda de excitação constante), por meio do uso de dois monocromadores. Assim, é factível a construção dos mapas espectrais variando ambos os módulos sucessivamente. A Figura 2.18 apresenta a estrutura genérica de um espectrofluorímetro. A radiação oriunda da fonte passa por um monocromador para seleção do comprimento de onda de excitação. O feixe é então dividido em duas direções, uma que passará pela amostra, e outra que será atenuada e direcionada à fotomultiplicadora de referência. A radiação fluorescente da amostra, por sua vez, passa por um monocromador para seleção do comprimento de onda de emissão e é detectada pela fotomultiplicadora de emissão. A emissão de fluorescência da amostra se dá em todas as direções, mas é geralmente medida em ângulo reto com o feixe de excitação, para atenuar interferências. A eficiência dos monocromadores determinará a resolução do equipamento.

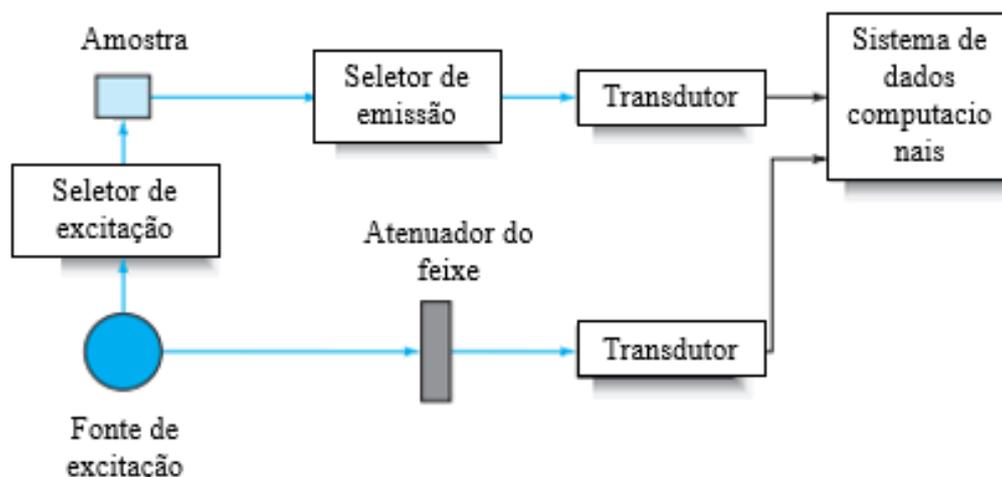


Figura 2.18. Representação genérica da estrutura de um espectrofluorímetro. Fonte: (SKOOG; HOLLER; CROUCH, 2009).

### 2.4.3 Supressão de Fluorescência - Quenching

A intensidade de fluorescência pode ser suprimida por uma ampla gama de processos, denominados de *quenching*. No *quenching* colisional, uma molécula excitada pode ser desativada ao colidir com outras moléculas da solução, denominados *quenchers*. Desta forma, ocorre troca de energia e o fluoróforo deixa de emitir fótons. Quanto maior a temperatura do meio, maiores as chances destas colisões acontecerem. Exemplos de *quenchers* comuns incluem oxigênio, halogênios, átomos pesados, aminas e moléculas eletro-deficientes. A fluorescência pode também ser suprimida pela formação de complexos não fluorescentes entre o fluoróforo e um *quencher*. Alguns fluoróforos também podem formar complexos consigo mesmos, como a formação de excímeros ou polímeros, alterando sua fluorescência a depender da concentração molecular. O pH também pode influenciar a fluorescência do meio, alterando as relações energéticas dos elétrons. Por fim, diversos mecanismos não moleculares podem causar *quenching*, como efeitos de filtro interno (absorção da emissão pelo próprio meio) e efeitos de barreira física, com atenuação da luz incidente (LAKOWICZ, 2006).

## 2.5 Referências

ABDULKADER, M. M. S.; GAJPAL, Y.; ELMKAWY, T. Y. Hybridized ant colony algorithm for the Multi Compartment Vehicle Routing Problem. **Applied Soft Computing**, v. 37, p. 196–203, 2015.

AMARBAYASGALAN, T.; JARGALSAIKHAN, B.; RYU, K. Unsupervised Novelty Detection Using Deep Autoencoders with Density Based Clustering. **Applied Sciences**, v. 8, 2018.

ASSAWAJARUWAN, S.; REINALTER, J.; HITZMANN, B. Comparison of methods for wavelength combination selection from multi-wavelength fluorescence spectra for on-line monitoring of yeast cultivations. **Analytical and Bioanalytical Chemistry**, v. 409, n. 3, p. 707–717, 2017.

BAHI, M.; BATOUCHE, M. **Deep semi-supervised learning for DTI prediction using large datasets and H2O-spark platform**. 2018 International Conference on Intelligent Systems and Computer Vision, ISCV 2018. **Anais...Institute of Electrical and Electronics Engineers Inc.**, 3 maio 2018

BAHRAM, M. et al. Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. **Journal of Chemometrics**, v. 20, n. 3–4, p. 99–105, 1

mar. 2006.

BALABIN, R. M.; SMIRNOV, S. V. Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data. **Analytica Chimica Acta**, v. 692, n. 1, p. 63–72, 2011.

BALSEIRO, S.; LOISEAU, I.; RAMONET, J. An Ant Colony algorithm hybridized with insertion heuristics for the Time Dependent Vehicle Routing Problem with Time Windows. **Computers & OR**, v. 38, p. 954–966, 2011.

BASHEER, I. A.; HAJMEER, M. Artificial neural networks: fundamentals, computing, design, and application. **Journal of Microbiological Methods**, v. 43, n. 1, p. 3–31, 2000.

BHATTACHARYYA, S. **Hybrid Metaheuristics**. [s.l.] WORLD SCIENTIFIC, 2018. v. 84

BLUM, C. et al. **Hybrid Metaheuristics: An Emerging Approach to Optimization**. [s.l.: s.n.].

BLUM, C. et al. Hybrid metaheuristics in combinatorial optimization: A survey. **Applied Soft Computing**, v. 11, n. 6, p. 4135–4151, 2011.

BOLÓN-CANEDO, V.; SÁNCHEZ-MAROÑO, N.; ALONSO-BETANZOS, A. A review of feature selection methods on synthetic data. **Knowledge and Information Systems**, v. 34, n. 3, p. 483–519, 2013.

BRITTAİN, H. G. Mid-Infrared Spectroscopy of Pharmaceutical Solids. In: **Profiles of Drug Substances, Excipients and Related Methodology**. [s.l.] Academic Press Inc., 2018. v. 43p. 321–358.

BRUCKER, P. NP-Complete operations research problems and approximation algorithms. **Zeitschrift für Operations Research**, v. 23, n. 3, p. 73–94, 1 jun. 1979.

CAMACHO, J.; PICÓ, J.; FERRER, A. Data understanding with PCA: Structural and Variance Information plots. **Chemometrics and Intelligent Laboratory Systems**, v. 100, n. 1, p. 48–56, 2010.

CAREY, P. Raman Crystallography, the Missing Link Between Biochemical Reactions and Crystallography. In: [s.l.] Springer, Dordrecht, 2014. p. 13–24.

CHANDRA MOHAN, B.; BASKARAN, R. A survey: Ant Colony Optimization based recent research and implementation on several engineering domain. **Expert Systems with Applications**, v. 39, n. 4, p. 4618–4627, 2012.

CHANDRASHEKAR, G.; SAHIN, F. A survey on feature selection methods. **Computers and Electrical Engineering**, v. 40, n. 1, p. 16–28, 1 jan. 2014.

CHARTE, D. et al. A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines. **Information Fusion**, v. 44, n. December 2017, p. 78–96, 2018.

CHI, Q. et al. A model predictive control approach with relevant identification in dynamic PLS framework. **Control Engineering Practice**, v. 22, n. 0, p. 181–193, 2014.

CHOLLET, F. **Deep Learning With Python**. [s.l.: s.n.]. v. 1

CIREŞAN, D. et al. Multi-column deep neural network for traffic sign classification. **Neural Networks**, v. 32, p. 333–338, ago. 2012.

CRAMER, J. A. et al. Ultra-low sulfur diesel classification with near-infrared spectroscopy and partial least squares. **Energy and Fuels**, v. 23, n. 2, p. 1132–1133, 2009.

CYBENKO, G. Approximation by superpositions of a sigmoidal function. **Mathematics of Control, Signals, and Systems**, v. 2, n. 4, p. 303–314, dez. 1989.

DORIGO, M.; BLUM, C. Ant colony optimization theory: A survey. **Theoretical Computer Science**, v. 344, n. 2–3, p. 243–278, 2005.

DORIGO, M.; GAMBARDELLA, L. M. Ant colony system: A cooperative learning approach to the traveling salesman problem. **IEEE Transactions on Evolutionary Computation**, v. 1, n. 1, p. 53–66, 1997.

DOSOVITSKIY, A.; BROX, T. Generating Images with Perceptual Similarity Metrics based on Deep Networks. **Advances in Neural Information Processing Systems**, p. 658–666, 2016.

DRESSELHAUS, M. S.; JORIO, A.; SAITO, R. Characterizing graphene, graphite, and carbon nanotubes by Raman spectroscopy. **Annual Review of Condensed Matter Physics**, 2010.

EFRON, B. et al. Least angle regression. **The Annals of Statistics**, v. 32, n. 2, p. 407–499, 2004.

ELLEN MACARTHUR FOUNDATION. **Towards the circular economy. Economic and business rationale for an accelerated transition.**

FAASSEN, S. M.; HITZMANN, B. Fluorescence Spectroscopy and Chemometric Modeling for Bioprocess Monitoring. p. 10271–10291, 2015.

FILZMOSER, P.; TODOROV, V. Review of robust multivariate statistical methods in high dimension. **Analytica Chimica Acta**, v. 705, n. 1–2, p. 2–14, 2011.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, v. 185, n. 0, p. 1–17, 1986.

GLOVER, F. Tabu Search—Part I. **ORSA Journal on Computing**, v. 1, n. 3, p. 190–206, 1 ago. 1989.

GLOVER, F. Tabu Search—Part II. **ORSA Journal on Computing**, v. 2, n. 1, p. 4–32, 1 fev. 1990.

GODOY, J. L.; VEGA, J. R.; MARCHETTI, J. L. Relationships between PCA and PLS-regression. **Chemometrics and Intelligent Laboratory Systems**, v. 130, p. 182–191, 15 jan. 2014.

GOLDBERG, D. E. **Genetic Algorithms in Search, Optimization and Machine Learning.** Boston: Kluwer Academic Publishers, 1989.

GUIDOTTI, R. et al. A Survey of Methods for Explaining Black Box Models. **ACM Computing Surveys**, v. 51, 2018.

GUIMET, F. et al. Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation–emission matrix fluorescence spectroscopy. **Analytica Chimica Acta**, v. 515, n. 1, p. 75–85, 2004.

GUYON, I. et al. Gene Selection for Cancer Classification using Support Vector Machines. **Machine Learning**, v. 46, n. 1, p. 389–422, 2002.

HASTIE, TREVOR, TIBSHIRANI, ROBERT, FRIEDMAN, J. **The Elements of Statistical Learning The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition.** [s.l: s.n.].

HE, K. et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 37, n. 9, p. 1904–1916, 1 set. 2015.

HEIYANTHUDUWAGE, M. A.; MOUNOURY, S.; KOVACEVIC, A. **Performance prediction methods for screw compressors.** Institution of Mechanical Engineers - 7th International Conference on Compressors and Their Systems 2011. **Anais...** Woodhead Publishing Limited, 1 jan. 2011

HEMMATEENEJAD, B.; MIRI, R.; ELYASI, M. A segmented principal component analysis—

regression approach to QSAR study of peptides. **Journal of Theoretical Biology**, v. 305, n. 0, p. 37–44, 2012.

HINTON, G. E.; SALAKHUTDINOV, R. R. Reducing the dimensionality of data with neural networks. **Science**, v. 313, n. 5786, p. 504–507, 28 jul. 2006.

HOERL, A. E.; KENNARD, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. **Technometrics**, v. 12, n. 1, p. 55–67, 1970.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. **Neural Networks**, v. 2, n. 5, p. 359–366, 1 jan. 1989.

HU, L. et al. Vis-NIR spectroscopy Combined with Wavelengths Selection by PSO Optimization Algorithm for Simultaneous Determination of Four Quality Parameters and Classification of Soy Sauce. **Food Analytical Methods**, v. 12, n. 3, p. 633–643, 2019.

IOFFE, S.; SZEGEDY, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2015.

ITAKURA, K. et al. Estimation of Citrus Maturity with Florescence Spectroscopy Using Deep Learning. **Horticulturae**, v. 5, n. 1, p. 2, 26 dez. 2018.

JACKSON, J. E. **A User's Guide to Principal Components**. Chichester: Wiley, 1991.

JES, F. et al. Deep Convolutional Autoencoders vs PCA in a Highly-Unbalanced Parkinson's Disease Dataset : A DaTSCAN Study. **Springer Nature**, p. 47–56, 2019.

JOHNSON, B.; XIE, Z. Classifying a high resolution image of an urban area using super-object information. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 83, p. 40–49, 2013.

JOLLIFFE, I. T. C. N.-S. C. I. Q. 5. J. 1986 H. L. Q. 5. J. 1986. **Principal component analysis**. New York: Springer-Verlag, 1986.

JONES, R. R. et al. **Raman Techniques: Fundamentals and Frontiers** **Nanoscale Research Letters** Springer New York LLC, , 1 dez. 2019. Disponível em: </pmc/articles/PMC6626094/?report=abstract>. Acesso em: 18 ago. 2020

KAROUI, R.; DE BAERDEMAEKER, J. **A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products** **Food Chemistry**, 2007.

KAVZOGLU, T. Chapter 33 - Object-Oriented Random Forest for High Resolution Land Cover Mapping Using Quickbird-2 Imagery. In: SAMUI, P.; SEKHAR, S.; BALAS, V. E. (Eds.). **Handbook of Neural Computation**. [s.l.] Academic Press, 2017. p. 607–619.

KENNEDY, J.; KENNEDY, J.; EBERHART, R. Particle swarm optimization. p. 4--1942, 1995.

KHAN, A. et al. A survey of the recent architectures of deep convolutional neural networks. **Artificial Intelligence Review**, p. 1–62, 21 abr. 2020.

KINGMA, D. P.; WELLING, M. **Auto-encoding variational bayes**. 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings. **Anais...** International Conference on Learning Representations, ICLR, 20 dez. 2014 Disponível em: <https://arxiv.org/abs/1312.6114v10>. Acesso em: 7 set. 2020

KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by simulated annealing. **Science**, 1983.

KISCHKAT, J. et al. Mid-infrared optical properties of thin films of aluminum oxide, titanium dioxide, silicon dioxide, aluminum nitride, and silicon nitride. **Applied Optics**, v. 51, n. 28, p. 6789–6798, 1 out. 2012.

KRIEG, T. Real-time monitoring of continuous fermentation by Raman spectroscopy. n. September, p. 30, 2014.

LAKOWICZ, J. R. C. N.-S. C. I. Q. F. L. 2006 N. E. T. M. I. T. A. O. \*\*SEE U. H. L. Q. F. L. 2006 H. L. C. Q. F. L. 2006. **Principles of fluorescence spectroscopy**. 3rd. ed. New York: Springer, 2006.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. **Convolutional networks and applications in vision**. ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems. **Anais...2010**

LEE, C.-Y.; GALLAGHER, P. W.; TU, Z. Generalizing Pooling Functions in Convolutional Neural Networks: Mixed, Gated, and Tree. **Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016**, p. 464–472, 29 set. 2015.

LEIJNEN, S.; VAN VEEN, F. The Neural Network Zoo †. **Proceedings 2020, Vol. 47, Page 9**, v. 47, n. 1, p. 9, 12 maio 2020.

LI, X. Y.; TIAN, P.; LEUNG, S. C. H. An ant colony optimization metaheuristic hybridized with tabu search for open vehicle routing problems. **Journal of the Operational Research Society**, v. 60, n. 7, p. 1012–1025, 2009.

LIU, X.; DENG, Z.; YANG, Y. Recent progress in semantic image segmentation. **Artificial Intelligence Review**, v. 52, n. 2, p. 1089–1106, 15 ago. 2019.

MANNING-DAHAN, T. PCA and Autoencoders. 2017.

MASCI, J. et al. **Stacked convolutional auto-encoders for hierarchical feature extraction**. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). **Anais...Springer**, Berlin, Heidelberg, 2011Disponível em: <[https://link.springer.com/chapter/10.1007/978-3-642-21735-7\\_7](https://link.springer.com/chapter/10.1007/978-3-642-21735-7_7)>. Acesso em: 7 set. 2020

MELO, P. M. A. S. Conceitos básicos da meta-heurística Tabu Search. **Faculdade de Engenharia da Universidade do Porto**, 2008.

MISRA, S. et al. **An Autoencoder Based Model for Detecting Fraudulent Credit Card Transaction**. Procedia Computer Science. **Anais...Elsevier B.V.**, 1 jan. 2020

MOHAPATRA, S.; PATRA, D.; SATPATHY, S. An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. **Neural Computing and Applications**, v. 24, n. 7, p. 1887–1904, 2014.

MURPHY, K. R. et al. Fluorescence spectroscopy and multi-way techniques. **PARAFAC. Analytical Methods**, v. 5, n. 23, p. 6557–6566, 2013.

MYSZKOWSKI, P. B. et al. Hybrid ant colony optimization in solving multi-skill resource-constrained project scheduling problem. **Soft Computing**, v. 19, n. 12, p. 3599–3619, 2015.

O. DUDA, R.; E. HART, P.; G.STORK, D. Pattern Classification / R.O. Duda, P.E. Hart, D.G. Stork. 2019.

OSTERTAGOVÁ, E. Modelling using polynomial regression. 2012.

PALUSZKIEWICZ, C. et al. **FT-IR study of montmorillonite-chitosan nanocomposite materials**. Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy. **Anais...15 ago. 2011**

PES, B. Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains. **Neural Computing and Applications**, 2019.

PESSOA, C. M. et al. Development of Ant Colony Optimization (ACO) Algorithms Based on Statistical Analysis and Hypothesis Testing for Variable Selection. **IFAC-PapersOnLine**, v. 48, n. 8, p. 900–905, 2015.

RANZAN, C. et al. Wheat flour characterization using NIR and spectral filter based on Ant Colony Optimization. **Chemometrics and Intelligent Laboratory Systems**, v. 132, n. 0, p. 133–140, 2014.

RANZAN, C. et al. Sulfur Determination in Diesel using 2D Fluorescence Spectroscopy and Linear Models. **IFAC-PapersOnLine**, v. 48, n. 8, p. 415–420, 2015.

RANZAN, L. et al. Classification of Diesel Fuel Using Two-Dimensional Fluorescence Spectroscopy. **Energy & Fuels**, v. 31, n. 9, p. 8942–8950, 2017.

REID, L. M.; O'DONNELL, C. P.; DOWNEY, G. Recent technological advances for the determination of food authenticity. **Trends in Food Science and Technology**, v. 17, n. 7, p. 344–353, 2006.

REUNANEN, J.; GUYON, I.; ELISSEEFF, A. Overfitting in Making Comparisons Between Variable Selection Methods. **Journal of Machine Learning Research**, v. 3, p. 1371–1382, 2003.

RÍOS-REINA, R. et al. Characterization and authentication of Spanish PDO wine vinegars using multidimensional fluorescence and chemometrics. **Food Chemistry**, v. 230, p. 108–116, 2017.

ROBERTSON, A. L. et al. **Using a partial least squares (PLS) method for estimating cyanobacterial pigments in eutrophic inland waters**. Remote Sensing and Modeling of Ecosystems for Sustainability VI. **Anais...SPIE**, 20 ago. 2009

ROBNIK-ŠIKONJA, M.; KONONENKO, I. Theoretical and Empirical Analysis of Relief and RRelief. **Machine Learning**, v. 53, n. 1, p. 23–69, 2003.

RODRIGUES E SILVA, S.; SCHIMIDT, F. Redução de variáveis de entrada de redes neurais artificiais a partir de dados de análise de componentes principais na modelagem de oxigênio dissolvido. **Química Nova**, v. 39, n. 3, p. 273–278, 1 abr. 2016.

RUTHERFORD, J. W. et al. Excitation emission matrix fluorescence spectroscopy for combustion generated particulate matter source identification. **Atmospheric Environment**, v. 220, p. 117065, 2020.

RYGULA, A. et al. **Raman spectroscopy of proteins: A review** *Journal of Raman Spectroscopy*, 2013.

SAKURADA, M.; YAIRI, T. Anomaly detection using autoencoders with nonlinear dimensionality reduction. **Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis**, p. 4, 2014.

ŠAŠIĆ, S. **Pharmaceutical Applications of Raman Spectroscopy**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2007.

SCHERER, D.; MÜLLER, A.; BEHNKE, S. **Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition**. [s.l.: s.n.]. Disponível em: <<http://www.ais.uni-bonn.de>>. Acesso em: 7 set. 2020.

SCHWENK, H.; BENGIO, Y. Boosting Neural Networks. **Neural Computation**, v. 12, n. 8, p. 1869–1887, 2000.

SEBBEN, J. A. et al. Development of a quantitative approach using Raman spectroscopy for carotenoids determination in processed sweet potato. **Food Chemistry**, v. 245, p. 1224–1231, 2018.

SERBENCU, A.; MINZU, V. **Hybridized Ant Colony System for Tasks to Workstations Assignment**. 2016 IEEE Symposium Series on Computational Intelligence (SSCI).

Anais...2016

SETTLE, F. **Handbook of Instrumental Techniques for Analytical Chemistry**. Upper Saddle River: Prentice-Hall, 1997.

SHINZAWA, H. et al. Accelerated weathering-induced degradation of poly (lactic acid) fiber studied by near-infrared (NIR) hyper spectral imaging. **Applied Spectroscopy**, v. 66, n. 4, p. 470–474, 2012.

SILVA, A. C. et al. Green chemistry method based on PARAFAC EEM data modeling for Benzo[a]pyrene quantitation in distilled spirit. **Journal of the Brazilian Chemical Society**, v. 30, n. 2, p. 398–405, 2019.

SILVA, A. C. DA et al. Two-dimensional linear discriminant analysis for classification of three-way chemical data. **Analytica Chimica Acta**, v. 938, p. 53–62, 2016.

SILVA, J. I. S. DA; SECCHI, A. R. AN APPROACH TO OPTIMIZE COSTS DURING ULTRA-LOW HYDRODESULFURIZATION OF A BLEND CONSISTING OF DIFFERENT OIL STREAMS. **Brazilian Journal of Chemical Engineering**, v. 35, p. 1293–1304, 2018.

SIMONYAN, K.; ZISSERMAN, A. **VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION**. [s.l: s.n.]. Disponível em: <<http://www.robots.ox.ac.uk/>>. Acesso em: 24 ago. 2020.

SINELLI, N. et al. Evaluation of quality and nutraceutical content of blueberries (*Vaccinium corymbosum* L.) by near and mid-infrared spectroscopy. **Postharvest Biology and Technology**, v. 50, n. 1, p. 31–36, out. 2008.

SKOOG, D. A.; HOLLER, J. F.; CROUCH, S. R. **Princípios de Análise Instrumental**. 6ª edição ed. [s.l.] Bookman, 2009.

SORIANO-DISLA, J. M. et al. **The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties** **Applied Spectroscopy Reviews**, 17 fev. 2014.

SRINIVAS, S.; SUBRAMANYA, A.; BABU, R. V. Training Sparse Neural Networks. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops**, v. 2017- July, p. 455–462, 21 nov. 2016.

SRIVASTAVA, N.; MANSIMOV, E.; SALAKHUTDINOV, R. Unsupervised Learning of Video Representations using LSTMs. **32nd International Conference on Machine Learning, ICML 2015**, v. 1, p. 843–852, 16 fev. 2015.

STÜTZLE, T.; LÓPEZ-IBÁÑEZ, M.; DORIGO, M. A Concise Overview of Applications of Ant Colony Optimization. In: [s.l: s.n.].

TANG, J.; ALELYANI, S.; LIU, H. Feature selection for classification: A review. In: **Data Classification: Algorithms and Applications**. [s.l: s.n.]. p. 37–64.

TIPPING, M. E.; BISHOP, C. M. Probabilistic Principal Component Analysis. **Journal of the Royal Statistical Society. Series B (Statistical Methodology)**, v. 61, n. 3, p. 611–622, 1999.

VALENTI, B. et al. Infrared spectroscopic methods for the discrimination of cows' milk according to the feeding system, cow breed and altitude of the dairy farm. **International Dairy Journal**, v. 32, n. 1, p. 26–32, 2013.

VERMA, K.; KUMAR SINGH, P. An Insight to Soft Computing based Defect Prediction Techniques in Software. **International Journal of Modern Education and Computer Science**, v. 7, n. 9, p. 52–58, 8 set. 2015.

VINCENT, P. et al. Extracting and composing robust features with denoising autoencoders. **Proceedings of the 25th international conference on Machine learning - ICML '08**, p. 1096–1103, 2008.

WANG, Y. et al. An Improved Ant Colony Optimization algorithm to the Periodic Vehicle

Routing Problem with Time Window and Service Choice. **Swarm and Evolutionary Computation**, v. 55, p. 100675, 1 jun. 2020.

WETZEL, S. J. Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. **Physical Review E**, v. 96, n. 2, p. 22140, 2017.

WIDROW, B.; LEHR, M. A. 30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation. **Proceedings of the IEEE**, v. 78, n. 9, p. 1415–1442, 1990.

WILSON, J. H.; ZHANG, C.; KOVACS, J. M. Separating crop species in Northeastern Ontario using hyperspectral data. **Remote Sensing**, v. 6, n. 2, p. 925–945, 2014.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, 2001.

XIAOBO, Z. et al. Variables selection methods in near-infrared spectroscopy. **Analytica Chimica Acta**, v. 667, n. 1, p. 14–32, 2010.

XU, L. et al. Variable-weighted PLS. **Chemometrics and Intelligent Laboratory Systems**, v. 85, n. 1, p. 140–143, 2007.

YADAV, L. D. S.; YADAV, L. D. S. Infrared (IR) Spectroscopy. In: **Organic Spectroscopy**. [s.l.] Springer Netherlands, 2005. p. 52–106.

YAMASHITA, R. et al. **Convolutional neural networks: an overview and application in radiology** Insights into Imaging Springer Verlag, , 1 ago. 2018. Disponível em: <<https://link.springer.com/articles/10.1007/s13244-018-0639-9>>. Acesso em: 7 set. 2020

YU, C.; YAO, W. Robust linear regression: A review and comparison. **Communications in Statistics - Simulation and Computation**, v. 46, n. 8, p. 6261–6282, 14 set. 2017.

YU, D. et al. Near-infrared fluorescent probe for detection of thiophenols in water samples and living cells. **Analytical Chemistry**, v. 86, n. 17, p. 8835–8841, 2014.

ZEILER, M. D.; FERGUS, R. **Visualizing and Understanding Convolutional Networks**. (D. Fleet et al., Eds.) Computer Vision – ECCV 2014. **Anais...** Cham: Springer International Publishing, 2014

ZHANG, G.; LI, H. **Effectiveness of Scaled Exponentially-Regularized Linear Units (SERLUs)**. [s.l.: s.n.].

ZHANG, P. (ED.). Front Matter. In: **Advanced Industrial Control Technology**. Oxford: William Andrew Publishing, 2010. p. iii.

ZHANG, X.; TANG, L. **A New Hybrid Ant Colony Optimization Algorithm for the Traveling Salesman Problem**. (D.-S. Huang et al., Eds.) Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence. **Anais...** Berlin, Heidelberg: Springer Berlin Heidelberg, 2008

ZHENG, C. et al. Feature selection for high-dimensional integrated data. In: **Proceedings of the 2012 SIAM International Conference on Data Mining**. Proceedings. [s.l.] Society for Industrial and Applied Mathematics, 2012. p. 1141–1150.

ZOU, J.; ZHANG, J.; JIANG, P. Credit Card Fraud Detection Using Autoencoder Neural Network. 30 ago. 2019.

# Capítulo 7 – Conclusões e Trabalhos Futuros

## 7.1 Conclusões

O estudo de modelos baseados em dados provenientes de espectroscopia por fluorescência se mostrou de grande valia para diversas aplicações no universo industrial. Apesar de amplamente pesquisados, ainda se identifica baixa penetrabilidade de sensores óticos em operações de larga escala nos ramos adjacentes da engenharia química. Um dos possíveis motivos se deve a complexidade no tratamento destes dados, que requer uma gama de técnicas especializadas para transformar a informação contida nos espectros em conhecimento útil de processo.

Neste âmbito, as pesquisas desenvolvidas neste trabalho são continuidade natural do desenvolvimento de sensores de processo personalizados, baseados em técnicas óticas. As vantagens associadas a este tipo de sensor são especialmente relevantes para processos químicos e biológicos. A coleta espectral é realizada de maneira rápida, não intrusiva, não destrutiva e não requer tratamento da amostra. Além disso, são técnicas sensíveis e com capacidade de correlacionar um único espectro a uma miríade de propriedades do meio. Em especial para espectroscopia por fluorescência, uma quantidade considerável dos analitos de interesse nestas áreas são fluoróforos naturais.

A metodologia para seleção de variáveis explicativas proposta neste trabalho, AnTSbe, se mostrou útil na área de quimiometria, resultando em modelos com capacidade preditiva superiores a metodologias consolidadas, como PLS, e também a versões anteriores de algoritmos Colônia de Formigas, quando comparados utilizando a mesma base de dados (quantificação de enxofre total em amostras de diesel combustível). Além de apresentar as variáveis selecionadas para o modelo com menores erros de predição, o algoritmo expõe, através das trilhas de feromônios, um ordenamento das variáveis de entrada com maior correlação com as saídas de interesse. Essa informação é proveitosa para técnicas de filtragem de dados, especialmente vantajosas para situações onde a base de dados é extensa e altamente colinear.

A introdução do conceito de lista tabu se mostrou benéfico para metodologia, evitando, no melhor dos casos, uma repetição redundante de 20% de todos os subgrupos testados pelo algoritmo. A real influência da hibridização com Busca Tabu é ainda maior do que a simples contagem de vezes que o algoritmo foi impedido de selecionar um grupo proibido: sem esta restrição, a ciclagem de subgrupos elevaria a quantidade de feromônios nestas variáveis com alta probabilidade de seleção, causando uma possível estagnação precoce. A introdução de subgrupos na lista de proibição garante, também, que a evaporação da trilha ao final de cada iteração melhor regule a relação de feromônios entre as variáveis de entrada.

A natureza dos dados de fluorescência acarreta em um detrimento nos benefícios da hibridização, uma vez que pares de fluorescência adjacentes carregam, muitas vezes, informação similar. Quanto menor o número de variáveis de entrada, e mais independentes elas forem, maior é a vantagem no uso da lista tabu, evitando desperdício de tempo computacional e melhorando os mecanismos de prospecção e exploração, tornando a metodologia mais consistente e eficiente.

A introdução da expansão polinomial e combinatória das variáveis de entrada, no interior da rotina de otimização, tornou possível a captura de uma maior gama de informação dos dados, melhorando as métricas de predição para modelos que não apresentavam bons resultados apenas com a base original. Este efeito pôde ser observado no estudo de caso dos modelos locais para Diesel S100, e é fortemente evidenciado para as predições de compostos fenólicos em cachaça envelhecida, assim como para inferidores de estado nas correntes da unidade de separação de propano/propeno. A introdução da expansão acarretou em uma diminuição de até 50% dos erros médios de predição de fenólicos totais, quando não consideradas amostras comerciais de cachaça. Para a predição dos perfis de concentração e impurezas da USPP, o uso da expansão de bases foi essencial para construção dos modelos, sendo inclusive desconsiderados os modelos sem seu uso.

O uso da expansão de bases, apesar de benéfico para a maioria dos modelos ajustados, evidenciou uma fragilidade na técnica. A expansão da ordem dos dados também ocasiona um aumento da sensibilidade dos modelos à distúrbios não medidos durante a etapa de calibração, fazendo com que os mesmos possuam menor poder de generalização. Isto pode ser observado quando amostras de cachaça comercial, bem distintas das amostras envelhecidas em laboratório, participavam dos subgrupos de teste para os modelos preditivos de fenólicos totais, com erros muito superiores as demais amostras.

O desenvolvimento da metodologia Adaptada AnTSbe, com introdução do conceito de Par Delta, obteve sucesso em aperfeiçoar a performance preditiva de modelos que utilizam expansão de base. O uso de um par de fluorescência como regulador do meio aprimorou a capacidade dos modelos ajustados em predizer amostras que apresentavam distinções significativas do grupo de calibração. Na predição de compostos fenólicos totais em amostras de cachaça envelhecida, o uso da metodologia adaptada apresentou erros até três vezes menores, em comparação a metodologia original, quando considerado o subgrupo de teste contendo cachaças comerciais. Excluindo as amostras comerciais, ambas metodologias apresentaram resultados satisfatórios similares nos subgrupos de teste.

A caracterização dos perfis de compostos fenólicos em processos de envelhecimento de cachaças em barris de madeira propiciou múltiplas contribuições. Primeiramente, não é corriqueiro encontrar na literatura trabalhos que acompanham processos de envelhecimento com a coleta frequente de espectros por fluorescência como o

apresentado. Isso se deve principalmente a dinâmica temporal necessária, tendo a pesquisa percorrido mais de três anos para sua conclusão. Este acompanhamento identificou as radicais mudanças nos espectros que acontecem ao longo do envelhecimento, como o deslocamento evidente dos picos de fluorescência. Outro fato interessante a ser pontuado foram as diferenças nas intensidades médias de fluorescência dos espectros entre os diferentes envelhecimentos.

Tanto os modelos ajustados com a metodologia AnTSbe original, quanto com a adaptada, foram capazes de prever satisfatoriamente a concentração de fenóis ao longo dos envelhecimentos. A análise dos dados e dos resultados mostrou necessário o ajuste de modelos locais que separassem os envelhecimentos CA2 e CA3 das amostras CA1, uma vez que tanto os espectros quanto as concentrações dos dois grupos eram significativamente distintas. A metodologia adaptada, por sua vez, foi mais eficiente em prever amostras comerciais de cachaça envelhecida, onde o algoritmo original foi falho. Ambas abordagens se mostraram úteis como base para o desenvolvimento de sensores capazes de refinar processos de envelhecimento, com acompanhamento do perfil fenólico, do tempo de vida útil dos barris, e um maior controle sobre a padronização do produto final.

Com relação a caracterização de bioprocessos, o uso de redes neurais convolucionais residuais se mostrou eficiente em representar de forma fiel as concentrações de interesse no meio ao longo dos processos fermentativos. Para as três fermentações estudadas, os erros de predição foram mínimos, com  $R^2$  superiores a 0.99. O uso de redes convolucionais com dados bidimensionais de fluorescência apresentou alto poder preditivo, e tem potencial para ser aplicado em outras pesquisas quimiométricas. Vale ressaltar que a qualidade do modelo neural está intimamente ligada a abundância de dados que o acoplamento de um espectrofluorômetro no sistema fermentativo possibilitou.

A metodologia de triagem e detecção de inconformidades baseada em redes autoencoder também se mostrou válida. A rede calibrada foi capaz de reconstruir os espectros de entrada dos sistemas fermentativos com erros médios inferiores a 0.65%. Utilizando a rede calibrada, e acompanhando o erro de reconstrução de um novo conjunto de pontos modificados artificialmente, foi possível identificar as amostras que receberam erros significativos, e também quantificar quanto essas amostras divergiam dos dados de calibração originais, baseado na magnitude do erro. Corroborando com a hipótese levantada, as amostras identificadas como anormais apresentaram erros médios de predição superiores aos esperados, quando preditos pela rede residual calibrada com a mesma base de dados original. O uso da metodologia pode atenuar um dos desafios no emprego de modelos empíricos, qualificando não supervisionadamente se novos dados a serem rodados por um modelo se enquadram nas características dos dados utilizados para sua calibração. Um influxo de novas leituras com grandes erros de reconstrução pode sinalizar algum problema na coleta de dados, ou então que o novo estado do sistema não é mais bem representado pelo grupo de calibração e que o modelo deve ser reajustado.

A expansão da aplicação da metodologia AnTSbe para dados industriais também apresentou bons resultados. A partir dos dados simulados baseados em uma unidade real de separação propano/propeno, foi possível prever as concentrações dos compostos de interesse em cada uma das três colunas de separação com precisão satisfatória. Os modelos ajustados para predição das impurezas nas correntes principais de cada coluna

(fundamental interesse do trabalho) demandaram o uso de modelos locais especificados por faixas de concentração, e atingiram predições com erros absolutos médios inferiores a 5%. Para a terceira coluna, onde é necessário o controle da especificação do produto final, a metodologia AnTSbe foi capaz de lidar com a grande quantidade de dados de entrada expandidos (superiores a 1500) de maneira mais eficiente que as demais técnicas apresentadas (Regressão *Ridge* e *LassoLars*).

De maneira geral, pode-se concluir com este trabalho que as metodologias propostas apresentam boa viabilidade para aplicação na caracterização de processos industriais. A combinação de espectroscopia por fluorescência bidimensional com técnicas de seleção de variáveis e redes neurais podem ser a base para construção de novos sensores capazes de mensurar rapidamente informações que hoje são dependentes de testes laboratoriais, abrindo possibilidades para novas estratégias de controle de processos. O uso de redes autoencoder, por sua vez, pode ter sua aplicação expandida dentro do ramo quimiométrico, evitando predições falhas e apurando a confiabilidade dos sistemas preditivos.

## 7.2 Sugestões para Trabalhos Futuros

As metodologias desenvolvidas neste trabalho para caracterização de processos a partir de dados de espectroscopia por fluorescência bidimensional se mostraram satisfatórias no intuito de selecionar e transcrever a informação espectral em conhecimento útil de processo. Apesar disso, diversas são as possibilidades de evolução e melhoria das técnicas.

A metodologia AnTSbe, ainda que apresentada majoritariamente neste trabalho com dados provenientes de espectroscopia por fluorescência, pode ser utilizada para a otimização de modelos e seleção de variáveis com quaisquer tipos de dados. Além disso, as matrizes de dados de espectroscopia podem ser acrescidas de outras fontes de informação, como, por exemplo, medidas de pH, temperatura e pressão, que já estão normalmente disponíveis no sistema. O algoritmo iria avaliar, então, se existe benefício em adicionar estas variáveis nos modelos de predição. Dados de espectroscopia vibracional também podem ser incluídos, uma vez que sua aquisição e benefícios muito se assemelham a fluorescência, aumentando a gama de informações disponíveis.

Outras hibridizações também podem ser apresentadas para ampliar as funcionalidades do algoritmo. Nos casos discutidos, a seleção de variáveis é realizada de forma discreta. Uma rotina interna de otimização utilizando meta-heurísticas como Enxame de Partículas pode ser sugerida para solução de problemas contínuos, inclusive com a proposição de ajuste de modelos não lineares.

O uso de redes residuais se provou eficiente para a caracterização de bioprocessos. Porém, sua aplicação está atrelada a disponibilidade de dados. Este é mais um incentivo para aprofundar estudos que apliquem coleta *online* de dados espectrais, com o acoplamento de espectrofluorômetros diretamente no sistema a ser estudado. O influxo de uma maior quantidade de dados espectrais é o caminho natural para confirmar de forma prática as funcionalidades teóricas apresentadas neste trabalho para dados laboratoriais, como para caracterização de produtos petroquímicos.

Com relação a aplicação de redes autoencoder para identificação de inconformidades, é sugerida a ampliação de estudos com o emprego em diferentes casos de estudos, para

explorar sua funcionalidade. A metodologia, em teoria, pode ser associada a modelos quimiométricos existentes de diferentes áreas, principalmente aquelas que possuem grande variabilidade de matéria-prima ou processual.

Considerando as características dos dados de fluorescência, outras abordagens para redes autoencoder podem ser aplicadas. Redes autoencoder do tipo *denoising* (geralmente construídas com arquitetura de redes convolucionais) podem ser utilizadas para, ao invés de identificar anormalidades, corrigir automaticamente espectros falhos e remover ruído da matriz de dados espectrais. Redes do tipo *variational* também podem ser úteis. Estas redes são análogas as redes autoencoder tradicionais, porém, ajustam um modelo generativo a partir da seção de decodificação, capaz de criar novos dados de entrada com características específicas. Essas redes são muito utilizadas para mesclar características dos dados. Esta abordagem é proveitosa quando a base de dados é restrita, tornando possível a criação computacional de uma plethora de espectros com ampla variabilidade, que seriam custosos ou impraticáveis laboratorialmente, a partir de uma quantidade reduzida de amostras chave.

Além das modificações conceituais sugeridas, é proposto a aplicação das metodologias para caracterização de outros processos com base em dados de espectroscopia, propiciando um melhor aproveitamento das ferramentas desenvolvidas e incentivando o progresso de técnicas óticas. Neste caso, aventamos o enfoque em sensores customizados, com coleta inteligente de dados (baseada nos princípios e vantagens da seleção de variáveis explicativas, amplamente debatidas neste trabalho). Com menor custo associado e uma maior especificidade, o uso destes sensores pode propiciar a automação de processos que atualmente não viabilizam economicamente o investimento em instrumentos óticos especializados.