Doctoral Thesis

Multiview pattern recognition methods for data visualization, embedding and clustering

Ph.D. Student: Samir Kanaan Izquierdo
Ph.D. Advisor: Alexandre Perera Lluna

Thesis Presented in Partial Fulfillment of the Requirements for the Degree of Ph.D.

Biomedical Engineering Ph.D. Program

Automatic Control Department

Universitat Politècnica de Catalunya

Barcelona, 2017

Abstract

Multiview data is defined as data for whose samples there exist several different data views, i.e. different data matrices obtained through different experiments, methods or situations. Multiview dimensionality reduction methods transform a high-dimensional, multiview dataset into a single, low-dimensional space or projection. Their goal is to provide a more manageable representation of the original data, either for data visualization or to simplify the following analysis stages. Multiview clustering methods receive a multiview dataset and propose a single clustering assignment of the data samples in the dataset, considering the information from all the input data views.

The main hypothesis defended in this work is that using multiview data along with methods able to exploit their information richness produces better dimensionality reduction and clustering results than simply using single views or concatenating all views into a single matrix.

Consequently, the objectives of this thesis are to develop and test multiview pattern recognition methods based on well known single-view dimensionality reduction and clustering methods. Three multiview pattern recognition methods are presented: multiview t-distributed stochastic neighbourhood embedding (MV-tSNE), multiview multimodal scaling (MV-MDS) and a novel formulation of multiview spectral clustering (MVSC-CEV). These methods can be applied both to dimensionality reduction tasks and to clustering tasks.

The MV-tSNE method computes a matrix of probabilities based on distances between samples for each input view. Then it merges the different probability matrices using results from expert opinion pooling theory to get a common matrix of probabilities, which is then used as reference to build a low-dimensional projection of the data whose probabilities are similar.

The MV-MDS method computes the common eigenvectors of all the normalized distance matrices in order to obtain a single low-dimensional space that embeds the essential information from all the input spaces, avoiding redundant information to be included.

The MVSC-CEV method computes the symmetric Laplacian matrices of the similarity matrices of all data views. Then it generates a single, low-dimensional representation of the input data by computing the common eigenvectors of the Laplacian matrices, obtaining a projection of the data that embeds the most relevant information of the input data views, also avoiding the addition of redundant information.

A thorough set of experiments has been designed and run in order to compare the proposed methods with their single view counterpart. Also, the proposed methods have been compared with all the available results of equivalent methods in the state of the art. Finally, a comparison between the three proposed methods is presented in order to provide guidelines on which method to use for a given task.

MVSC-CEV consistently produces better clustering results than other multiview methods in the state of the art. MV-MDS produces overall better results than the reference methods in dimensionality reduction

experiments. MV-tSNE does not excel on any of these tasks. As a consequence, for multiview clustering tasks it is recommended to use MVSC-CEV, and MV-MDS for multiview dimensionality reduction tasks.

Although several multiview dimensionality reduction or clustering methods have been proposed in the state of the art, there is no software implementation available. In order to compensate for this fact and to provide the community with a potentially useful set of multiview pattern recognition methods, an R software package containg the proposed methods has been developed and released to the public.

Unesco codes: 120304, 120804, 120110

Resumen

Los datos multivista se definen como aquellos datos para cuyas muestras existen varias vistas de datos distintas, es decir diferentes matrices de datos obtenidas mediante diferentes experimentos, métodos o situaciones. Los métodos multivista de reducción de la dimensionalidad transforman un conjunto de datos multivista y de alta dimensionalidad en un único espacio o proyección de baja dimensionalidad. Su objetivo es producir una representación más manejable de los datos originales, bien para su visualización o para simplificar las etapas de análisis subsiguientes. Los métodos de agrupamiento multivista reciben un conjunto de datos multivista y proponen una única asignación de grupos para sus muestras, considerando la información de todas las vistas de datos de entrada.

La principal hipótesis defendida en este trabajo es que el uso de datos multivista junto con métodos capaces de aprovechar su riqueza informativa producen mejores resultados en reducción de la dimensionalidad y agrupamiento frente al uso de vistas únicas o la concatenación de varias vistas en una única matriz.

Por lo tanto, los objetivos de esta tesis son desarrollar y probar métodos multivista de reconocimiento de patrones basados en métodos univista reconocidos. Se presentan tres métodos multivista de reconocimiento de patrones: proyección estocástica de vecinos multivista (MV-tSNE), escalado multidimensional multivista (MV-MDS) y una nueva formulación de agrupamiento espectral multivista (MVSC-CEV). Estos métodos pueden aplicarse tanto a tareas de reducción de la dimensionalidad como de agrupamiento.

MV-tSNE calcula una matriz de probabilidades basada en distancias entre muestras para cada vista de datos. A continuación combina las matrices de probabilidad usando resultados de la teoría de combinación de expertos para obtener una matriz común de probabilidades, que se usa como referencia para construir una proyección de baja dimensionalidad de los datos.

MV-MDS calcula los vectores propios comunes de todas las matrices normalizadas de distancia para obtener un único espacio de baja dimen-

sionalidad que integre la información esencial de todos los espacios de entrada, evitando información redundante.

MVSC-CEV calcula las matrices Laplacianas de las matrices de similitud de los datos. A continuación genera una única representación de baja dimensionalidad calculando los vectores propios comunes de las Laplacianas. Así obtiene una proyección de los datos que integra la información más relevante y evita añadir información redundante.

Se ha diseñado y ejecutado una batería de experimentos completa para comparar los métodos propuestos con sus equivalentes univista. Además los métodos propuestos se han comparado con los resultados disponibles en la literatura. Finalmente, se presenta una comparación entre los tres métodos para proporcionar orientaciones sobre el método más adecuado para cada tarea.

MVSC-CEV produce mejores agrupamientos que los métodos equivalentes en la literatura. MV-MDS produce en general mejores resultados que los métodos de referencia en experimentos de reducción de la dimensionalidad. MV-tSNE no destaca en ninguna de esas tareas. Consecuentemente, para agrupamiento multivista se recomienda usar MVSC-CEV, y para reducción de la dimensionalidad multivista MV-MDS.

Aunque se han propuesto varios métodos multivista en la literatura, no existen programas disponibles públicamente. Para remediar este hecho y para dotar a la comunidad de un conjunto de métodos potencialmente útil, se ha desarrollado un paquete de programas en R y se ha puesto a disposición del público.

Códigos Unesco: 120304, 120804, 120110

To J&B, with love.

Acknowledgements

First I would like to express my gratitude to my advisor Dr. Alexandre Perera for his support in the creation of this thesis. As my advisor, he has shown me the way to go when I was disoriented, he has encouraged me when I was stranded in my research, and he has also taught me to realize that, eventually, things get working. As a researcher I have learnt a lot from him as well: he is careful with his work, always willing to learn new things, and has a keen eye for spotting flaws! Finally, he is a kind person and it has been a pleasure to work with him all these years.

In second place I want to thank my colleagues at my workplace, the EEBE (former EUETIB), for their support, advice and encouragement during all these years. Very specially I want to thank Dr. Gerard Escudero for his generous help assuming so many common tasks to relieve me and let me have more spare time to devote to my thesis.

I also want to thank Drs. Pedro Gomis and Montserrat Vallverdú for introducing me to my advisor. Seemingly they had a perfect notion of what I could and wanted to do and I think they made a perfect match.

Raimon Jané and Noemí Zapata also deserve my gratitude for their help with all the academic requirements and the paperwork. At times I have caused them extra work and I am very thankful for this.

Of course I am not forgetting my fellow mates, with whom I have shared all these years of work, exchange of ideas and hopes. And some mountain hiking too! Thank you Andrey, Francesc, Helena, Maria, Gio, Sergi and many others along the way.

Finally I want to thank my family and friends for supporting (and standing) me during all these long years. Hopefully they (and me) will be finally relieved from the old question “How’s your thesis going?”. It is about time to move on to new questions.

Contents

List of Tables

List of Figures

Chapter 1

Multiview unsupervised pattern recognition methods

1.1 Introduction

1.1.1 Multiview data

The development of information and communication technologies has led to ever-increasing data production in most areas of human activity. The difference is not only quantitative, but it is also qualitative, as today it is relatively easy to capture different aspects or features from a given entity or experiment. New pattern recognition methods should be designed, not only to deal with large amounts of information, in the sense of high number of data samples, but also with information of heterogeneous nature even in a single dataset.

In the context of this thesis, focused on multiview datasets and methods, a **view** of an entity is defined as a set of variables acquired by means of an instrument applied to the entity. In this sense, a view can be a picture, an audio or video recording, the results of a poll, survey or interview, physical variables of any kind, clinical variables, etc. Therefore, a **multiview dataset** is a dataset that comprises data matrices (data views) from different instruments or experiments on the same entities.

There are some concepts closely related that require a precise specification. While a **view** is the most general term in this area, it actually specifies the data directly acquired from an observation instrument (camera, microphone, x-ray machine, interviewer, electronic survey, etc.). This is also known as **sensory** information. On the other hand, when different characteristics are computed from a given sensory input (for example, different image descriptors from a picture), this data is usually defined as **feature sets**, and consequently the dataset is qualified as a **multifeature** dataset. It is possible to have hybrid datasets, with several sensory inputs (views) and several feature sets extracted from them.

Finally, in the context of multimedia information management and retrieval, it is often used the term **multimodal dataset** to refer to datasets that combine data from different media, like video, audio, still image, etc.

The methods presented in this thesis are generical and do not depend on a specific design or origin of the dataset, as they are conceived to process any kind of multiview, multifeature, or hybrid datasets. Throughout this thesis, the methods and the datasets used to test them will be qualified as multiview.

The relevance of multiview datasets is increasing due to the high availability of data acquisition instruments. However, most pattern recognition methods are designed to process a single data view. A naïve option is to discard all data views but one; a second option is to concatenate all the input views into a single data matrix; the third option is to use a proper multiview pattern recognition method. Many experiments have shown that the latter option renders better results [66, 62, 106, 11, 117, 106, 112, 70]. In other words, using multiview data is a challenge, as it requires new processing methods, but it is also an opportunity, as the potential results are better. As a consequence, multiview processing methods have become an important tool for data processing tasks. This is the goal of this work.

The multiview or multifeature quality is intrinsically heterogeneous, as the different aspects of the entities or subjects under study can also be heterogeneous. Next, examples of multiview datasets are given in order to highlight this heterogeneity and to illustrate the assortment of multiview dataset designs and the varied nature of the data views.

1.1.1.1 Image datasets

The University of California at Irvine (UCI) multiple features digits dataset [9], available at the UCI machine learning repository,[1] is a multifeature dataset created from a set of handwritten numerals (from '0' to '9'), scanned as 15×16 grayscale pixels images. There are 200 samples of each numeral, resulting in a total of 2,000 samples. The multifeature aspect of this dataset lies in the different image descriptors that have been applied to the images. More specifically, this dataset has six feature sets: (1) the pixel averages in 2×3 windows, (2) 76 Fourier coefficients of the character shapes, (3) 216 profile correlations, (4) 64 Karhunen-Love coefficients [99], (5) 47 Zernike moments [71], and (6) 6 morphological features (not specified). As each image descriptor captures a different aspect of the images, having multiple views or feature sets of the data is assumed to contain more information about the data samples.

Strictly speaking, this is a multifeature dataset, as several features have been extracted or computed from the same sensory input (the digit images). Figure 1.1 shows an example of the original digits of this dataset.

[1] `https://archive.ics.uci.edu/ml/datasets/Multiple+Features`

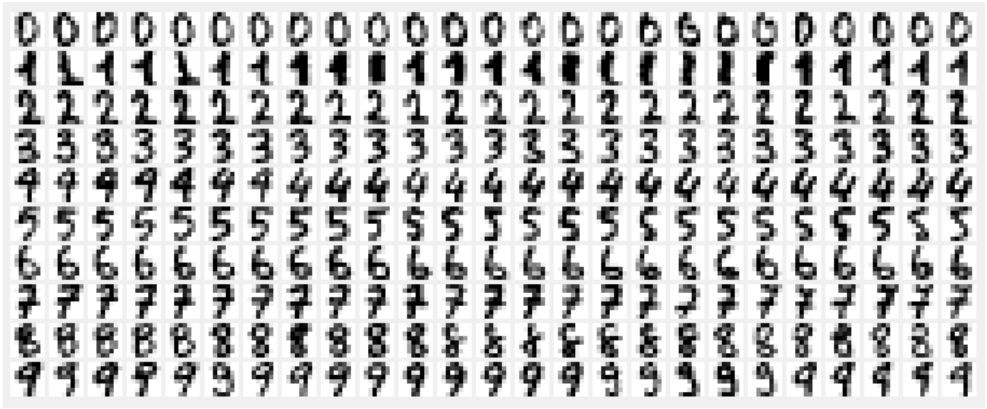


Figure 1.1: Example handwritten digits from the "Multiple features" dataset.

A similar multifeature image dataset is the Animal with attributes (AWA) dataset[65][2]. In this dataset the original data are $30,475$ photographies of animals, divided in 50 animal classes. As the input images have higher resolution than in the case of the digits dataset and they are in color, the image descriptors extracted from the pictures are different. However, the overall design of the dataset is the same: compute a series of image descriptors from the input images. These descriptors are: (1) $2,688$ color histogram features, (2) $2,000$ self-similarity features, (3) 252 pyramid histogram of oriented gradients features (PHOG) [25], (4) $2,000$ scale-invariant feature transform values (SIFT) [72], (5) $2,000$ colour SIFT values, and (6) $2,000$ speeded-up robust features (SURF)[6]. Figure 1.2 shows an example of the original images from which the dataset has been generated.

Another source of multiview datasets are hyperspectral images, where the same location is photographed using cameras that can capture light wavelengths other than those of visible light. A known example of this kind of datasets is the Hydice dataset [3], with a 191 band hyperspectral image of a mall in Washington DC. In this case, each band of the image can be considered a different view of the same entities, and therefore multiview methods can be useful to process it. Figure 1.3 shows some fragments of this image.

A qualitatively different multiview image dataset are the Columbia object image libraries (COIL-20 and COIL-100)[83][4], which respectively are collections of pictures of 20 or 100 objects. The multiview aspect in these datasets lies in the fact that there are 72 pictures of each object in the collection, taken

[2] `http://attributes.kyb.tuebingen.mpg.de/`

[3] `https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html`

[4] `http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php`

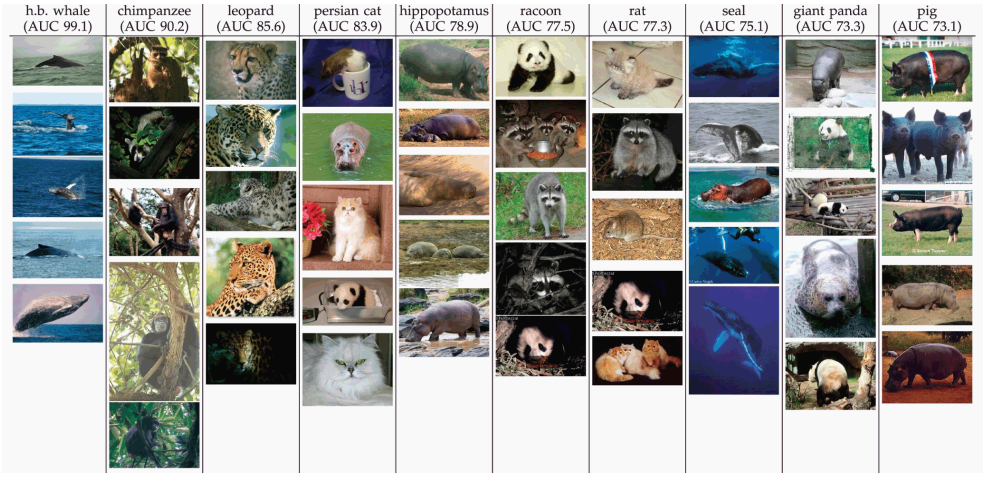


Figure 1.2: Example images from the "Animal with attributes" dataset. Taken from [65].

from different angles. Figure 1.4 shows an example of some of these pictures. No image descriptors are provided in this dataset, but simply the images of the objects with the backgrounds removed.

A similar dataset is the extended Yale face database B [42][5], that contains gray level images of 28 human subjects, each with 9 poses and 64 different lighting conditions. The goal of this dataset is to train face recognition systems robust to varying situations. Figure 1.5 shows some example images of one of the subjects in the dataset.

1.1.1.2 Text datasets

There are multiple ways in which a text can be analyzed to extract features from it. As a consequence, there are several approaches to text multiview dataset. Some of the most well known among them are described next.

The BBC News multiview text collection [44, 43] [6] comprises 2,225 news articles from the BBC news website (years 2004-2005) labelled with one of five possible topics (*business*, *entertainment*, *politics*, *sport* or *tech*). There are several subsets, but the overall dataset design is to split each news article in segments (ranging from 2 to 4) and use each of the segments as a different data view. The features of a segment are the non-stop words it contains. The

[5] `http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html`
[6] `http://mlg.ucd.ie/datasets/bbc.html`

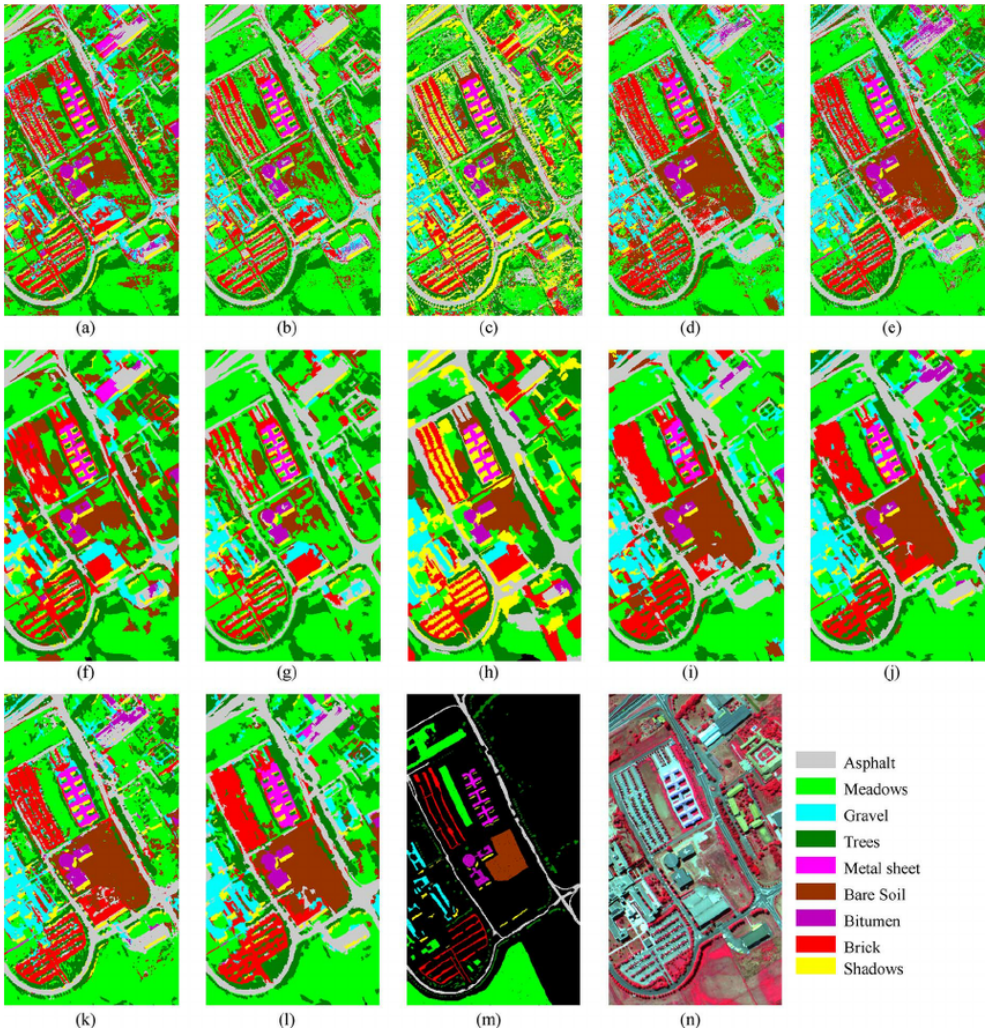


Figure 1.3: Example fragments from the Hydice dataset. Taken from [69].

complete feature matrix of each view is a matrix of documents × words.

A different approach to build multiview datasets is exemplified by the Reuters multilingual corpus [3][7], a set of 18,758 news articles labelled in six categories. In this case the multiview aspect lies in the fact that these articles are available in five different languages (English, French, German, Italian and Spanish).

[7] `https://archive.ics.uci.edu/ml/datasets/Reuters+RCV1+RCV2+Multilingual+Multiview+Text+Categorization+Test+collection`

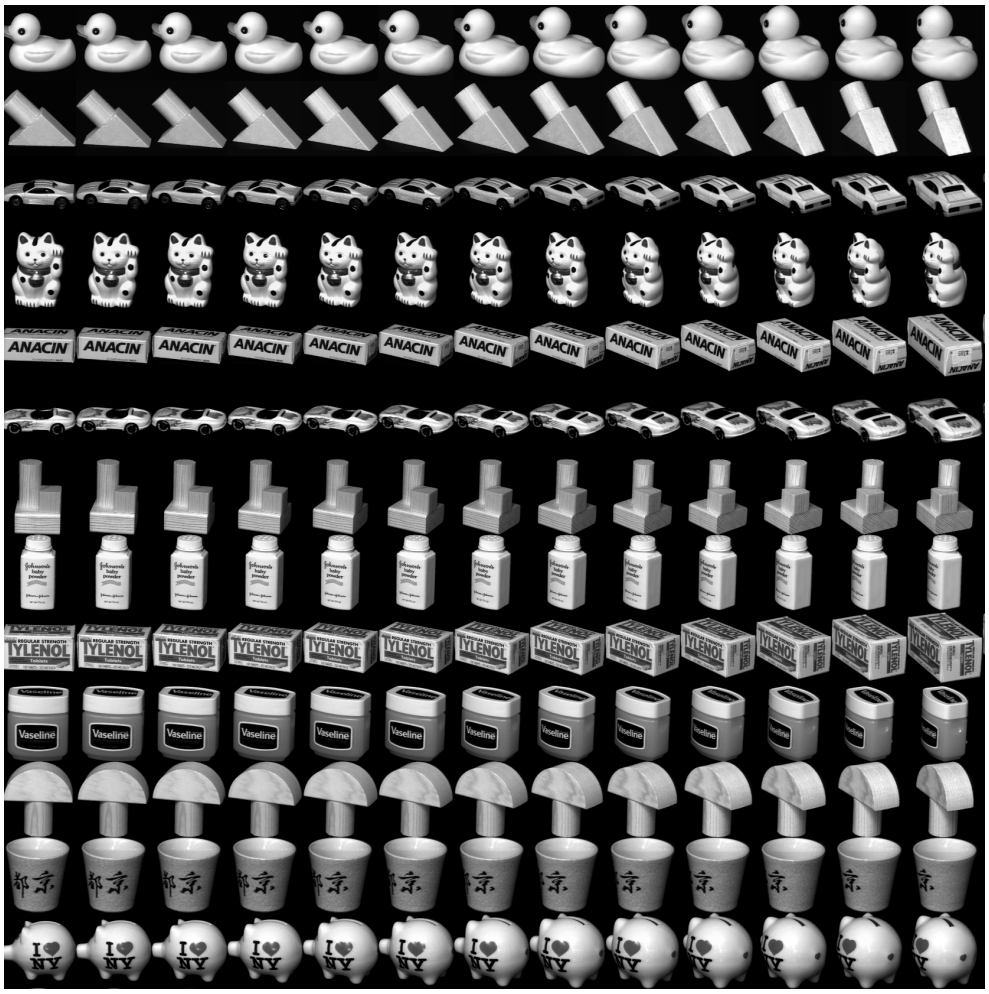


Figure 1.4: Example images from the COIL-20 dataset.

Figure 1.5: Example images from one subject in the extended Yale face database B.

	cod	diagnosis	sex	age	BT	remission	CR	date.cr	t(4;11)	t(9;22)	cyto.normal	citog
01005	1005	5/21/1997	M	53	B2	CR	CR	8/6/1997	FALSE	TRUE	FALSE	t(9;22)
01010	1010	3/29/2000	M	19	B2	CR	CR	6/27/2000	FALSE	FALSE	FALSE	simple alt.
03002	3002	6/24/1998	F	52	B4	CR	CR	8/17/1998	NA	NA	NA	<NA>
04006	4006	7/17/1997	M	38	B1	CR	CR	9/8/1997	TRUE	FALSE	FALSE	t(4;11)
04007	4007	7/22/1997	M	57	B2	CR	CR	9/17/1997	FALSE	FALSE	FALSE	del(6q)
04008	4008	7/30/1997	M	17	B1	CR	CR	9/27/1997	FALSE	FALSE	FALSE	complex alt.
04010	4010	10/30/1997	F	18	B1	CR	CR	1/7/1998	FALSE	FALSE	FALSE	complex alt.
04016	4016	2/10/2000	M	16	B1	CR	CR	4/17/2000	FALSE	FALSE	FALSE	simple alt.
06002	6002	3/19/1997	M	15	B2	CR	CR	6/9/1997	FALSE	FALSE	TRUE	normal
08001	8001	1/15/1997	M	40	B2	CR	CR	3/26/1997	FALSE	FALSE	FALSE	del(p15)
08011	8011	8/21/1998	M	33	B3	CR	CR	10/8/1998	FALSE	FALSE	FALSE	del(p15/p16)

Figure 1.6: Excerpt from the phenotypic data in the ALL dataset.

The Citeseer dataset [73][8] consists of 3,312 scientific publications from the Citeseer database [10] labelled with one of six classes according to their subject area. It features two data views: (1) a binary dictionary of 3,703 words, where each publication has a 1 if it contains the word or a 0 otherwise, and (2) a symmetrical citation network, as a matrix of $3,312 \times 3,312$ elements whose value $c_{ij} = 1$ if document i cites document j or vice versa, or $c_{ij} = 0$ if there are no references between these documents. In this case, the dictionary view is defined in feature space while the citation view is defined in graph space.

There exist several datasets with the same structure, as for example the Cora dataset [79], which contains $2,708$ scientific publications classified into one of seven classes. This dataset also has two views, one bag of words with $1,433$ words and a reference graph that represents $5,429$ links between documents.

WebKB [24] has a similar structure, but the source data are the words in web pages from four universities and their hyperlinks.

1.1.1.3 Biological datasets

Biology, medicine and related areas are also a natural source of multiview data, as the subjects themselves are of complex nature and there are numerous kinds of tests and instruments that produce different data. Some examples of these datasets are presented next.

The acute lymphocytic leukemia dataset (ALL) [18] [9] compiles information from 128 subjects. One of the views includes 21 phenotypic features (age, gender, including biological markers and clinical diagnosis). The other view has the expression level of $12,625$ genes. Figures 1.6 and 1.7 illustrate the two views of this dataset.

[8] http://www.cs.umd.edu/~sen/lbc-proj/LBC.html

[9] http://bioconductor.org/packages/release/data/experiment/html/ALL.html

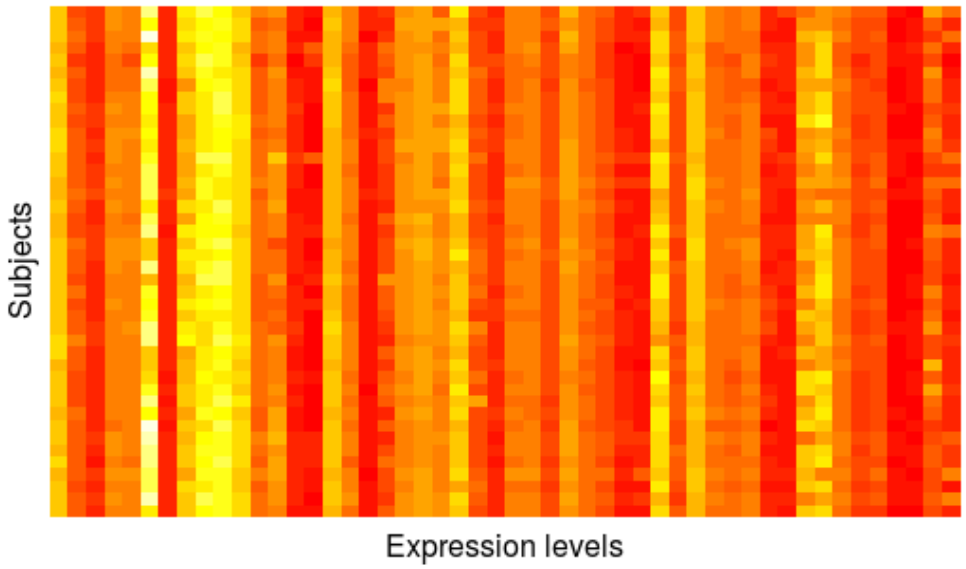


Figure 1.7: Partial heatmap of the gene expression data in the ALL dataset.

The Berkeley protein dataset for genomic data fusion [66][10] is a multiview dataset whose samples are $1,040$ yeast proteins. The proteins are labeled according to their location, as either membrane proteins, ribosomal proteins, or other. This dataset comprises 8 data views or feature sets, intended to grant knowledge on different aspects of the proteins so as to improve the predictive power of machine learning methods. The 8 views of this dataset are:

- K_{SW}: Smith-Waterman[93] distance kernel on protein sequences.
- K_B: BLAST[2] distance kernel on protein sequences.
- K_{Pfam}: Pfam database [95] hidden markov model kernel on protein sequences.
- K_{FFT}: fast fourier transform of the hydrophobicity profile [64] extracted from the protein sequences. Useful to recognize membrane proteins.
- K_{LI}: linear kernel on protein interactions.
- K_D: diffusion kernel on protein interactions.
- K_E: radial basis kernel on gene expression (microarray gene expression on 441 genes).
- K_{RND}: linear kernel on a matrix of random numbers, used as baseline to compare the other kernels.

[10] `http://noble.gs.washington.edu/proj/sdp-svm/`

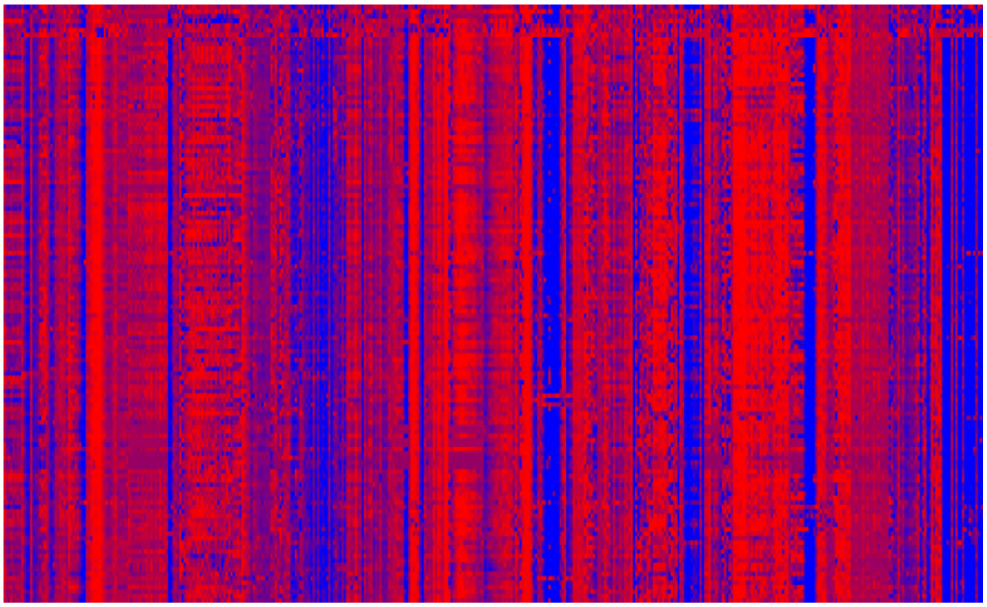


Figure 1.8: Expression profiles of the ribosomal genes. Rows correspond to the ribosomal genes, and columns to the microarray experiments. Taken from [66]

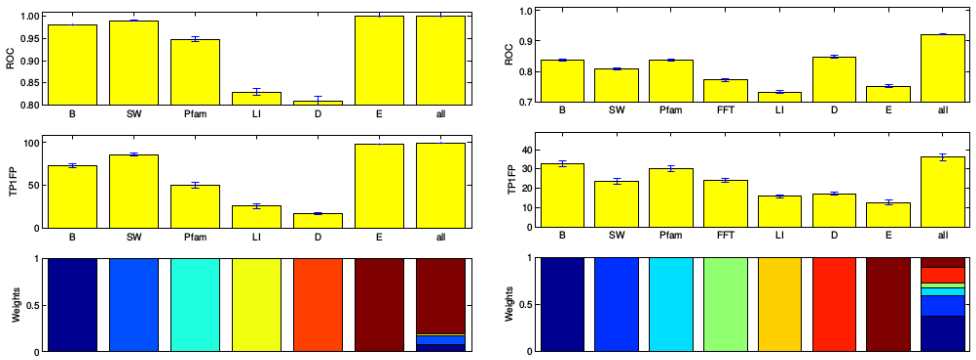
.

An important aspect of these data views is that some are defined in feature space (for example the expression levels) while others are relationships between proteins and therefore are defined in graph space. Also, as the focus of [66] is to present different kernel methods, some of the feature sets are different kernel matrices of the same data (protein sequences or interactions). Strictly speaking, this is a multiview dataset, as three different views of the proteins are used (sequences, interactions, expression), and also a multifeature dataset, as several features are extracted from the original views (the different kernels applied).

Figure 1.8 shows the expression profiles of the ribosomal genes. Figure 1.9 shows the improvement on classification accuracy of using all the data views relative to using only one view.

1.1.1.4 Multimodal datasets

Multimodal datasets combine information from different media sources, such as image and audio, in order to increase the amount of information available to the learning methods.



(A) Ribosomal proteins (B) Membrane proteins

Figure 1.9: Comparison of single-view versus multiview classification. The first row shows the ROC classification score. The second row shows the percentage of true positives at one percent false positives. The third row shows the relative weights of the kernel matrices for the linear combination used in the experiments. Taken from [66].

The Wikipedia articles cross-modal dataset [22][11] contains both images and texts from 2,866 Wikipedia articles, classified into one of ten categories. This dataset is designed to test cross-modal multimedia retrieval methods, where information from a single view is deemed insufficient for the task and the combination of the different views is expected to increase the quality of the results.

There exist several datasets composed of images and their associated annotation tags. Among the most popular in the literature are: Corel5k [34], that contains 5,000 images from a Corel image CD that are manually annotated with different tags; ESP Game [105], that includes 20,000 images extracted from the profiles of the users of an online game, along with user-defined textual information; and NUS-WIDE [19][12], that contains 269,648 extracted from the social network Flickr, along with six image feature sets, and the ground truth for 81 concepts or tags, where a given image can have more than one concept. Figure 1.10 shows some example images and their tags from the NUS-WIDE dataset.

[11] `http://www.cs.umd.edu/~sen/lbc-proj/LBC.html`
[12] `http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm`

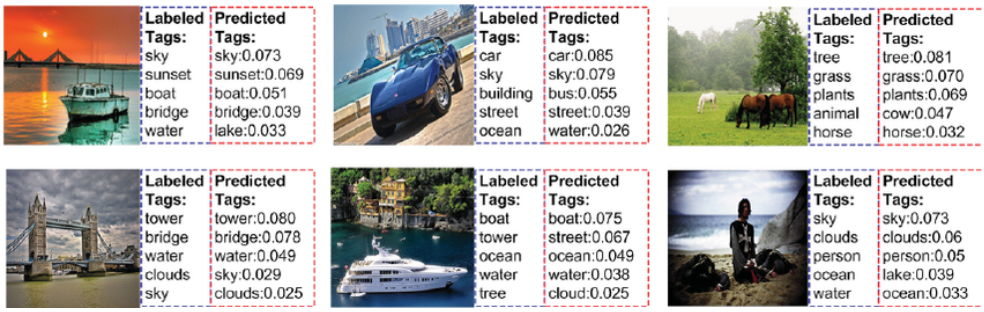


Figure 1.10: Example figures and tags from NUS-WIDE dataset. Taken from [17].

1.1.2 Unsupervised pattern recognition methods

> "Development of unsupervised pattern recognition methods is the most challenging and promising area of research today." Andrew Ng, NIPS Conference, Dec. 2016

Pattern recognition methods can be classified in two categories: supervised and unsupervised. Supervised methods require the existence of some kind of data sample labeling that indicates the class or other relevant property of each sample. Building labeled datasets is usually expensive, as in most cases a human expert (or a team of them) is required to generate the labeling. This limits the number of samples that can be labeled by the availability of human experts, usually in the thousands of samples. Moreover, labeling a dataset can be intrinsically difficult or laborious, as in image partition tasks where each object in the image has to be isolated, for example by drawing its contour, or in medical diagnosis.

It is important to note that the nature of the labels depends on the task that has to be solved. For example, on image processing there are numerous tasks to be performed: image segmentation, identification of a specific object type, identification of letters or digits, of faces, car models, etc. In each case, the labeling will be different, even if the image set is the same.

Nowadays, in the era of the big data, it is possible to acquire millions or even billions of data samples more easily than it has ever been. But in most cases, this data is not labeled. In general, manually labeling millions of data samples is unaffordable. Although hybrid methods exist that can use partially labeled data, known as semi-supervised methods, they are not always suitable.

Unsupervised methods do not require any kind of human annotation on the data in order to perform their pattern recognition tasks. Two of the most relevant unsupervised pattern recognition tasks are dimensionality reduction and data clustering.

1.1.2.1 Dimensionality reduction

Dimensionality reduction methods, also known as data embedding methods or data projection methods, have as their goal to transform an input dataset with high dimensionality (i.e. high number of variables), into a lower dimensionality space, i.e. with fewer variables. This process must keep as much information as possible from the original dataset. Application of dimensionality reduction methods often is one of the first steps in the data analysis workflow. The dimensionality reduction offers several advantages: it gives researchers a better understanding of the data, it sometimes improves the usefulness of the data (removing noise or irrelevant information), and it also reduces the computational complexity of the next data processing steps.

Some dimensionality reduction methods, also called data projection methods, are specifically designed to generate a graphical representation of high dimensional data. Reducing the data to a 2 or 3-dimensional space allows to graphically display the data, giving researchers an insight into the structure of the data that may be difficult to attain otherwise.

Among the most relevant dimensionality reduction methods are: principal components analysis (PCA)[53], multidimensional scaling (MDS) [60, 61, 23], t-distributed stochastic neighbour embedding (t-SNE) [102, 74, 75, 76], canonical correlation analysis (CCA)[46, 47].

1.1.2.2 Clustering algorithms

The goal of clustering algorithms is to find groups or clusters of samples given a data set of which no previous information about groups or classes is assumed. This is one of the first steps when analyzing new data, as it suggests a structure for initially unstructured data.

Most clustering algorithm have a parameter that controls the granularity of the clustering, be it the number of clusters to be obtained, the minimum distance to group two points together, or any other equivalent granularity controlling parameter. There is no single solution of the clustering of a dataset, but rather it can be done at different levels of detail or granularity.

There is a wide range of clustering algorithms of which some representative examples are: K-means[48, 81], hierarchical clustering[52, 82], partition around medoids (PAM)[59, 31], DBScan [35], spectral clustering[87, 84].

1.2 State of the art: multiview dimensionality reduction

The goal of multiview dimensionality reduction methods is to reduce a dataset with multiple, high-dimensional data views $\{X_1, X_2, \dots X_v\}$ into a single, lower-dimensional space or data projection, while keeping the most relevant properties of the original data.

In general, dimensionality reduction methods are expected to preserve the relative distances between the data samples. This is a relatively difficult problem to solve with a single data view, and it obviously becomes harder to solve when there are several data views, as the distances between data samples may vary from view to view and a consensus solution must be found.

When analyzing single-view dimensionality reduction methods, the quality of the resulting data space with respect to the original data is analyzed from two points of view. First, the **local data structure** is evaluated: do the points in the low-dimensional space have the same neighbours as in the original, high-dimensional space? A method whose projections satisfy this condition are said to keep the local data structure. However, this is not the only desirable property of a dimensionality reduction method, as there is a second question to be assessed: if points a and b are far from each other in the input space, are they also far from each other in the output space? This is also an important feature, and the methods that satisfy it are said to preserve the **global data structure**.

These two properties (preservation of local and global data structure) are also central to multiview dimensionality reduction methods, with the added difficulty of greater computational complexity and potential conflicts between views. A nave solution to this problem is to concatenate the input views into a single feature matrix, but this method does not account for the intrinsic structure of each data view and therefore does not fully exploit the richness of the multiview data.

1.2.1 Multiview dimensionality reduction methods

1.2.1.1 Low-rank approximation methods

Laplacian eigenmaps[7], also known as spectral embedding, is a well known single-view dimensionality reduction method. Multiview spectral embedding (MSE)[108] is an extension of spectral embedding to multiview datasets. Its goal is to find a low-dimensional and smooth embedding of the high-dimensional, multiple input views. More specifically, if the multiview dataset X is composed of V data matrices such that $X = \{X^{(v)} \in \mathbb{R}^{n \times m_v}\},\ \forall\ 1 \leq v \leq V$, then the resulting low-dimensional representation of X is $Y \in \mathbb{R}^{n \times d}$, where d is a user defined parameter such that $d < m_v\ \forall\ 1 \leq v \leq V$. In other words, the dimensionality of the low-dimensional representation has to be lower than the dimensionality of any of the input views. MSE comprises three steps: part optimization, global coordinate alignment, and alternating optimization. The first stage, part optimization, is based on the patch alignment framework described in [116]. The goal of this stage is to independently obtain a low-dimensional representation of each view that preserves the locality of each sample, i.e. the relationship with its closest neighbours. In order to find the patch of the i-th sample in the v-th view, denoted as $X_i^{(v)}$, the following

objective function has to be minimized

$$\min_{Y=\{Y_i^{(v)}\},\alpha} \sum_{v=1}^{V} \alpha_v \mathrm{tr}\Big(Y_i^{(v)} L_i^{(v)} Y_i^{(v)^T}\Big) \tag{1.1}$$

where $Y_i^{(v)}$ is the low-dimensional representation of $X_i^{(v)}$, $L_i^{(v)}$ is the Laplacian of $X_i^{(v)}$, and α_v is a weight value associated to each view that increases with the relative relevance of view v.

On the second stage, the different low-dimensional representations $Y^{(v)}$ are aligned by assuming that they can be unified from the global coordinate matrix Y by means of a selection matrix that encodes the spatial relationship with the samples in the high dimensional space, such that $Y_j^{(v)} = Y S_j^{(v)}$. In other words, the low-dimensional representations of each view are consistent with each other globally. The solution to these transformations is given by the following objective function

$$\begin{aligned} &\min_{Y,\alpha} \sum_{v=1}^{V} \alpha_v^r \mathrm{tr}\Big(Y L_n^{(v)} Y^T\Big) \\ &\text{s.t. } YY^T = I;\ \sum_{v=1}^{V} \alpha_v = 1;\ \alpha_v \geq 0 \end{aligned} \tag{1.2}$$

The third and last stage of MSE is an alternating optimization algorithm that minimizes the previous objective function in a computationally efficient way. Matrix Y is the resulting global low-dimensional projection.

The method described in [88] aims at generating a common, low-dimensional projection of several data views. The goal is to apply this projection with cross-media document retrieval, looking for the nearest neighbours in the projection space, saving time and memory. Their premise is to make two samples appear close to each other in the projection if and only if they are close in all the input views. In order to obtain the common projection, a projection function is defined for each data view such that (two view case, such as described in the paper; it can be extended to more views):

$$g_1 : \mathbb{R}^{d_1} \to \mathbb{R}^D \text{ and } g_2 : \mathbb{R}^{d_2} \to \mathbb{R}^D \tag{1.3}$$

where d_i is the dimensionality of view i and D is the dimensionality of the desired common projection; in general it is assumed $D \ll \max(d_1, d_2)$. A linear parametrization of the above functions is assumed, such that $g_1^w := \langle w_1, \phi(x_i)\rangle$ and $g_2^w := \langle w_2, \psi(y_i)\rangle$. Functions g_1 and g_2 are determined by optimizing the following objective function

$$L(w_1, w_2, \mathcal{X}, \mathcal{Y}, \mathcal{S}) := \sum_{i,j=1}^{m} L^{i,j}(w_1, w_2, x_i, y_j, \mathcal{S}_{x_i}) + \eta\Omega(w_1) + \gamma\Omega(w_2) \tag{1.4}$$

where the loss function L is defined as follows

$$L^{i,j}(w_1, w_2, x_i, y_j, \mathcal{S}_{x_i}) = \frac{I_{y_j \in \mathcal{S}_{x_i}}}{2} \times L_1^{i,j} + \frac{1 - I_{y_j \in \mathcal{S}_{x_i}}}{2} \times L_2^{i,j} \tag{1.5}$$

with

$$L_1^{i,j} = \|g_1^{w_1}(x_i) - g_2^{w_2}(y_j)\|_F^2 \tag{1.6}$$

$$L_2^{i,j}(\beta_d) = \begin{cases} -\frac{1}{2}\beta_d^2 + \frac{a\lambda^2}{2}, & \text{if } 0 \leq |\beta_d| < \lambda \\ \frac{|\beta_d|^2 - 2a\lambda|\beta_d| + a^2\lambda^2}{2(a-1)}, & \text{if } \lambda \leq |\beta_d| \leq a\lambda \\ 0, & \text{if } |\beta_d| \geq a\lambda \end{cases} \tag{1.7}$$

where a and λ are heuristically chosen constants. This optimization problem is further decomposed in two lesser problems and resolver iteratively. The resulting projection functions g_1 and g_2 can then be used to generated the desired common projection.

The convex multi-view subspace learning method (MSL) [107] aims at learning a subspace representation of a multiview dataset while assuming and enforcing conditional independence between the different views. A convex regularizer that finds the subspace is proposed. This method is also designed to achieve the best possible reconstruction of the original data from the low-dimensional representation. Summarizing the method, it applies an optimization algorithm to the objective function

$$\min_{K \in \text{conv}\mathcal{G}} f(K), \; f(K) = \|\hat{Z} - K\|_F^2 \tag{1.8}$$

where K is the desired low-dimensional representation, conv$\mathcal{G}$ is the convex hull of the set of possible subspaces for the input views, $\hat{Z} = CH$ is the product of the concatenated feature matrix C and its common $2,1$-norm H. Figure 1.11 compares the quality of an image reconstructed using MSL with respect to the same image reconstructed using two alternative methods: local multiview subspace learning (LSL) [88] and single-view subspace learning (SSL) [13].

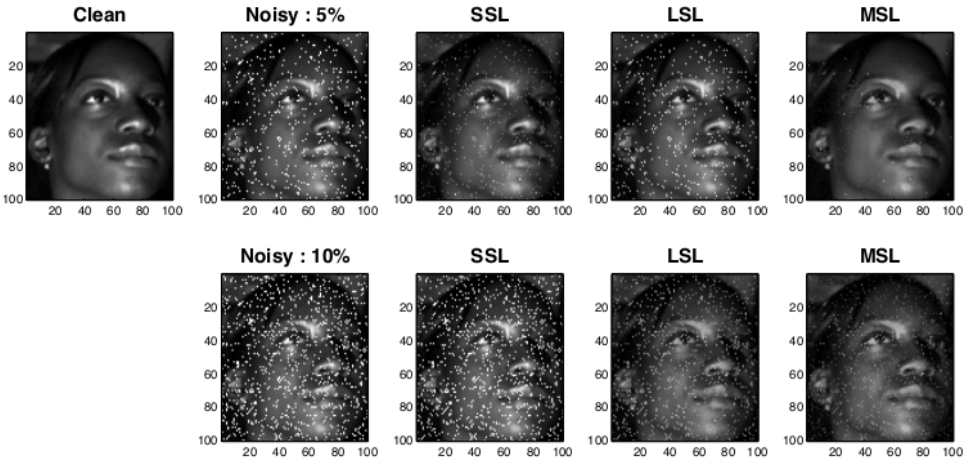


Figure 1.11: Reconstructed image comparison between different dimensionality reduction methods. Taken from [107].

The ensemble manifold regularized sparse low-rank approximation (EMR-SLRA) algorithm [114] uses the framework of least-squares component analysis [27] to obtain a low-rank approximation of the concatenated multiview feature matrix. This algorithm comprises three steps: first it computes the low-rank approximation of the multiview matrix, second it regularizes the ensemble manifold, and finally it determines and applies a group sparsity constraint. Let $X = \{X_1, X_2, \dots X_V\}$ be the concatenated feature matrix. The low-rank approximation Y of X is given by the following expression

$$\begin{aligned} &\min_{U,Y} \|X - UY\|^2 \\ &\text{s.t. } U^T U = I \end{aligned} \tag{1.9}$$

On the second step, the low-rank approximation Y obtained has to be regularized, as from its definition all the features in X have been processed uniformly, and the intrinsic structure of each of the input views has not been considered. EMR-SLRA uses ensemble manifold regularization [41] and heat kernels [7] to find a vector of view coefficients $\beta = \{\beta_1, \beta_2, \dots, \beta_V\}$, so that input view X_i is multiplied by that factor in order to account for their different relevance. The expression to find β is

$$\begin{aligned} \min_{Y,\beta} & \sum_{v=1}^{V} (\beta_v)^r \mathrm{tr}(Y L^{(v)} Y^T) \\ & \text{s.t. } \sum_{v=1}^{V} \beta_i = 1, \beta_v > 0 \end{aligned} \tag{1.10}$$

where $L^{(i)}$ is the Laplacian matrix of input view X_i and r is a user defined parameter. Finally, a group sparsity constraint is introduced in order to minimize the potential noise introduced in Y from X. To minimize the effects of that noise, an ideal multiview feature matrix $\hat{X}$ is obtained by applying the $\ell_{2,1}$-norm fitting constraint [85], as it is robust against noise in the data [109]. The expression to compute $\hat{X}$ is

$$\min_{\hat{X}} \|\hat{X} - X\|_{2,1} \tag{1.11}$$

The $\ell_{2,1}$-norm regularized is defined as:

$$\|X\|_{2,1} = \sum_i \sqrt{\sum_j X_{ij}^2} = \sum_i \|x_{i,:}\|_2 \tag{1.12}$$

Replacing X by $\hat{X}$ in Equation 1.9 along a number of iterations produces a low-rank projection matrix Y that is more robust to the noise in the input views.

Multitask multiview feature embedding (MMFE) [115] generates a low-rank approximation matrix to the multiview data. The overall structure of the method is similar to that of EMR-SLRA [114], although the integration of the multiview data is now based on multitask learning methods [14, 36, 86, 33]. More specifically, each input data view is considered a different learning task. The objective function used to find the integrated low-rank representation Y is

$$\begin{aligned} \min_{U,U^{(v)},Y,Y^{(v)}} & \|X - UY\|^2 + \sum_{v=1}^{V} \|X^{(v)} - U^{(v)} Y^{(v)}\|^2 \\ & + \alpha \sum_{v=1}^{V} \|Y - Y^{(v)}\|^2 \\ & \text{s.t. } U^T U = I,\ {U^{(v)}}^T U^{(v)} = I \end{aligned} \tag{1.13}$$

where $\{X_1, X_2, \ldots X_V\}$ are the input views, X is the concatenated feature matrix, $Y^{(v)}$ is the low-rank representation of $X^{(v)}$, Y is the low-rank representation of X, $U^{(v)}$ and U are the projection of the data points into the corresponding low-rank space, and α is a regularization user defined parameter.

The resulting low-rank matrix Y is the desired low-dimensional representation of the original data.

1.2.1.2 Probabilistic methods

The multiview stochastic neighbor embedding method (m-SNE) [110] is a multiview method based on stochastic neighbor embedding (SNE) [49] and t-distributed SNE (t-SNE) [102, 74]. t-SNE computes a symmetric joint probability distribution from the pairwise distances of the data samples. The conditional probability between points x_i and x_j in the input high-dimensional space is

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma^2)}{\sum_{k \neq l} \exp(-\|x_k - x_l\|^2/2\sigma^2)} \tag{1.14}$$

where σ is an hyperparameter of the method. The symmetric joint probability distribution, designed to avoid the effect of outlier points, is $p_{ij} = \frac{p_{j|i}+p_{i|j}}{2n}$, where n is the number of data samples. m-SNE generates one symmetric joint probability matrix for each input view and combines them using the next formula

$$p_{ij} = \sum_{v=1}^{V} \alpha^{(v)} p_{ij}^{(v)} \tag{1.15}$$

where $\alpha^{(v)}$ is a coefficient associated with each view and $p_{ij}^{(v)}$ is the joint probability matrix of view v. This produces a single joint probability matrix that is used as input for the standard t-SNE method.

1.2.1.3 Kernel-based methods

The kernelized multiview projection (KMP) [91] is a kernel method to reduce a multiview dataset of human actions, where each frame of a video capture of a subject performing an action is a view of the dataset. The approach of this method is to use a kernel that reduces the multiple, high-dimensional input views (video frames) to a single, low-dimensional space. Another requisite of this method is that it should be computationally efficient. The method comprises three steps. First, it applies an image filter called incremental nave Bayes filter (INBF) to remove the noise in the frames and try to keep only the information relevant to the task. This step is specific to the human action recognition problem. Second, it computes a kernel matrix for each input view using dynamic time warping (DTW)[8] using the following kernel function between videos v_p and v_q in view i:

$$k_i(v_p, v_q) = \exp\Big(- \frac{\mathrm{DTW}(X_p^i, X_q^i)^2}{2\sigma^2} \Big) \tag{1.16}$$

this produces a set of kernel matrices $K_1, K_2, \ldots, K_M$, where M is the number of views. After computing the Laplacian of these kernel matrices, $L_1, L_2, \ldots, L_M$, an iterative procedure is applied to compute the fused kernel matrix $K = \sum_{i=1}^{M} \alpha_I K_i$ and the fused Laplacian matrix $L = \sum_{i=1}^{M} \alpha_I L_i$. The coefficients initially are $\alpha_i = \frac{1}{M}$ $1 \leq i \leq M$, but are iteratively adjusted by solving the generalized eigenvalue problem

$$KLK_p = \lambda KDK_p \tag{1.17}$$

where D is the diagonal matrix of K_p. This process is repeated until K_p and α_i become stable; the final common projection P is derived from the stable value of K_p.

1.2.1.4 Co-training methods

The method presented in [111] is specifically designed to improve image tagging and classification. It assumes a two-view dataset, with images as one view and annotation tags as the other. From the image view, several feature sets can be extracted using different image descriptors. The main strategy of this method is to have the information on each view guide the information extraction from the other view, in order to obtain an improved set of image tags. In the end, an improved common subspace Z is obtained, on which the image tags are predicted. First, the geometric structure of the different views is modeled by a corresponding k-nearest neighbour graph $\{W^v\}_{v=1}^{V}$, where V is the number of image feature views. $W^S = T^T T$ is used to model the semantic structure of the images, where T is the matrix of image tags. The common subspace Z is obtained by optimization of the following expression

$$\min_{Z} f(Z), \text{ s.t. } Z^T Z = I \tag{1.18}$$

where $f(Z)$ is defined as

$$f(Z) = \frac{1}{\gamma} log\Big\{ \sum_{v=1}^{V} \exp[\gamma \|Z - (W^h \odot W^S)\|_F^2] \Big\} + \eta \|T - ZZ^T T\|_F^2 \tag{1.19}$$

where $\odot$ is the Hadamard product [39] and $\|A\|_F$ is the Frobenius norm of matrix A. The common subspace Z is then used to train a classifier to decide on the presence or absence of each possible image tag in the dataset.

1.3 State of the art: multiview clustering

In general terms, clustering a dataset with multiple data views $\{V_1, V_2, \dots V_c\}$ involves the following steps:

1. Obtain a similarity matrix S_i for each view V_i.
2. Compute a projection P_i of each S_i into a space suitable for clustering.
3. Produce a clustering assignment.

The main structural difference between the multiview clustering methods proposed in the literature lies in the step where the information from the multiple views is collapsed into a single view in order to produce the final clustering assignment.

The first category of multiview clustering methods merges the similarity matrices to obtain a combined similarity matrix S' that minimizes the differences between the input similarity matrices S_i, i.e. views are merged in Step 1. Afterwards, a standard clustering algorithm is applied to S' in order to obtain the final clustering.

The second category of multiview clustering methods merge the input views during Step 2 to generate a compatible projection for all views (P'). Afterwards, a standard clustering method is applied to the merged projection P'.

Ensemble clustering methods are designed to overcome the randomness of clustering methods such as K-means by combining clustering assignments from several runs in order to find a stable assignment. Although not strictly considered as multiview clustering methods, they can be used for multiview clustering if they are applied to the clustering assignments of different views. Thus, they would produce a clustering assignment compatible with all views. These methods merge the information from the different views after Step 3.

There exist several multiview clustering methods in the literature that are described next, classified according to the step where they merge the information in order to produce the final clustering assignment.

1.3.1 View merging methods

The first category of multiview clustering methods merge all the input views into a single similarity matrix S', then apply a standard clustering algorithm to it.

The method described in [29] is designed to merge exactly two input views by minimizing the disagreement between them, applying the Minimizing-disagreement algorithm [28]. This method generates a weighted graph where each data sample is a node and the edges between nodes i and j are weighted

using a Gaussian function. The input value to the Gaussian function, i.e. the similarity value, is high if nodes i and j are relatively close at least in one of the input views. More specifically, the similarity between data points i and j in the weighted graph is given by the expression

$$w_{ij} = \sum_k \exp(-\frac{\|x_i^{(1)} - x_k^{(1)}\|^2}{2\sigma_1^2}) \exp(-\frac{\|x_i^{(2)} - x_k^{(2)}\|^2}{2\sigma_2^2}) \tag{1.20}$$

where $x_k^{(v)}$ is the vector of coordinates of data sample k in view v, and σ_1 and σ_2 are two user defined parameters that control the radius of the Gaussian on each of the input views. This gives as a result a combined adjacency matrix A_{12}. Then, a variant of spectral clustering [84] is applied to A_{12}.

A co-training approach is proposed in [62] for multiview spectral clustering. Co-training is a technique developed for semi-supervised methods, where the method is first trained on labeled data, then run on unlabeled data in order to suggest a possible label. In the co-training for multiview spectral clustering, the co-training is trained first with the clustering assignment of one of the input data views. The assumption of the method is that if points a and b belong to the same cluster in a view V_i, then they should belong to the same cluster in the remaining views; complementarily, if a and b do not belong to the same cluster in a view V_j then they should not belong to the same cluster in the other views. This method is designed to work on two data views V_1 and V_2. First, standard spectral clustering is run on V_1. The clustering assignment obtained is used to adjust the geometry of the adjacency graph of V_2. In parallel, the same process is performed on the opposite views. This process is repeated for a user-defined number of iterations, with the goal of obtaining two adjacency matrices as similar as possible. Let S_1 be the adjacency matrix of input view V_1; the update of S_1 on iteration i, denoted by S_1^i, is given by

$$S_1^i = \text{sym}\Big(U_2^{i-1} U_2^{i-1^T} S_1^{i-1}\Big) \tag{1.21}$$

where the symmetrization operator $\text{sym}(S) = (S + S^T)/2$, and U_2^{i-1} are the eigenvectors of S_2 computed on iteration $i-1$. The expression to compute S_2^i is equivalent. This way, both adjacency matrices tend to converge on a common adjacency matrix S' on which standard spectral clustering is applied in order to obtain the final clustering assignment.

The method proposed in [106] is both a multiview clustering method and a multiview feature learning method. Its goal is to assign a weight to each feature of each input view in order to induce a structured sparsity on the data samples that reflects the underlying clustering structure. This method uses the group ℓ_1-norm regularization [113] to learn the effect of each data view

on each cluster, and the $\ell_{2,1}$-norm regularization to learn the effect of each data view on multiple clusters. This way, the task becomes an optimization problem whose objective function is

$$\min_{W,F^TF=I} \|X^TW + 1_n b^T - F\|_F^2 + \gamma_1\|W\|_{G_1} + \gamma_2\|W\|_{2,1} \tag{1.22}$$

where W is a vector of feature weights, X is a data matrix with all the features of all the input data matrices, 1_n is a vector of n ones, b is an intercept vector and γ_1 and γ_2 are user-defined parameters. The result of this optimization procedure is a vector of weights to be applied to the input features. Afterwards, a standard clustering method can be applied on this modified matrix of features.

The method proposed in [112] comprises two steps. First, it uses the prior information in all data views to generate a sparse representation for each view of the data samples. More specifically, two constraints labelled "must-link" and "cannot-link" are defined to control the way each sample is related to each other. The objective function designed to compute these sparse representations is

$$\begin{gathered}\min_{Z_i^v} \|x_i^v - X_{-i}^v Z_i^v\|^2 + \alpha\|Z_i^v\|_1 + \beta\sum_{w\neq v}\|Z_i^w Z_i^v\|_1 \\ s.t.\ \mathrm{diag}(Z^v) = 0\end{gathered} \tag{1.23}$$

where X^v is the v-th data view, a negative subindex means all rows in the matrix except the one in the subindex, and α and β are parameters of the method. The result of this optimization is the sparse representation Z^v of each input view X^v, with $Z \in \mathbb{R}^{n\times n}$, where n is the number of samples in the dataset. Then, a unique affinity matrix is constructed using any of the Z^v matrices or their average *(sic)* according to the expression $A = \frac{1}{2}\big(|Z^v|^T + |Z^v|\big)$. Standard spectral clustering is applied to A in order to find the clustering of the input data.

1.3.2 Intermediate merging methods

An implementation of this approach using Canonical Correlation Analysis in order to maximize the correlation of samples across the projected views can be found in [16].

Co-regularized multiview spectral clustering [63] is a method that also combines the adjacency matrices of the input views in order to produce a single adjacency matrix on which to apply standard spectral clustering. This

proposal is based on the semi-supervised technique of co-regularization [80], where the clustering assignments from the different views are expected to be equal. This produces an optimization problem whose objective function is

$$\max_{U^{(1)},U^{(2)},\ldots,U^{(M)}} \sum_{v=1}^{M} \mathrm{tr}\Big(U^{(v)^T}\mathcal{L}^{(v)}U^{(v)}\Big) + \lambda \sum_{1\leq v,w\leq M, v\neq w} \mathrm{tr}\Big(U^{(v)}U^{(v)^T}U^{(w)}U^{(w)^T}\Big) \tag{1.24}$$

where M is the number of input views, $U^{(i)}$ is the matrix of eigenvectors of the input view V_i, and $\mathcal{L}^{(i)}$ is the Laplacian matrix of the input view V_i. Standard spectral clustering is applied to the co-regularized matrix of eigenvectors. This method can be applied to any number of input views.

A partial multiview clustering method is described in [70], by partial meaning that it is able to process datasets with incomplete views. This method is analyzed for two view datasets but it can also be applied to a larger number of input views. For this case, the partial multiview dataset is split in three subsets, $X = \{\hat{X}^{(1,2)}, \hat{X}^{(1)}, \hat{X}^{(2)}\}$, where $\hat{X}^{(1,2)}$ contains the samples that appear in both input views, $\hat{X}^{(1)}$ contains the samples that only appear on the first input view, and $\hat{X}^{(2)}$ contains the samples that appear only in the second input view. The main idea behind this method is to try to learn a common latent subspace for the two views. Applying non-negative matrix factorization (NMF) [67], a basis matrix for each view's latent space can be learnt, denoted as $U^{(1)}$ and $U^{(2)}$. The representation of the data samples in these latent subspaces are two matrices, respectively $P^{(1)}$ and $P^{(1)}$. In order to determine $U^{(1)}$, $U^{(2)}$, $P^{(1)}$ and $P^{(2)}$, an iterative algorithm is proposed, where the goal is to find a common data sample representation P_c. The first step of the algorithm is to find an estimate of $U^{(1)}$ and $U^{(2)}$ applying NMF, as described by the following objective function

$$\min_{U^{(1)},U^{(2)},P_c} \|\hat{X}^{(1)} - P_c U^{(1)}\|_F^2 + \|\hat{X}^{(2)} - P_c U^{(2)}\|_F^2 + \lambda\|P_c\|_1 \tag{1.25}$$

where λ is a user defined parameter. This minimization stage produces the initial values for $U^{(1)}$, $U^{(2)}$ and P_c. Afterwards, an iterative adjustment of either $\{U^{(1)}, U^{(2)}\}$ or $\{P^{(1)}, P^{(2)}, P_c\}$ is performed until all three sample representations match, i.e. $P^{(1)} = P^{(2)} = P_c$. On each iteration, the first step is to adjust $P^{(1)}$, $P^{(2)}$ and P_c by fixing $U^{(1)}$ and $U^{(2)}$ and optimizing the following objective functions

$$\min_{P^{(1)}\geq 0} \|\hat{X}^{(1)} - P^{(1)}U^{(1)}\|_F^2 + \lambda\|P^{(1)}\|_1 \tag{1.26}$$

$$\min_{P^{(2)}\geq 0} \|\hat{X}^{(2)} - P^{(2)}U^{(2)}\|_F^2 + \lambda\|P^{(2)}\|_1 \tag{1.27}$$

$$\min_{P_c\geq 0} \|\hat{X}^{(1)} - P_cU^{(1)}\|_F^2 + \|\hat{X}^{(2)} - P_cU^{(2)}\|_F^2 + \lambda\|P_c\|_1 \tag{1.28}$$

afterwards, $P^{(1)}$, $P^{(2)}$ and P_c are fixed in order to adjust $U^{(1)}$ and $U^{(2)}$ by optimizing the following objective functions

$$\min_{U^{(1)}\geq 0} \|\hat{X}^{(1)} - P^{(1)}U^{(1)}\|_F^2 \tag{1.29}$$

$$\min_{U^{(2)}\geq 0} \|\hat{X}^{(2)} - P^{(2)}U^{(2)}\|_F^2 \tag{1.30}$$

This iterative process is repeated until the aforementioned condition $P^{(1)} = P^{(2)} = P_c$ is met. The common and complete representation of the data samples P_c is then passed to a standard clustering method in order to obtain the clustering assignment.

Large scale multiview spectral clustering via bipartite graph [70] is a proposed method focused on processing large datasets and allowing out of sample operation, which standard spectral clustering does not allow. This method works as follows. First, a set of m *salient* points U is selected from the dataset using K-means on the concatenated features from all the input views. These points are supposed to capture the structure of the manifold while requiring much less data to represent it. Then, a bipartite graph is generated between the complete set of data points and the points in U. This graph is constructed as a k-NN graph between the original data points X and the salient points U. The weight of each edge of such graph is given by the expression

$$w_{ij} = \frac{K(x_i, u_j)}{\sum_{k\in\Phi_i} K(x_i, u_k)} \tag{1.31}$$

where $x_i \in X$ are the original data points, $u_j \in U$ are the salient points, K is a kernel function (the Gaussian kernel in the experiments described in the paper), and Φ_i are the indices of the s-nearest neighbours of points x_i in U. s is a user defined parameter. The result is a weight matrix W. Then, the first n eigenvectors of the Laplacian matrix of W are computed and K-means is applied to the matrix of eigenvectors in order to obtain the final clustering assignment with n clusters.

1.3.3 Ensemble clustering methods

Ensemble clustering methods independently run a standard, single view clustering algorithm on each of the input data views, and then merge the resulting

clustering assignments. In other words, they merge the multiview information in the last step of the clustering process. A survey of ensemble clustering methods can be found in [104].

A multiview K-means algorithm is proposed in [11], where the ℓ_2-norm is replaced by the $\ell_{2,1}$-norm in order to avoid the influence of the outliers in the performance of the K-means algorithm. The K-means problem is reformulated into the following expression

$$\min_{F^{(v)},G,\alpha^{(v)}} \sum_{v=1}^{M} (\alpha^{(v)})^{\gamma} \|X^{(v)^T} - GF^{(v)^T}\|_{2,1} \tag{1.32}$$

where M is the number of input feature matrices, $X^{(v)}$ are each of the input feature matrices, $\alpha^{(v)}$ is a weight coefficient associated to the v-th matrix, $F^{(v)}$ is the matrix of centroids of the v-th matrix and G is the clustering assignment vector. Therefore, this algorithm requires solving an optimization problem that requires finding the optimal values for $\alpha^{(v)}$, G and $F^{(v)}$. Moreover, parameter γ has to be chosen by the user in order to tune the behaviour of the algorithm.

1.3.4 Other multiview clustering methods

The method presented in [117] does not fall into the categories presented above, as it iterates over the different multiview clustering phases (views, projection, clustering) in order to achieve its goal. This method can only process two-view datasets. The cotraining framework for multiview clustering (CoKmLDA) algorithm is detailed in Algorithm 1.

The multi-feature spectral clustering with minimax optimization method [106] tries to find a common feature embedding that unifies the different input features while minimizing the disagreement cases among them. This is performed in four steps. First, the normalized Laplacian matrix and standard spectral clustering is computed for each input view, obtaining a possibly different clustering assignment per view $V \in \mathbb{R}^{N \times K}$, where N is the number of samples and K the number of input views. Second, the regularized data-cluster similarity is computed as $P_V(U) = UU^TV$, where U are the K first eigenvectors of the normalized Laplacian matrix of each input view as computed in the first step. On the third step, an agreement among the different P_V matrices is searched. Deriving from the previous definitions,the normalized Laplacian matrix of a specific view i with respect to view j is

$$\mathcal{L}_{ij} = I - \text{sym}(U_i U_i^T U_j U_j^T) \tag{1.33}$$

Algorithm 1 . Cotraining framework for multiview clustering

Input: dataset $\mathcal{X} = \{x_1, x_2\}$,

expected number of clusters: K,

Output: clustering assignment vectors (one for each input view)

function CoKmLDA($\mathcal{X}, K$)

Perform K-means on each view to obtain cluster assignments $H^{(v)}$

For each view v, identify the data samples closest to each of the K cluster centroids, obtaining $S^{(v)} = s_1^{(v)}, s_2^{(v)}, \ldots, s_K^{(v)}$

for $t \leftarrow 1$ **to** $numIters$ **do**

for $v \leftarrow 1$ **to** 2 **do**

Use $X^{(v)}$ and $H^{(3-v)}$ to train LDA; project the samples on LDA space

Use $S^{(v)}$ to perform K-means on the projected samples to estimate new cluster assignments $H^{(v)}$

Update $S^{(v)}$

end for

end for

end function

where $\text{sym}(A) = (A + A^T)/2$. Then the pairwise disagreement costs between samples i and j are defined as

$$\mathcal{Q}_{ij} = \text{tr}(V^T \mathcal{L}_{ij} V) \tag{1.34}$$

The objective function to find the optimal $\mathcal{Q}_{ij}$ is

$$\begin{aligned} \min_{\{U_m\}_{m=1}^M} \max_{\{\alpha_{ij}\}_{j\geq i}^M} & \sum_{j=i}^{M} \sum_{i=1}^{M} \alpha_{ij}^{\gamma} \mathcal{Q}_{ij} \\ \text{subject to} \quad & \alpha_{ij} \in \mathbb{R}^+, \ \sum_{j=i}^{M} \sum_{i=1}^{M} \alpha_{ij} = 1, \\ & U_m \in \mathbb{R}^{N\times K}, \ U_m^T U_m = I, \\ & V \in \mathbb{R}^{N\times K}, \ V^T V = I \end{aligned} \tag{1.35}$$

where γ is a user-defined parameter. The fourth and final step of the method is an iterative algorithm to optimize the previous objective function and find an optimal low-dimensional representation of all views V. Finally, k-means is executed on V to obtain the final clustering assignment.

1.4 Open issues in multiview unsupervised pattern recognition methods

There exist a series of open issues regarding the application of unsupervised pattern recognition methods to multiview data. These issues condition the methods designed in this thesis, as well as the experiments performed therein. The most relevant of these open issues are:

- There is no formal framework that demonstrates the advantages of using multiview datasets and methods in front of using single-view datasets and methods. The experimental comparisons in the state of the art are not complete and general enough.
- There are no evaluation metrics specifically designed to evaluate the performance of multiview pattern recognition methods. Moreover, there is not even a consensus on which evaluation metric to use to evaluate dimensionality reduction experiments.
- There are no clear criteria to decide, given a multiview dataset, which views to use and which to exclude in order to achieve better results. This would be an equivalent step to variable selection techniques, but applied on a per-view basis.

1.5 Thesis objectives

The general objective of the present thesis is to develop a set of unsupervised pattern recognition methods that can be applied to any multiview dataset, and to study if these methods involve a meaningful improvement over existing single-view methods. More specifically,

- To develop a multiview extension to the t-SNE dimensionality reduction method so that it can be applied to multiview datasets.
- To develop a multiview extension to the MDS dimensionality reduction method so that it can be applied to multiview datasets.
- To develop a multiview extension to the spectral clustering and Laplacian eigenmaps methods (which are very closely related) so that they can be applied to multiview datasets.
- To design and execute a set of experiments to properly test the proposed multiview methods, including the selection of the proper datasets, evaluation metrics, baseline reference methods and comparable methods in the state of the art.

- To analyze the performance of the proposed multiview methods with respect to single-view equivalent methods, in order to determine the potential advantages of the multiview approaches.
- To analyze the performance of the proposed multiview methods with respect to equivalent multiview methods in the state of the art, in order to assess its usefulness for the community.
- To compare the performance of the three proposed methods, in order to provide some guidelines for deciding which method is more suitable for a given pattern recognition problem.
- To generate a software package that makes the proposed methods readily available to the community using a widespread software environment.

1.6 Structure of this thesis

This thesis is organized into eight chapters plus three appendices. Although three novel methods are proposed in this thesis, the experimental setup is the same for all of them. Therefore, Chapter 2 explains the design and setup of the experiments that will be applied to all the methods.

Chapter 3 describes the first proposed method, a multiview extension of the t-SNE dimensionality reduction method, along with the results of this method on the experiments and a discussion of these results.

Chapter 4 describes the second proposed method, a multiview extension of the MDS dimensionality reduction method, along with the results of this method on the experiments and a discussion of these results.

Chapter 5 describes the third proposed method, a multiview extension of both spectral clustering and Laplacian eigenmaps, along with the results of this method on the experiments and a discussion of these results.

After describing and analyzing the three proposed methods individually, Chapter 6 presents a comparison between all three methods and a discussion of these results.

The software package created to release the three proposed methods to the community is described in Chapter 7.

The overall conclusions of the present thesis are presented in Chapter 8.

Finally, the three appendices contain the tables with the detailed results of the different experiments performed throughout the thesis. As the number of tables is considerable, they are presented in the appendices to keep the corresponding chapters in a manageable size, clear and straightforward.

Chapter 2

Experimental setup

2.1 Motivation

In the present thesis three methods are presented. All three methods perform dimensionality reduction of multiview data into a single, low-dimensional space, therefore condensing a considerable amount of information into a much lower number of dimensions.

The fact that all three methods perform the same tasks makes it convenient to run the same experiments on all of them. Moreover this will allow to make a final comparison among all three methods, as presented in Chapter 6. These are the reasons why the design of the experiments is presented first.

2.1.1 Description of the experiments

The methods presented in this thesis are, in essence, dimensionality reduction methods of multiview data. Therefore, some of the experiments presented are dimensionality reduction experiments on multiview datasets. However, the unique, low-dimensional data representation that is produced by the methods is also convenient for the clustering of the original data samples. As a consequence, clustering experiments have also been made in order to assess the validity of the methods in the multiview clustering problem.

The experiments presented in this thesis are divided in two categories. The first category are the **baseline experiments**, where the proposed multiview methods are compared with the respective single-view versions of the algorithms. More specifically, the single-view algorithms are applied to all the input views of each dataset independently, plus to the concatenation of all the input view matrices (referred as the *stacked* configuration). The goal of these baseline experiments is to analyze if the multiview methods proposed involve an improvement over their single-view counterparts, therefore justifying the use of multiview data and methods. The design of these baseline experiments is thoroughly described in Section 2.4.

The second category of experiments are the **state of the art experiments**, where the proposed multiview methods are compared with equivalent multiview methods in the state of the art. The goal of these experiments is to assess if the proposed multiview methods are an improvement over other multiview methods in the state of the art. The design of these state of the art experiments is described in Section 2.5.

2.1.2 Structure of this chapter

The remainder of this chapter includes the following contents. First, a description of the multiview datasets used in the experiments. Second, the evaluation metrics used to assess the quality of the different results are presented. Third, the detailed design of the baseline experiments and their goals is given. Finally, the design of the state of the art experiments is presented.

2.2 Dataset description

The multiview datasets used in the experiments belong to three categories: text documents, image and biological. However, each of the datasets on each category have different features that make them convenient to test different aspects of the performance of the algorithms.

The criteria used to select these datasets have been the following. First, they are multiview datasets, i.e. they are published with multiple views or feature matrices, thus allowing anyone interested in reproducing the present experiments to use exactly the same data. Second, they are provided to the community by a recognized institution and research team and are backed by peer-reviewed publications. Finally, they are used by other methods in the state of the art so the results can be compared

2.2.1 Text datasets

Obtaining multiple views from text documents can be accomplished in several ways, and the datasets used in the experiments reflect a different multiview approach. Their quantitative details are given in table 2.1.

First, the BBC News multiview text collection [44, 43] [1]. It comprises 2,225 news articles labelled with one of five possible topics (*business*, *entertainment*, *politics*, *sport* or *tech*). The input texts are split into several segments. The term frequencies on each segment become the different input views. There are several subsets in the original data set. The two-segment subset has been chosen to allow direct comparison with the results in the literature. The number of terms in each view, i.e. the number of attributes, is 6,838 and 6,790 respectively, although only the 500 most frequent terms on each segment

[1] `http://mlg.ucd.ie/datasets/bbc.html`

	BBC	Reuters	Cora
View 1	Seg. A (500/6,838)	English (500/21,531)	Bag of words (1,433)
View 2	Seg. B (500/6,790)	French (500/24,892)	References (2,708)
View 3	—	German (500/34,251)	—
View 4	—	Italian (500/15506)	—
View 5	—	Spanish (500/11547)	—
No. of samples	2,112	18,758	2,708
No. of samples used	2112	6000	2,708
No. of classes	5	6	7

Feature name (used variables/number of variables in the feature matrix). A single number means that all available variables have been used.

Table 2.1: Summary of the text multiview datasets

are used as the less frequent terms do not contribute to the quality of text classification [51]. The *tf.idf* (term frequency / inverse document frequency) [78] is computed on each of the input segments, and the cosine similarity is used instead of the euclidean distance because of the high sparsity of the feature matrices.

The second text dataset is the Reuters multilingual corpus [3][2], a set of 18,758 news articles available in five different languages (English, French, German, Italian and Spanish). The subset of original English news articles has been used; the term matrices of the remaining languages come from machine-translated texts. The texts belong to one out of six news categories. For each input view (language), a matrix with term frequencies is given. As with the *BBC news* dataset, only the 500 most frequent terms of each language have been used. Their *tf.idf* value has been computed and finally the cosine similarity has been employed to find the similarity matrices.

The third text dataset used in the experiments is the Cora dataset [79][3], which contains $2,708$ scientific publications classified into one of seven classes. This dataset has two views. The first one is a bag of words with $1,433$ words. The second view is a reference graph that represents $5,429$ links between the documents.

2.2.2 Image datasets

Although the number of image datasets used in the literature is huge, few of them are specifically multiview or multifeature in the sense of providing different sets of features for each image; often these datasets simply contain raw images. The two multiview image datasets selected, on the contrary, pro-

[2] `https://archive.ics.uci.edu/ml/datasets/Reuters+RCV1+RCV2+Multilingual+Multiview+Text+Categorization+Test+collection`

[3] `https://linqs.soe.ucsc.edu/`

Table 2.2: Summary of the image datasets

	Digits	AWA
View 1	Pixels (240)	CQ (2,688)
View 2	Fourier coeffs. (76)	LSS (2,000)
View 3	Profile correl. (216)	PHOG (252)
View 4	Zernike coeffs. (47)	SIFT (2,000)
View 5	Karhunen moments (64)	RGSIFT (2,000)
View 6	Morph. feats. (6)	SURF (2,000)
No. of samples	2,000	30,475
No. of samples used	2,112	4,000
No. of classes	10	50

Feature name (number of variables in the feature matrix).

vide different image features. The main difference between these two datasets stems from the original images: the first dataset (Digits), derives from handwritten numerals in grayscale tonalities, while the second dataset (Animal with attributes, or AWA) contains features extracted from real-world, color photographs. As a consequence, the specific feature types extracted and their values greatly differ from one dataset to the other. The details of these datasets are given in table 2.2.

The University of California at Irvine (UCI) multiple features digits dataset [9], available at the UCI machine learning repository,[4] is created from a set of handwritten numerals (from '0' to '9'), scanned as 15×16 grayscale pixels images. There are 200 samples of each numeral, resulting in a total of 2,000 samples. The data set provides six different views or feature sets of the original image data: (1) the pixel averages in 2×3 windows, (2) 76 Fourier coefficients of the character shapes, (3) 216 profile correlations, (4) 64 Karhunen-Love coefficients [99], (5) 47 Zernike moments [71], and (6) 6 morphological features (not specified).

The other image dataset used in the experiments is the Animal with attributes data set (AWA)[65], [5] which is a multiple feature data set with six standard image features extracted from animal photographs. This dataset includes photographs from 50 different animal species, which become the classes of the data samples. Due to the high number of classes, it is particularly hard to achieve high evaluation scores with this dataset in its original configuration.

[4] `https://archive.ics.uci.edu/ml/datasets/Multiple+Features`

[5] `http://attributes.kyb.tuebingen.mpg.de/`

Table 2.3: Summary of the biological dataset

	Protein
View 1	Hydrophobicity FFT (fft) (1,040)
View 2	Gene expression (expr) (1,040)
View 3	Pfam hidden Markov model (pfam) (1,040)
No. of samples	1,040
No. of classes	3

Feature name (number of variables in the feature matrix).

2.2.3 Biological dataset

The Berkeley protein dataset [66] [6] is a multiview dataset of 1,040 yeast proteins. These proteins are labeled by their location, as either membrane proteins, ribosomal proteins, or other. This dataset comprises 8 data views or feature sets. In order to reproduce the experiments in the literature, only three feature sets have been used: the hydrophobicity fast Fourier transform (fft), the gene expression (expr) and the expectation values derived from hidden Markov models in the Pfam database [95] (pfam). All these feature sets are expressed as relationships between proteins, therefore they are similarity matrices of $1,040 \times 1,040$ values. Table 2.3 presents the details of the views used in the experiments.

2.3 Evaluation of the experiments

As stated above, the methods presented in this thesis will be used as both multiview dimensionality reduction methods and multiview clustering methods. This implies that the evaluation of the experiments has to use quality metrics suited to each of these two tasks. In order to increase the coverage of the evaluation tests, supervised and unsupervised metrics have also been used. The set of evaluation metrics used in the experiments are described next.

2.3.1 Dimensionality reduction metrics

The quality of the low-dimensional projection of a dataset generated by a dimensionality reduction method can be assessed in several ways. If the data samples are labeled, then a supervised evaluation method can be used. More specifically a classifier is trained on the projection and the classification accuracy measured in order to produce an estimate of the quality of the projection with respect to the labeling. The advantage of this method is that it can be equally applied to multiview data and methods, as only the class labels are considered.

[6] `http://noble.gs.washington.edu/proj/sdp-svm/`

Another, unsupervised approach, measures some kind of similarity of the low-dimensional projection to the original, high-dimensional data. Although this is a straightforward procedure for single-view data, in multiview data the similarity measure can be computed with respect to the different input views. The solution adopted is to compute the similarity with respect to all the input views and give the average similarity as overall quality measure.

The three methods to evaluate the quality of dimensionality reduction results that are used in the experiments are described next.

2.3.1.1 One-vs-one classifier

The first method presented in this Section is one of the standard procedures in the literature to evaluate the quality of a low-dimensional projection of a dataset. It comprises two steps. First, a one-vs-one support vector machine (SVM) [50] is used to classify all the points in the dataset. One-vs-one SVMs train one classifier for each possible pair of classes in the dataset, i.e. they train $c(c-1)$ classifiers, where c is the number of classes in the dataset. The second step is to measure the classification accuracy. A point is well classified if the majority of the $c(c-1)$ classifiers trained assign it to its true class. The overall classification accuracy is the final evaluation metric for the low-dimensional projection of the data. The value of the accuracy ranges from 0 to 1, with 1 being a perfect classification. This method will be referred as **SVM** in the results. This evaluation method is used, among others, in [114, 115].

2.3.1.2 Cophenetic correlation

As the second evaluation method, the cophenetic correlation between the distances in the high-dimensional input space and the low-dimensional projection space is computed. The cophenetic distance [94] measures how faithfully the distances between a set of points is kept in an alternate representation. This method was originally developed for distance in dendrograms, in the field of biostatistics. Given two data points i, j, if the distance between them in the input high-dimensional space is δ_{ij} and the distance between them in the projection low-dimensional space is d_{ij}, then the cophenetic correlation is

$$c = \frac{\sum_{i<j}(\delta_{ij} - \bar{\delta})(d_{ij} - \bar{d})}{\sqrt{[\sum_{i<j}(\delta_{ij} - \bar{\delta})^2][\sum_{i<j}(d_{ij} - \bar{d})^2]}} \tag{2.1}$$

Given that the datasets used in the experiments have multiple views, the average cophenetic correlation of all the input data views with respect to the output projection is computed and referred as **coph**.

2.3.1.3 Area under the local neighbourhood agreement curve

The third evaluation method is proposed by [68], and the implementation of R package *dimRed*[7] is used. Its goal is to measure the conservation of the local K-ary neighbourhoods of the original data points in the low-dimensional projection. The rank of data point p_j with respect to point p_i in the original high-dimensional space is defined as

$$\rho_{ij} = |\{k : \delta_{ik} < \delta_{ij} \text{ or } (\delta_{ik} = \delta_{ij} \text{ and } 1 \leq k \leq j \leq N\}| \tag{2.2}$$

where δ_{ij} is the distance between points p_i and p_j in the high dimensional space and N is the number of points in the dataset. The rank r_ij between points p_j and p_i in the low-dimensional space is similarly defined, using the corresponding distances in that space. The K-ary neighbourhood of point p_i in the high-dimensional space is the set of points defined by $v_i^K = \{j : 1 \leq \rho_{ij} \leq K\}$. The K-ary neighbourhood n_i^K of p_i in the low-dimensional space is similarly defined using r_{ij} instead of ρ_{ij}. The performance index Q_{NX} is defined as

$$Q_{NX}(K) = \sum_{i=1}^{N} \frac{|v_i^K \cap n_i^K|}{KN} \tag{2.3}$$

which gives a normalized average agreement between the neighbourhoods in high and low-dimensional spaces, ranging from 0 (no agreement at all) to 1 (perfect agreement). Another index derived from Q_{NX} is given next, that accounts for the random neighbour matches that may exist

$$R_{NX} = \frac{(N-1)Q_{NX}(K) - K}{N - 1 - K} \tag{2.4}$$

in this case, a random neighbourhood assignment (i.e. a random projection) would produce an R_{NX} value of 0. Iterating over different values of K produces a curve. The area under this curve (AUC) for $1 \leq K \leq \kappa$, where κ is the number of dimensions of the original data space, is proposed as a measure of the quality and consistence of the dimensionality reduction method.

As this index depends on the points in the input space, a different index will be obtained for each input view in a multiview setting. The results given, referred as **AUC-RNX**, are the average area under the curve of the R_{NX} index of the low-dimensional space with respect to each of the input high-dimensional spaces.

2.3.2 Clustering quality metrics

In order to evaluate the quality of a clustering assignment, two approaches are possible. If the data is labeled then an *external* evaluation is feasible, where

[7] `https://cran.r-project.org/web/packages/dimRed/index.html`

the clustering assignment produced is compared with the original class labels of the data. On the other hand, *internal* clustering evaluation methods do not rely on class labels but rather on the intrinsic properties of the data and the clustering assignment. Both kinds of evaluation methods are used in the present experiments.

As there are multiple input matrices on each dataset but a single class labeling, external methods can be applied as they are, while internal methods have to be applied to each input data view. In the latter case, the average indices obtained are given as final result.

2.3.2.1 Clustering purity

Clustering *purity* [78] is an external measure of clustering quality that measures the agreement of the clustering assignment with the original classes. In order to compute the clustering purity, each cluster w_i is assigned to the class most frequent in the cluster (c_j). The accuracy of this assignment is measured by counting the number of correctly assigned data samples, and dividing by the number of data samples N:

$$purity(\Omega, \mathbb{C}) = \frac{1}{N} \sum_i \max_j |w_i \cap c_j| \tag{2.5}$$

where $\Omega = \{w_1, w_2, \ldots, w_k\}$ is the cluster assignment and $\mathbb{C} = \{c_1, c_2, \ldots, c_k\}$ are the original class labels. The clustering purity is referred in the results as **purity**.

2.3.2.2 Normalized mutual information

The normalized mutual information (NMI) [97] is another external clustering evaluation metric. It is defined as

$$NMI(\Omega, \mathbb{C}) = \frac{I(\Omega, \mathbb{C})}{[H(\Omega) + H(\mathbb{C})]/2} \tag{2.6}$$

where I is the mutual information, H is Shannon's entropy, $\Omega = \{w_1, w_2, \ldots, w_k\}$ is the cluster assignment and $\mathbb{C} = \{c_1, c_2, \ldots, c_k\}$ are the original class labels. The mutual information numerator computes how informative is the clustering assignment of the real classes. The problem is that larger numbers of clusters produce better results. For this reason, the normalization denominator penalizes having a high number of clusters (as the entropy tends to increase with the number of clusters). This way, a NMI of zero implies a random clustering assignment, while a NMI of 1 implies a perfect clustering assignment.

2.3.2.3 Davies-Bouldin index

The Davies-Bouldin index (DBI) [26] is an internal clustering quality measure that measures the compactness of each cluster with respect to the other clusters in the assignment. It is defined as

$$\text{DBI} = \frac{1}{K}\sum_{i=1}^{K}\max_{j\neq i}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right) \tag{2.7}$$

where K is the number of clusters, c_n is the centroid of cluster n, σ_n is the average distance of the points in cluster n to its centroid, and $d(c_i, c_j)$ is the distance between centroids of clusters i and j.

Better clustering assignments yield lower DBI values. However, the reliability of the DBI is limited by the meaningfulness of the distance metric (Euclidean by default) on the input data.

2.4 Design of the baseline experiments

The experiments required to test the methods presented in this work with the baseline, single-view methods must account for several factors involved.

In general lines, the experiments presented in this thesis are designed with the following objectives or principles in mind:

- Demonstrate the performance of the methods on an adequate number of datasets. These datasets must have heterogeneous characteristics so as to account for a wide range of cases.
- Thoroughly test the three methods presented in this work.
- Test the methods as both dimensionality reduction and clustering methods.
- Use several evaluation metrics to obtain more solid results.
- Test the response of the method with different dimensionalities of the projections generated.
- Carefully design the experiments to account for possible sources of randomness in the methods or the evaluation metrics.
- Properly compare the proposed methods with single-view counterparts, either on individual views or on matrices of stacked or concatenated features.
- Properly compare the proposed methods with other multiview methods in the literature.

Consequently, in order to fulfill these objectives the experiments presented in this work are designed as follows. Table 2.4 summarizes the factors involved in the design of the experiments.

Datasets Six datasets are used in the experiments, and their subjects are: image, text and biology. Moreover, even if image and text categories have more than one dataset, they are not redundant as on each dataset the design of the multiview aspect is different, as explained in Section 2.2.

Methods The three methods presented in this thesis are used in all the experiments, in order to perform a thorough experimental analysis and comparison of all three. Moreover, the methods are used in both dimensionality reduction and clustering tasks to assess their performance on each of these tasks.

Evaluation Six evaluation metrics, presented in Section 2.3, are used in all the experiments in order to properly capture as many aspects as possible about the quality of the output of each method. More specifically, three metrics are used to evaluate the dimensionality reduction task, and other three metrics are used to evaluate the clustering task. In each category, both supervised and unsupervised metrics are used to increase the scope of the evaluation.

As a side note, the unsupervised evaluation methods cannot be applied to the Cora dataset because they are not compatible with graph space input views (Cora's references between articles).

Dimensionality of the projection Both in dimensionality reduction and clustering tasks, the dimensionality of the projection generated is a decisive factor in the performance of the methods and the quality indicators obtained. As a consequence, all the experiments are run on 20 different dimensionality values, ranging from 2 to 100.

Account for randomness Some of the methods presented or analyzed in this work have an intrinsically random behaviour, specifically SNE [49] and their derivations. This causes them to output possibly different results on each execution. On the other hand, some evaluation metrics also have a random component, as 1vs1-SVM (Section 2.3), because it requires to partition the data in train and test subsets. This partition is randomly executed, and consequently the results of different executions may vary.

To account for the randomness of methods and evaluation metrics, the affected experiments are repeated ten times and the mean and standard deviation are measured to try to minimize the impact of the random factor in the final results.

Factor	Magnitude
Datasets	6
Methods proposed	$\times(2 + 10 \times 1)$
Baseline methods	$\times((2+1)+(5+1)+(2+1)+(6+1)+(6+1)+(3+1))= \times 28$
Tasks	$\times 2$
Evaluation metrics	$\times(5 + 10 \times 1)/2$
Dimensions	$\times 20$
Total experiments	604800

Table 2.4: Factors in the design of the baseline experiments

Exhaustive comparisons The multiview methods presented in this work are carefully compared with other methods in order to assess their performance and potential usefulness. First, they are compared with their single-view counterparts, either using a single view or using a matrix with the stacked features of all the input views. As a consequence the number of experiments is equal to the number of views in the dataset (each single view) plus one (the stacked feature matrix). The goal of this comparison is to estimate if the multiview approach is better than the standard single-view methodology, and also to estimate if processing each view separately is better than concatenating all the views and processing them as a single input view.

On the other hand, the proposed methods are compared with equivalent multiview methods in the literature in order to measure their performance with respect to existing multiview methods and to decide if they are a significative contribution. This experiment design subgoal is addressed in the state of the art experiments, described in Section 2.5.

The experiments performed for the present work result from the combination of the above factors. The results obtained by each method are presented in the respective chapter where each method is described. Moreover, Chapter 6 includes a comparison between the three methods presented in this thesis.

2.5 Design of the state of the art experiments

In order to allow the comparison of the method presented in this thesis with equivalent methods in the literature, a second set of experiments has been designed. These experiments reproduce as faithfully as possible the experimental setup of the reported papers in the state of the art.

The most frequently cited datasets in the literature are used for these comparison experiments. These datasets are the BBC News text collection, the Reuters multilingual corpus, the multiple features digits and the animal with

attributes (AWA) datasets. They are described in Section 2.2. However, the AWA experiments in the state of the art differ from the general experiments with AWA presented in this thesis as they follow the preprocessing proposed in [106]: (1) use only the 10 most populated classes, and (2) use all available samples, giving a total of $18,450$ samples. These changes cause the higher results yielded by these experiments, specially regarding supervised evaluation metrics.

The multiview clustering papers in the literature follow the evaluation methodology described in [78] and use the clustering purity and the clustering normalized mutual information as clustering evaluation metrics. These supervised quality metrics are described in Section 2.3.

The proposed methods are compared with the most relevant multiview clustering methods in the state of the art, namely: co-regularized spectral clustering [62] (**CoregSC**), multi-modal spectral clustering [12] (**MMSC**), multiview clustering via structured sparsity [106] (**MVC-SS**), multiview K-means clustering [11] (**MV-KMeans**), subspace co-training for multiview clustering [117] (**CoKmLDA**), multi-feature spectral clustering with minimax optimization [106] (**MFSC-MO**), multiview clustering via pairwise sparse subspace representation [112] (**MVC-PSS**), and large-scale multiview spectral clustering via bipartite graph [70] (**MVSC-BG**).

No reproducible results have been found in the state of the art for multiview dimensionality reduction methods, and as a consequence no state of the art comparison experiments on dimensionality reduction have been performed.

Chapter 3

Multiview t-distributed stochastic neighbour embedding

3.1 Motivation

t-distributed stochastic neighbour embedding (t-SNE) [101, 102, 74] is one of the most popular dimensionality reduction methods among the community. It is specifically designed to achieve good two or three dimensional projections of complex data, intended for data visualization. Its goals are to keep both the local structure (neighbourhood of each sample) and the global structure (keep distant samples away in the low-dimensional representation).

t-SNE is itself an evolution of the stochastic neighbour embedding method (SNE) [49] that solves some of its limitations by changing the method to compute the output, low-dimensional space.

Given the popularity and usefulness of t-SNE, an extension of this method to multiview datasets may be useful to the community. Its objective would be to take several input data spaces on the same samples and generate a single, low-dimensional output space while preserving both the local and global structure of all the input views as much as possible.

In this chapter, two multiview extensions to t-SNE are proposed, one using multi-objective optimization techniques, the other using expert opinion pooling theory methods. Although both proposals solve the given problem, the first algorithm is too expensive in computational terms and consequently only the second algorithm is considered at the experimenal phase.

3.2 Related work

3.2.1 Stochastic neighbour embedding

Stochastic neighbour embedding (SNE)[49] is a non-linear dimensionality reduction method. The main intuitions behind SNE are (a) to view the distances between data points as probabilities of being neighbours, and (b) to try to find a low dimensional space $\mathcal{Y}$ whose neighbouring probabilities are as close as possible to the probabilities of the original, high dimensional space $\mathcal{X}$. The formal description of the method follows.

First step. SNE begins by converting the Euclidean distances between the points of the original space $\mathcal{X}$ into conditional probabilities. Given two points $x_i, x_j \in \mathcal{X}$, the conditional probability $p_{i|j}$ is computed using a Gaussian probability density centered at x_i, with a variance σ_i whose determination is described later in this section. $p_{i|j}$ can be interpreted as the probability that x_i would have x_j as its closest neighbour. Closer points have a high probability, while points far apart have an almost zero probability. The mathematical expression of $p_{i|j}$ is:

$$p_{j|i} = \frac{exp(-\|x_i - x_j\|^2/2\sigma_i^2)}{\sum_{k \neq i} exp(-\|x_i - x_k\|^2/2\sigma_i^2)} \tag{3.1}$$

as a result, for each $x_i \in \mathcal{X}$ a vector of probabilities P_i of length $|\mathcal{X}|$ is obtained, where $p_{i|i}$ is set to 0; otherwise the high similarity of each point with itself would dominate over the other neighbouring probabilities. Stacking the vectors of probabilities produces a matrix P sof $|\mathcal{X}| \times |\mathcal{X}|$ probabilities whose diagonal is zero.

As stated above, σ_i has to be defined for each datapoint x_i. This value is not defined directly, but through *perplexity*, a user-defined global parameter. Intuitively, the perplexity is the average number of neighbours per point, with typical values between 5 and 50. A different σ_i is chosen for each point to account for the different density of points on the different areas of the input space.

The perplexity is defined as follows:

$$Perp(P_i) = 2^{H(P_i)} \tag{3.2}$$

where $H(P_i)$ is the Shannon entropy [90] of P_i measured in bits, i.e.

$$H(P_i) = -\sum_i p_{j|i} log_2 p_{j|i} \tag{3.3}$$

SNE performs a binary search of each σ_i so that the perplexity condition is satisfied. Points in sparse regions will naturally have a higher σ_i than points in denser regions.

Second step. On the low dimensional space to be produced, $\mathcal{Y}$, a similar conditional probability is computed for all the pairs of projected points y_i and y_j, denoted by $q_{i|j}$. The main difference with respect to $p_{i|j}$ is that SNE computes $q_{i|j}$ using a constant $\sigma = \frac{1}{\sqrt{2}}$ for all the points in the dataset. Therefore the formulation of this conditional probability is:

$$q_{j|i} = \frac{exp(-\|y_i - y_j\|^2)}{\sum_{k \neq i} exp(-\|y_i - y_k\|^2)} \tag{3.4}$$

As in the previous case, $q_{i|i} = 0$. The vector of conditional probabilities associated with each projected point y_i is referred to as Y_i, and the whole matrix of conditional probabilities is Y, with the elements in its diagonal equal zero.

Third step. The hypothesis of SNE is that if the conditional probabilities $p_{i|j}$ and $q_{i|j}$ are similar then the low dimensional space $\mathcal{Y}$ should be a faithful representation of the original, high dimensional space $\mathcal{X}$. A suitable measure to compare probability densities is the Kullback-Leibler divergence (KL) [103] between each probability density vector P_i and Q_i. SNE has to find a Q that minimizes the following the sum of such KL divergences, i.e. the following cost function:

$$C = \sum_i KL(P_i \| Q_i) = \sum_i \sum_j p_{j|i} log \frac{p_{j|i}}{q_{j|i}} \tag{3.5}$$

SNE uses gradient descent optimization to minimize C, i.e. it has to find the projection $\mathcal{Y}$ of the original data points whose conditional probability matrix Q minimizes C. The gradient of the cost function defined in Equation 3.5 is:

$$\frac{\delta C}{\delta y_i} = 2 \sum_j (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})(y_i - y_j) \tag{3.6}$$

The gradient descent algorithm is initialized by using a random Gaussian distribution of points for y_i, which progressively gets modified until either a minimum is reached or a predefined number of iterations is executed.

Analysis of the algorithm. As a consequence of the asimmetry of the KL divergence, not all differences between the pairwise conditional probabilities in P and Q are equally relevant. More specifically, having two close points x_i and x_j projected in distant positions y_i y_j has a higher cost than the inverse situation, i.e. representing two separate points in relative proximity in the projection space $\mathcal{Y}$. In other words, SNE is designed to keep the local structure of the data (close points remain close to each other) but not as much the global structure (far points may be represented close to each other).

SNE cost function is non-convex, what leads to the necessity of running the algorithm several times in order to avoid local minima and possibly obtain better data projections.

3.2.2 t-distributed stochastic neighbour embedding

An inherent problem when reducing the dimensionality from high-dimensional spaces to low-dimensional spaces (often two or three dimensions if the objective is to plot the points) is the problem of point crowding. In high-dimensional spaces there is much more room for points to be moderately apart from each other. However, representing such intermediate distances in a low-dimensional space implies using relatively large distances to properly represent the points and avoid projecting them in nearby positions. In turn, neighbouring points in the original space would have much smaller distances in the projected space, leading them to be almost overlapped an irrealistically high concentration of points in the center of the plot.

Several solutions have been proposed to address this issue. However none of them focuses on the cause of the problem: the conditional neighbouring probabilities in the low-dimensional space do not follow the same probability density function as the probabilities in the high-dimensional space. This stems from the fact that distant points are required to have a proportionally larger distance in the low-dimensional space to avoid the crowding effect. Therefore, [102, 74] proposed to use a different conditional probability density function to model the low-dimensional distances. This improvement on SNE, called t-Distributed Stocastic Neighbour Embedding (tSNE), uses a t-distribution with one degree of freedom to model the conditional probabilities in the low-dimensional space. The t-distribution has heavier tails than the Gaussian, therefore it allows to model moderate distances in the original space as larger distances in the low-dimensional space. In turn, this helps alleviate the crowding problem mentioned before, as in practice it provides more useable space in the projection.

From a theoretical point of view, t-SNE also justifies the use of the t-distribution in the low-dimensional space because it is an infinite mixture of Gaussian distributions. Moreover, its computational cost is lower as it elminates the exponential from the expression.

The conditional neighbouring probabilities in the low-dimensional space using t-distribution are given by the following expression:

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|y_k - y_l\|^2)^{-1}} \tag{3.7}$$

And the gradient of the new cost function, i.e. the Kullback-Leibler divergence between P and the t-distribution based Q is:

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|^2)^{-1} \tag{3.8}$$

In order to improve the efficiency of the gradient descent optimization, a momentum term is added to the gradient. Thus, an exponentially decaying sum of the previous gradients is added to the current gradient, so that improvement along consistent directions is enhanced. The expression that incorporates the gradient update with the momentum, considering the last two gradient matrices, is given by

$$\mathcal{Y}^{(t)} \leftarrow \mathcal{Y}^{(t-1)} + \eta \frac{\delta C}{\delta \mathcal{Y}} + \alpha(t)(\mathcal{Y}^{(t-1)} - \mathcal{Y}^{(t-2)}) \tag{3.9}$$

where $\mathcal{Y}^{(t)}$ is the low-dimensional embedding of the data at iteration t, η is the learning rate of the gradient descent algorithm, and $\alpha(t)$ is a coefficient that regulates the desired amount of momentum at iteration t.

The algorithmic specification of tSNE is given in Algorithm 2. The main source of computational cost in tSNE is found inside the main optimization loop, more specifically in the computation of the Kullback-Leibler divergence between P and Q, that is $\mathcal{O}(n^2)$, where n is the number of points in the dataset. Therefore the overall computational cost of the algorithm is $\mathcal{O}(n^2 T)$.

Algorithm 2 . t-Distributed Stochastic Neighbour Embedding

Input: dataset $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$,
cost function parameters: perplexity $Perp$,
optimization parameters: number of iterations T, learning rate η, momentum $\alpha(t)$
Output: low-dimensional data representation $\mathcal{Y}^{(T)} = \{y_1, y_2, \ldots, y_n\}$
function TSNE($\mathcal{X}, Perp, T, \eta, \alpha(t)$)
    compute pairwise affinities $p_{j|i}$ with perplexity Perp (using Equation 3.1)
    $p_{ij} \leftarrow \frac{p_{j|i} + p_{i|j}}{2n}$
    sample initial solution $\mathcal{Y}^{(0)} = \{y_1, y_2, \ldots y_n\}$ from $\mathcal{N}(0, 10^{-4})$
    for $t \leftarrow 1$ **to** T **do**
        compute low-dimensional affinities q_{ij} (using Equation 3.7)
        compute gradient $\frac{\delta C}{\delta \mathcal{Y}}$ (using Equation 3.8)
        $\mathcal{Y}^{(t)} \leftarrow \mathcal{Y}^{(t-1)} + \eta \frac{\delta C}{\delta \mathcal{Y}} + \alpha(t)(\mathcal{Y}^{(t-1)} - \mathcal{Y}^{(t-2)})$
    end for
end function

3.2.3 Optimization of multiple objectives

The dimensionality reduction algorithm tSNE, described in Section 3.2.2, uses gradient descent optimization to find a reasonably good projection of the points in their original, high-dimensional space into a low-dimensional space. Gradient descent optimization is well defined for single objective problems where the cost function is analitycally differentiable. However, when the problem at hand is multi-objective, i.e. there are several cost functions to minimize that may be contradictory with each other, there is no clear solution for the gradient descent optimization method.

A fundamental criterium in multi-objective optimization is the Pareto optimality [98]. Given a problem defined on a domain $\Upsilon \subseteq \mathbb{R}^N$, with multiple objective functions to be minimized $J_i(Y)|Y \in \Upsilon, 1 \leq i \leq n$, solution Y^1 dominates Y^2 in efficiency, $Y^1 \succ Y^2$, if and only if

$$J_i(Y^1) \leq J_i(Y^2) \forall i = 1, \ldots, n \tag{3.10}$$

and at least one inequality is strict, i.e. all the objective functions on Y^1 are less or equal than on Y^2 and at least one of them is strictly minor. The solution associated with point Y^1 is said to be Pareto efficient with respect to the solution associated with point Y^2.

From the previous definition, a point $Y^o \in \Upsilon$ is said to be Pareto optimal if there is no other point that dominates it. The set of Pareto-optimal points is referred to as the Pareto set. The Pareto front or frontier is the manifold or shape defined by the points in the Pareto set. Most multi-objective gradient descent methods are designed with the objective of finding the Pareto front, that would be the equivalent of the set of minimal points in a single objective problem.

Several multi-objective gradient descent methods have been proposed in the literature, as described in [32], [77], and in [38]. There are alternative solutions also, like the non-dominated sorting genetic algorithm (NSGA)[30] that uses genetic algorithms to optimize a multi-objective problem.

The Multiple Gradient Descent Algorithm (MGDA) [32] is based on the following definitions. Let $Y^0 \in \Upsilon$ be the current point in a multi-objective gradient descent optimization problem. As it is a multi-objective problem, there exists one gradient matrix for each objective function at point Y^0: $\nabla J_i(Y^0)$. Each such gradients implies an optimization direction or vector u_i. Let $\overline{U}$ be the *convex hull* defined by the vectors u_i:

$$\overline{U} = \left\{ u \in \mathbb{R}^N / u = \sum_{i=1}^{n} \alpha_i u_i; \alpha_i \geq 0 (\forall i); \sum_{i=1}^{n} \alpha_i = 1 \right\} \tag{3.11}$$

then there exists a unique vector $\omega \in \overline{U}$ with minimum norm:

$$\forall u \in \overline{U} : (u, \omega) \geq \|\omega\|^2 \tag{3.12}$$

[32] demonstrates that using ω as descent vector on Y^0 guarantees the Pareto efficiency of the new point $Y^1 = Y^0 + \omega$. In this way, MGDA can optimize a multi-objective problem using gradient descent steps that satisfy the Pareto efficiency criterium.

At each iteration of MGDA, given that the current solution point of the algorithm is Y^t, determining ω implies finding the vector characterized by the following expression:

$$min\|\sum_{i=1}^{n}\alpha_i\nabla J_i(Y^t)\|^2 \tag{3.13}$$

where $alpha_i \geq 0 \forall i$ and $\sum_{i_1}^{n}\alpha_i = 1$. In the case of a two objective problem ($n = 2$), there exists an analytical expression that allows to determine ω. However, in the more general case of $n > 2$, determining the ω that satisfies Equation 3.13 implies solving an optimization problem in itself.

3.2.4 Expert opinion pooling

Let us imagine we have a decision problem where a certain entity θ has to be classified into one among k classes. The opinion of an expert can be requested, so that he or she provides a vector of k probability values $p_j, j = 1, \ldots, k$ such that p_j is the probability that θ belongs to class j. Moreover, $\sum_j p_j = 1$. However, it may be the case that the opinion of several experts is requested. Therefore each expert would provide a possibly different vector of probabilities. Then a problem arises: what is the best way of combining the different opinions, i.e. the different vectors of probabilities, in order to make the best use of all the opinions available? This is the subject of the *expert opinion pooling* or *expert opinion aggregation* problem, whose origins are in the areas of economy and market studies.

The approaches to solve this problem are classified in two families according to [20]: axiomatic and Bayesian. Bayesian methods compute a Bayesian estimate from a set of expert opinions and usually yield good results, but require a high number of experts to be considered. This makes them inappropriate for the problems to be treated here.

On the other hand, the axiomatic approaches assume a prior criterium and apply a combination formula that does not take into consideration the actual probability values. One of the most used axiomatic approaches is the *linear opinion pool* ([4]):

$$p(\theta) = \sum_{i=1}^{n}\omega_i p_i(\theta) \tag{3.14}$$

where n is the number of expert opinions considered, $p_i(\theta)$ is the vector of probabilities given by expert i, and ω_i is a weight or confidence value assigned to expert i. In order to simplify the following operations, $\sum_{i=1}^{n} \omega_i = 1$.

Another common approach is the *log-linear opinion pool* [20], defined as:

$$p(\theta) = r \prod_{i=1}^{n} p_i(\theta)^{\omega_i} \tag{3.15}$$

where r is a normalizing constant in order to make $p(\theta)$ a probability distribution, and the remaining values have the same meaning as in the linear opinion pool.

However an open question is the assignment and interpretation of experts' weights (w_i). A straightforward solution proposed by [21] is to assign the same weight to all the experts. [5] propose to choose the weight combination that minimizes the variance of the resulting probability distribution $p(\theta)$. [45] propose to use the Kullback-Leibler divergence (KL) ([103]) to assign weights to the experts, making them inversely proportional to the maximum KL with respect to the other experts' opinions. This way the more dissimilar an expert is from any other expert, the less weight its opinion has. [40] review other expert weight assignment strategies. [15] propose a method to choose the experts' weights designed to be used with the log-linear opinion pooling method. Their method (1) maximizes the entropy of the resulting probability distribution $p(\theta)$ and (2) minimizes the KL divergence between $p(\theta)$ and the individual expert probability distributions $p_i(\theta)$.

Finally, [1] analyzes the criteria for choosing between the linear and the log-linear opinion pooling method. Summarizing the results of his work, the following expression has to be computed:

$$H(p(\theta)) - \sum_{i=1}^{n} \omega_i H(p_i(\theta)) - \log(r) \tag{3.16}$$

where H is Shannon's entropy function. If the value of this expression is positive then the log-linear pooling should be used; otherwise, it is recommended to use the linear pooling. However, as it can be observed, the result of this expression depends on the weights used, which in turn affect the entropy of $p(\theta)$. Combining this criterium with the weight selection method proposed in [15], it is safe to use the log-linear as default pooling method, as Equation 3.16 will be positive in most cases.

3.3 Multiview tSNE

There exist several possible approaches to design an extension to the t-SNE algorithm that can process multiview data. Two approaches are presented in

this thesis. The first solution proposed is to use a multi-objective optimization gradient descent method to find a projection of the original data points that minimizes the divergence between the projection and the different input data views. The second solution uses expert opinion pooling to aggregate the conditional probability distributions of all the input views and therefore transform the multiview problem into a standard tSNE problem applied to the pooled probability matrix. Both approaches are described next.

3.3.1 MV-tSNE as a multiobjective optimization problem

The tSNE dimensionality reduction method finds a reasonably good projection of a set of points in a high-dimensional space to a low-dimensional space by minimizing the Kullback-Leibler divergence (KL) between two conditional probability distributions, as explained in Section 3.2.2. It uses gradient descent optimization in order to find the most convenient arrangement of points in the low-dimensional space. On each iteration of the gradient descent, it computes the gradient of the KL between the conditional probability distribution matrices with respect to the position of the points in the low-dimensional space.

The multiview tSNE extension presented in this Section (MV-tSNE1) differs from the standard tSNE algorithm in the following aspects.

Conditional probability distributions. SNE and tSNE (see Sections 3.2.1 and 3.2.2) convert the Euclidean distances between the points of the input space $\mathcal{X}$ into a matrix of conditional probabilities according to a Gaussian distribution of the distances. In a multiview setting, there exist v input spaces $\{\mathcal{X}^1, \mathcal{X}^2, \ldots, \mathcal{X}^v\}$. MV-tSNE1 computes the conditional probability matrix of each $\mathcal{X}^i$ using Equation 3.1, thus obtaining v probability matrices $P^k, k = 1, \ldots, v$.

However, there is a single conditional probability distribution matrix Q for the low-dimensional space $\mathcal{Y}$, as the goal of MV-tSNE1 is to produce a unique data projection common to all the input views $\mathcal{X}^k$. Q is computed as in tSNE algorithm, using Equation 3.7.

Cost function. The cost function of MV-tSNE1 is the sum of the KL divergences of all the input conditional probability matrices P^k with respect to the low-dimensional conditional probability matrix Q:

$$C = \sum_{k=1}^{v} \sum_{i=1}^{n} KL(P_i^k \| Q_i) = \sum_{k=1}^{v} \sum_{i=1}^{n} \sum_{j=1}^{n} p_{j|i}^k log \frac{p_{j|i}^k}{q_{j|i}} \tag{3.17}$$

however the gradient of the combined objectives is not required, as the multi-objective optimization algorithm used works with the gradients of each objec-

tive (each input view). As a consequence, equation 3.8 applied to each matrix P^k still holds in this algorithm.

Multi-objective gradient descent optimization. In order to minimize the KL divergence of the low-dimensional points with respect to the high-dimensional input views, this multiview dimensionality reduction problem requires a multi-objective gradient descent method, more specifically the Multiobjective Gradient Descent Algorithm (MGDA) described in Section 3.2.3. On each iteration of the optimization algorithm, the gradients of the different input views are computed and combined in a way such that the Pareto efficiency criterium holds. In other words, the change on each iteration never worsens the partial cost value of a specific input view (problem objective).

The use of a momentum vector to improve the performance of the gradient descent algorithm is not defined in MGDA, in fact applying a momentum would often collide with the Pareto-compliant direction of change ω. Therefore that improvement from tSNE is removed in MV-tSNE1. The specification of MV-tSNE1 is presented in Algorithm 3.

Algorithm 3 . Multiview t-Distributed Stochastic Neighbour Embedding 1

Input: v data views of the same n entities $\mathcal{X}^k = \{x_1^k, x_2^k, \ldots, x_n^k\}$, where $k = 1, 2, \ldots, v$,

cost function parameters: perplexity $Perp$,

optimization parameters: number of iterations T, learning rate η

Output: low-dimensional data representation $\mathcal{Y}^{(T)} = \{y_1, y_2, \ldots, y_n\}$

function TSNE($\mathcal{X}^1, \mathcal{X}^2, \ldots, \mathcal{X}^v, Perp, T, \eta$)

compute pairwise affinities $p_{j|i}^k$ with perplexity Perp (using Equation 3.1) for each $\mathcal{X}^k, k = 1, \ldots, v$

$p_{ij}^k \leftarrow \frac{p_{j|i}^k + p_{i|j}^k}{2n}$

sample initial solution $\mathcal{Y}^{(0)} = \{y_1, y_2, \ldots y_n\}$ from $\mathcal{N}(0, 10^{-4})$

for $t \leftarrow 1$ **to** T **do**

compute low-dimensional affinities q_{ij} (using Equation 3.7)

compute gradients $\frac{\delta C^k}{\delta \mathcal{Y}}$ (using Equation 3.8 on the partial cost C^k associated with each p_{ij}^k)

compute vector of change ω as the minimum-norm vector in the convex hull defined by the gradients $\frac{\delta C^k}{\delta \mathcal{Y}}$ (using algorithm MGDA)

$\mathcal{Y}^{(t)} \leftarrow \mathcal{Y}^{(t-1)} + \eta\omega$

end for

end function

Limitations of MV-tSNE1 MV-tSNE1 presents several practical limitations that make it unusable in real datasets. The most relevant limitations regard computational cost, caused by the following reasons:

- On each iteration the KL divergence between each input view and the projection has to be computed. This multiplies by v the computational cost of each iteration, as the KL computation is the most expensive operation on each iteration as discussed in Section 3.2.2.

- The execution of the MGDA algorithm on each iteration also adds an important computational cost per iteration. This is specially the case in problems with more than two objectives (i.e. data views), where it requires executing an optimization algorithm on each iteration to find the common gradient, in order to find the descent vector characterized by Equation 3.13.

- The fact that MGDA does not support the use of momenta in the optimization loop makes it necessary to run the main loop in Algorithm 3 for more iterations (an order of magnitude more as seen in experimental trials).

- The strong Pareto condition makes MGDA halt very often in clearly suboptimal points, as it cannot find a ω vector that satisfies the Pareto efficiency criterium and therefore stalls the optimization process. In other words, the algorithm stops in a local minimum of the multi-objective problem. As a consequence the whole MV-tSNE1 algorithm has to be executed several times in order to hopefully find better solutions.

All these factors combined make MV-tSNE1 extremely expensive in computational terms, what only allows it to be run on "toy" datasets to test its behaviour. Therefore it has been excluded from the main experiments presented in this work.

3.3.2 MV-tSNE as an expert opinion pooling problem

As seen in Sections 3.2.1 and 3.2.2, SNE and tSNE model the input, high-dimensional space as a matrix of conditional distance probabilities according to a Gaussian distribution. The expression for this matrix P is given in Equation 3.1.

In a multiview scenario there are k input high-dimensional spaces instead of only one. A viable strategy for extending tSNE to multiview datasets is as follows. First, to compute a different matrix of conditional distance probabilities P^k for each input space $\mathcal{X}^k, k = 1, \ldots, n$ using Equation 3.1. These k probability matrices can be seen as the probability opinions of k different experts on the distribution of distances of the input samples. Therefore, the

second step of the algorithm is to compute a pooled opinion probability matrix $\hat{P}$ using the log-linear method exposed in Section 3.2.4. From that point, the problem is reduced to finding a low-dimensional space that minimizes the KL divergence between $\hat{P}$ and the conditional probability matrix of the low-dimensional space Q using the approach in tSNE algorithm.

Therefore the cost function of the newly defined optimization problem is:

$$C = \sum_i KL(\hat{P}_i \| Q_i) = \sum_i \sum_j \hat{p}_{j|i} log \frac{\hat{p}_{j|i}}{q_{j|i}} \tag{3.18}$$

And its gradient with respect to the low-dimensional projection space $\mathcal{Y}$ is:

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (\hat{p}_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|^2)^{-1} \tag{3.19}$$

This multiview dimensionality reduction algorithm will be referred as MV-tSNE2, and it is specified in Algorithm 4.

Analysis of the algorithm An important feature of MV-tSNE2 is that it only requires to compute the pooled opinion matrix once, and from that step on its complexity equals that of the single view tSNE algorithm, as all the information from the different input spaces is condensed into a single probability matrix, $\hat{P}$. This makes MV-tSNE2 computationally efficient. Also, MV-tSNE2 is compatible with the use of momentum in the gradient descent optimization stage, leading to better performance relative to the number of iterations.

For the previous reasons, the method used in the experiments is MV-tSNE2 and for simplicity it will be simply referred to as MV-tSNE.

3.4 Results

Although MV-tSNE is a multiview dimensionality reduction method, its application to multiview clustering tasks is straightforward. It simply requires to use the single, low-dimensional projection generated as input to a standard, single view clustering algorithm as K-means. The multiview dimensionality reduction (1) reduces all input views into a single output view, and (2) it summarizes the most relevant information of all input views into a low-dimensional projection. As a consequence applying a clustering algorithm to such projection is equivalent to finding a clustering assignment given the common structure of the multiple input views.

Algorithm 4 . Multiview t-Distributed Stochastic Neighbour Embedding 2

Input: v data views of the same n entities $\mathcal{X}^k = \{x_1^k, x_2^k, \ldots, x_n^k\}$, where $k = 1, 2, \ldots, v$,

cost function parameters: perplexity $Perp$,

optimization parameters: number of iterations T, learning rate η, momentum $\alpha(t)$

Output: low-dimensional data representation $\mathcal{Y}^{(T)} = \{y_1, y_2, \ldots, y_n\}$

function TSNE($\mathcal{X}^1, \mathcal{X}^2, \ldots, \mathcal{X}^v, Perp, T, \eta, \alpha(t)$)

for $k \leftarrow 1$ **to** v **do**

compute pairwise affinities $p^k_{j|i}$ of input matrix $\mathcal{X}^k$ with perplexity Perp using Equation 3.1

$P^k \leftarrow \frac{p^k_{j|i} + p^k_{i|j}}{2n}$

end for

compute the weight ω_k of each affinity matrix P^k using the method described in [15]

$\hat{P} \leftarrow \prod_{k=1}^{v} (P^k)^{\omega_k}$

$\hat{P} \leftarrow \frac{\hat{P}}{sum(\hat{P})}$

sample initial solution $\mathcal{Y}^{(0)} = \{y_1, y_2, \ldots y_n\}$ from $\mathcal{N}(0, 10^{-4})$

for $t \leftarrow 1$ **to** T **do**

compute low-dimensional affinities q_{ij} (using Equation 3.7)

compute gradient $\frac{\delta C}{\delta \mathcal{Y}}$ (using Equation 3.19)

$\mathcal{Y}^{(t)} \leftarrow \mathcal{Y}^{(t-1)} + \eta \frac{\delta C}{\delta \mathcal{Y}} + \alpha(t)(\mathcal{Y}^{(t-1)} - \mathcal{Y}^{(t-2)})$

end for

end function

3.4.1 MV-tSNE with respect to SC baseline

The first block of experiments compare the proposed MV-tSNE method with the t-SNE baseline method, either applied to each input view independently, or applied to all the input views stacked into a single data matrix. The goal of these experiments is to assess the advantages of the multiview method proposed with respect to single view approaches.

In order to synthesize the numerous results (different metrics, datasets and embedding dimensionalities), the results of each evaluation metric are summarized in a set of graph plots, one for each dataset in the experiments. Therefore, six graphs per evaluation metric are produced (five on the unsupervised metrics, as the Cora dataset has a graph space view that is not compatible with these metrics).

However, the detailed numerical results are given in Appendix A, with one table for each combination of evaluation metric and dataset, in a total of 33 tables to evaluate the MV-tSNE method.

3.4.1.1 Dimensionality reduction evaluation

There are three evaluation metrics for the dimensionality reduction task: SVM classification, cophenetic correlation (average on all the input views) and area under the curve of the R_{NX} value (average on all input views).

Figure 3.1 shows the two-dimensional projections of two example datasets using MV-tSNE.

SVM classification (Figure 3.2). The results on the animal with attributes (AWA) dataset are low and irregular in general, given the specific difficulty of the task (there are 50 classes in this dataset). There is an initial peak and a posterior descent around K=10, probably induced by added information that is not related to the classification task and misguides the algorithm. The best single view performs slightly better than the others, although all configurations are quite overlapped.

On the BBC segmented news dataset, both the stacked and the MV-tSNE configurations give better results than the pure single view setups. This dataset has two views that actually are two text segments of the same document, and as a consequence both views are broadly equivalent. This is reflected in the SVM results for BBC, where MV-tSNE and stacked-SC results overlap on most K's, although stacked t-SNE shows higher peaks for some dimensionalities.

On the handwritten digits dataset, there is a clear influence of the best single view, that overlaps almost exactly with both the stacked view and the MV-tSNE. The stability of the results is also remarkable. The Berkeley protein dataset shows very similar results. On the Cora dataset there is also an almost exact overlap of the best single view with MV-tSNE, although the results are not as stable.

Finally on the Reuters dataset the highest score is achieved by the best single view, also at low dimensionalities.

Cophenetic correlation (Figure 3.3). The cophenetic correlation measures the similarity of the distances of the new space with respect to the original input spaces. The average cophenetic correlation on all input views is provided. On the AWA and protein datasets, MV-tSNE matches the best single view, while on the BBC, digits and Reuters datasets MV-tSNE tends to the results of the stacked views. In general MV-tSNE does not produce the best results except on the BBC dataset.

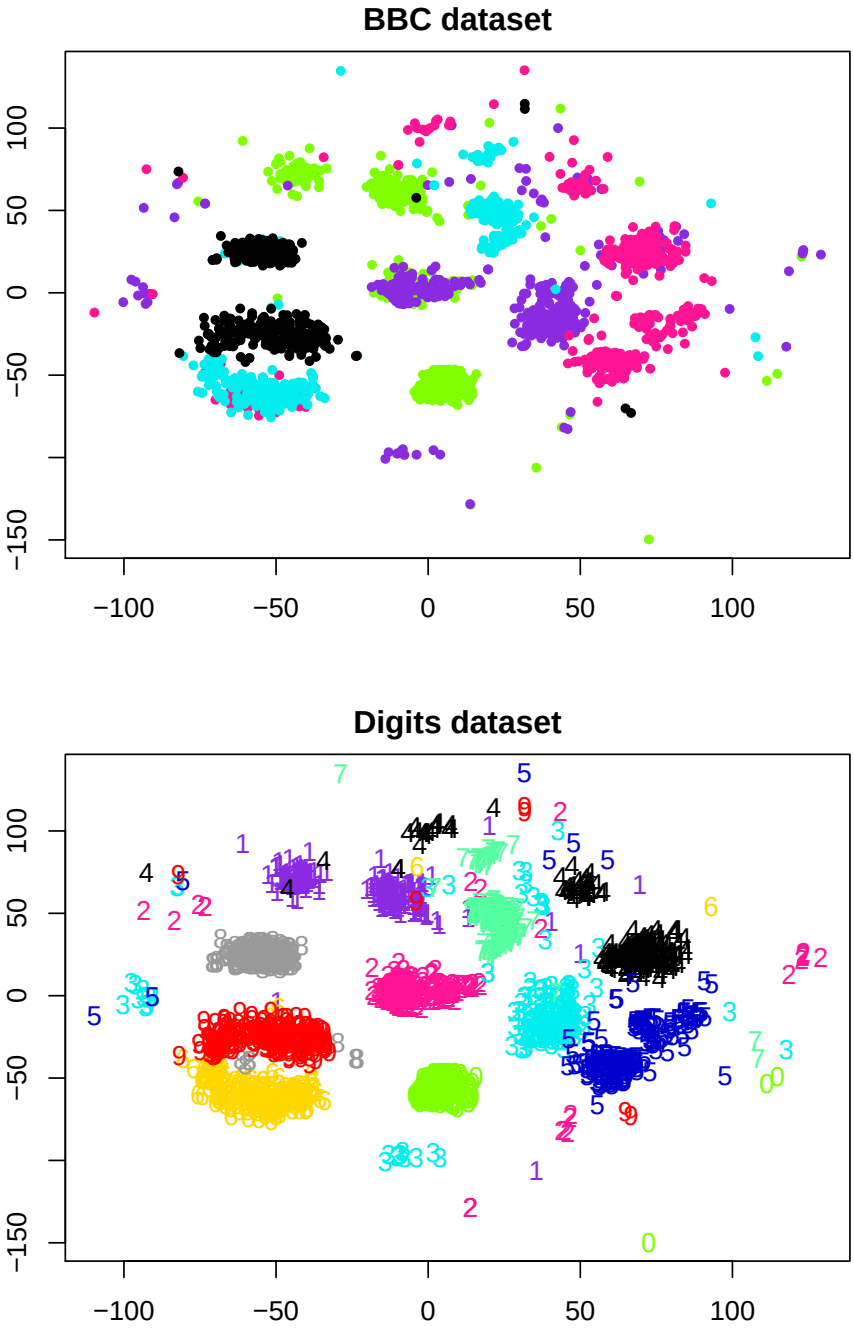


Figure 3.1: MV-tSNE projection of two example datasets.

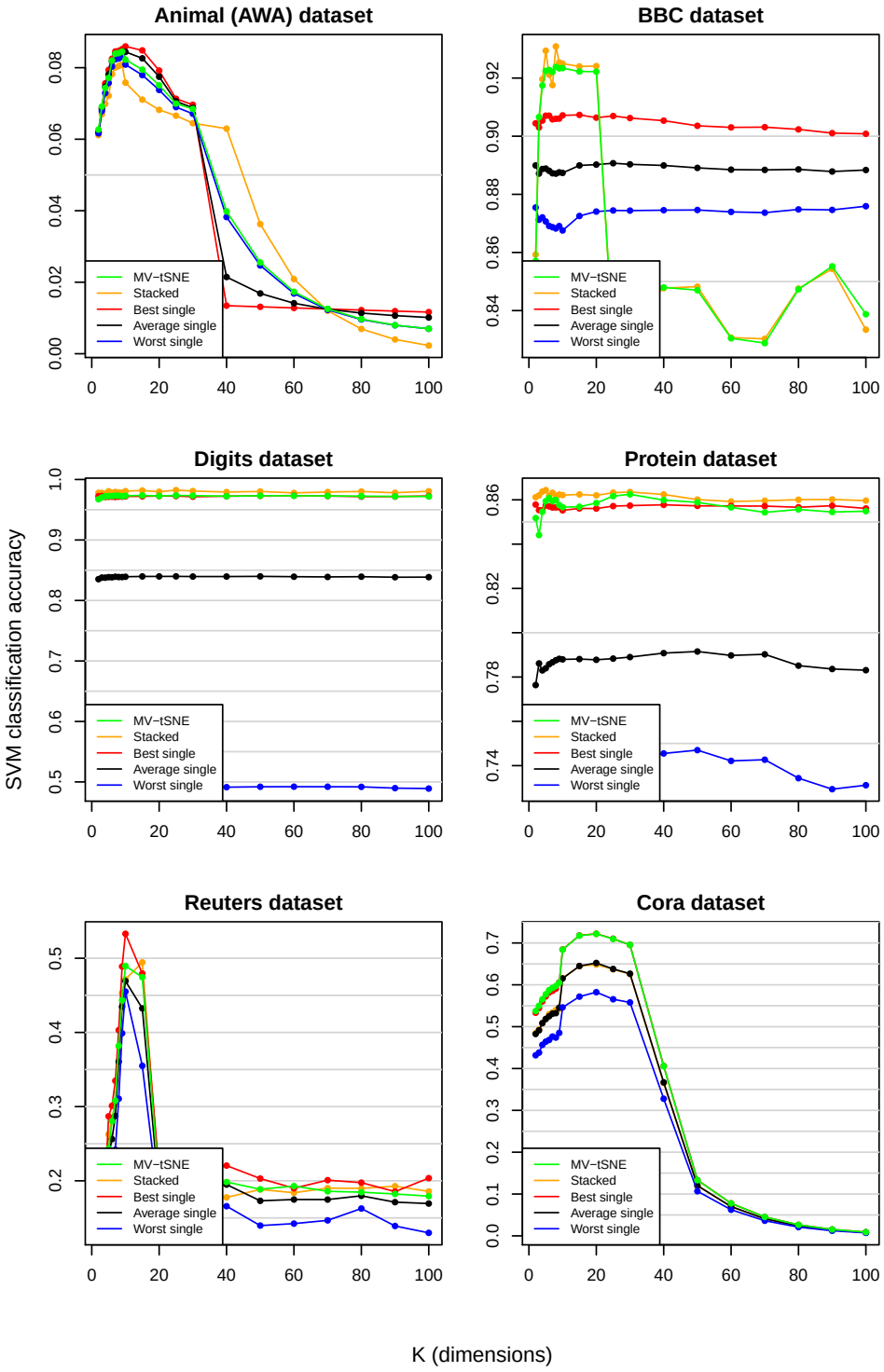


Figure 3.2: MV-tSNE dimensionality reduction evaluation with SVM classification.

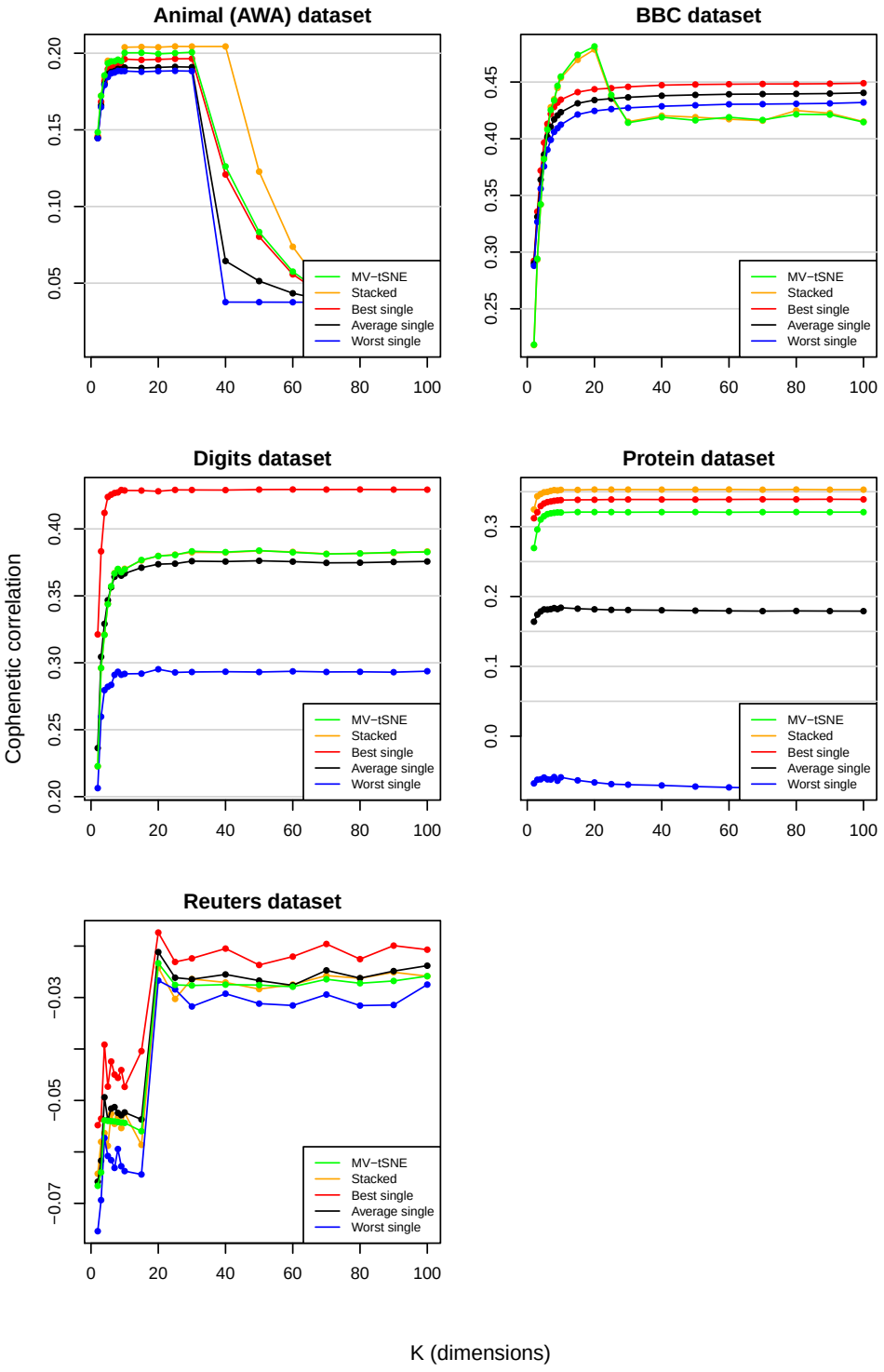


Figure 3.3: MV-tSNE dimensionality reduction evaluation with cophenetic correlation (average on all input views).

Area under the R_{NX} curve (Figure 3.4). The area under the R_{NX} curve (AUC-RNX) also measures the similarity of the sample neighbourhoods in the projected space with respect to the original input view. Here, the average AUC-RNX over all input views is given.

The AUC-RNX results are mostly comparable to the cophenetic correlation results, although here the highest value on the BBC dataset is achieved by the stacked views.

3.4.1.2 Clustering evaluation

There are three evaluation metrics for the clustering task: the clustering purity, the clustering normalized mutual information (NMI), and the Davies-Bouldin index (DBI) (average on all input views).

Clustering purity (Figure 3.5). On the AWA and BBC datasets MV-tSNE initially shows the best results, although in general all results drop with the dimensionality. On the digits dataset it is the stacked views configuration that achieves the maximum values at low K's.

On the protein dataset, the best single view is clearly above the stacked views and MV-tSNE. The difference between single views is noticeable, althogh the vertical scale is small.

On the Reuters dataset, the best single view produces the highest result.

Finally, on the Cora dataset MV-tSNE overlaps with the best single view with the best results.

Clustering normalized mutual information (Figure 3.6). The clustering NMI results are quite similar to the clustering purity results, with MV-tSNE showing the best results on animal and BBC datasets. Again, the stacked view performs better on the digits dataset, and also on the protein dataset in this case.

On the Reuters dataset the best single view achieves the highest NMI. On the Cora dataset, MV-tSNE matches the best single view on the top positions.

Davies-Bouldin index (Figure 3.7). The Davies-Bouldin index (DBI) measures the internal properties of the clusters with respect to other clusters, considering the distances in the input spaces. Here, the average Davies-Bouldin index over all input views is given. For DBI, less is better.

On the animal, digits, protein and Reuters datasets, the best single view is the best configuration. On the BBC dataset both MV-tSNE and stacked views produce the best results, with a remarkable difference from the single views configurations.

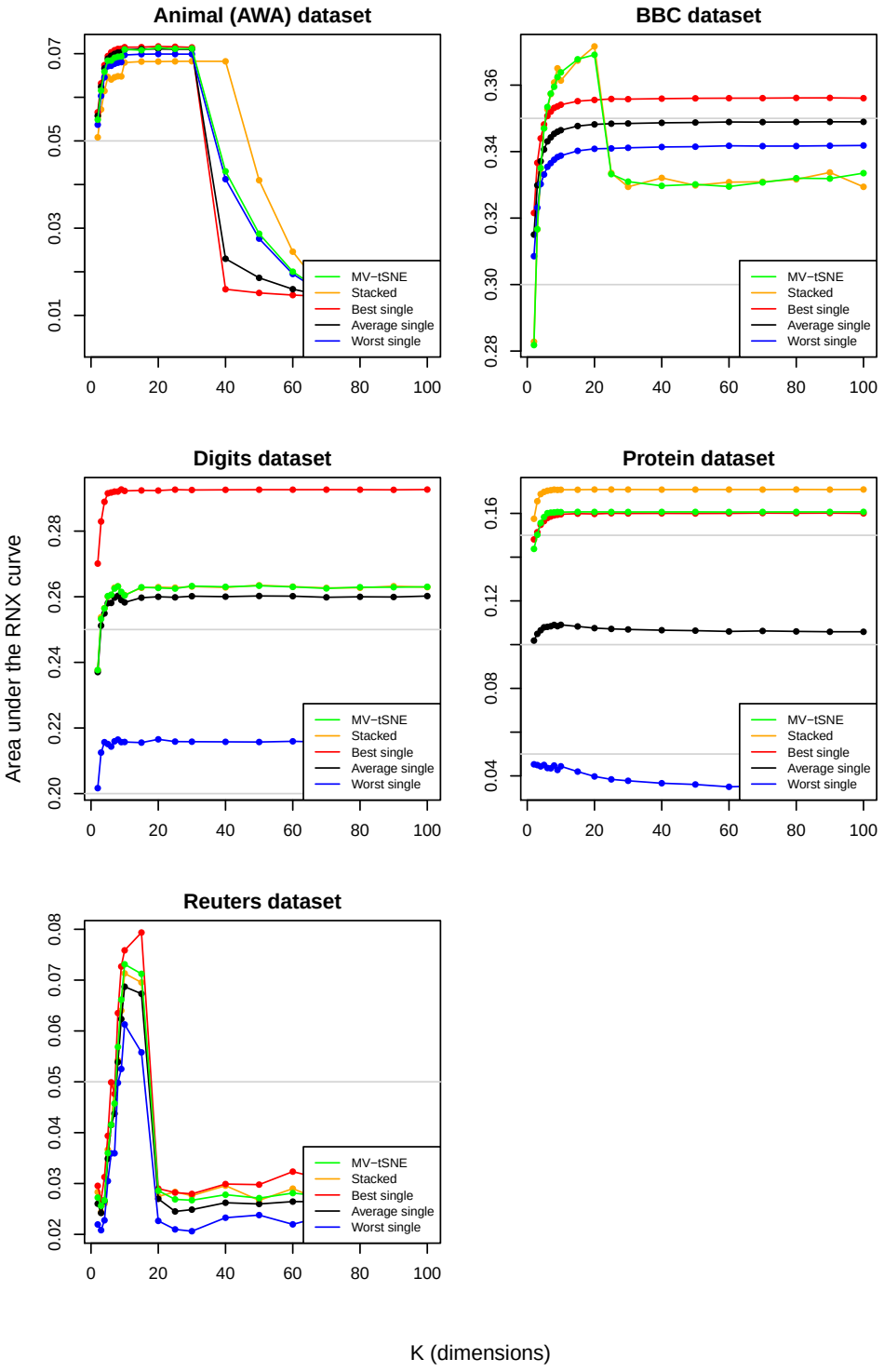


Figure 3.4: MV-tSNE dimensionality reduction evaluation with area under the R_{NX} curve (average on all input views).

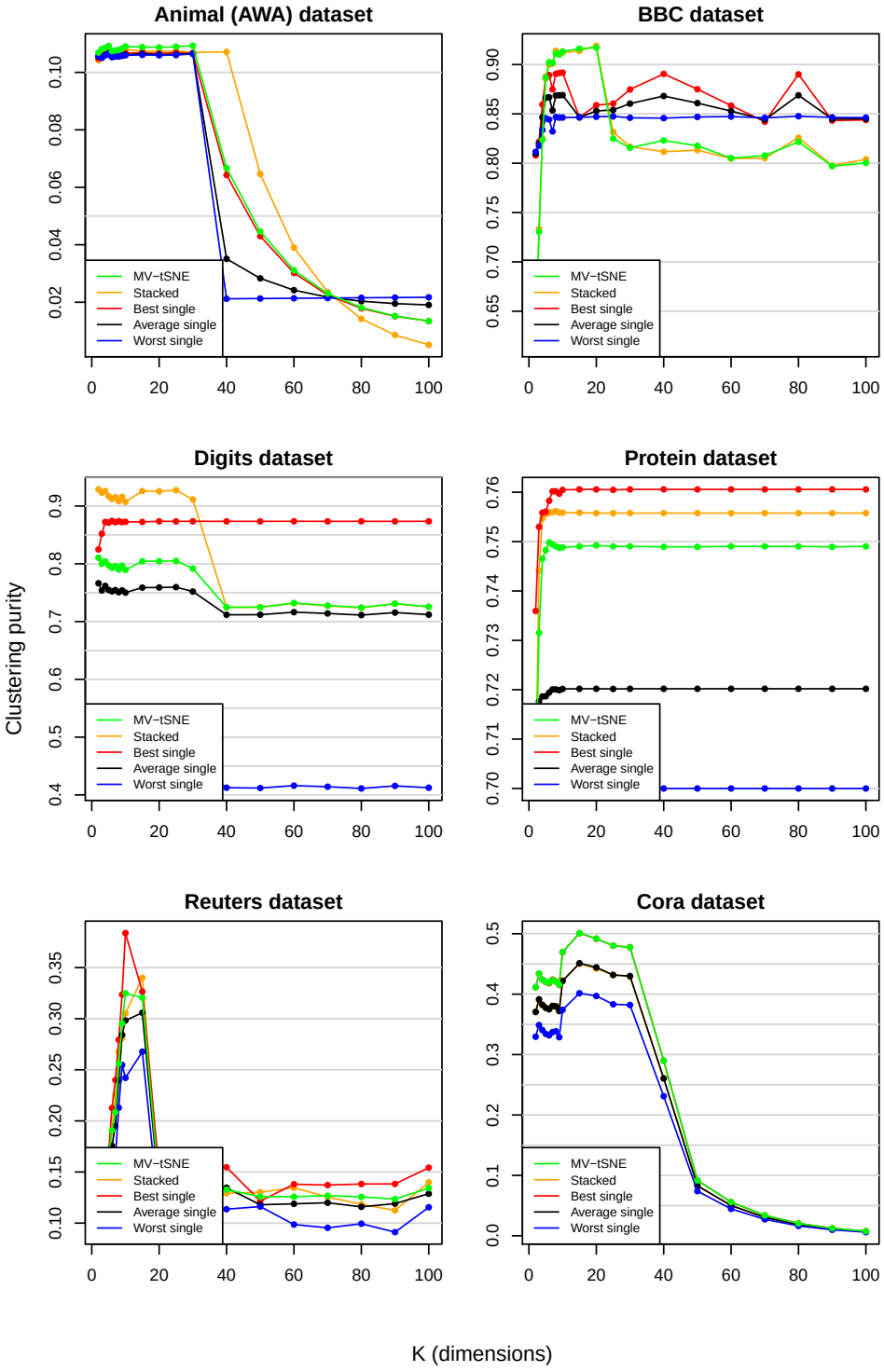


Figure 3.5: MV-tSNE clustering evaluation with clustering purity.

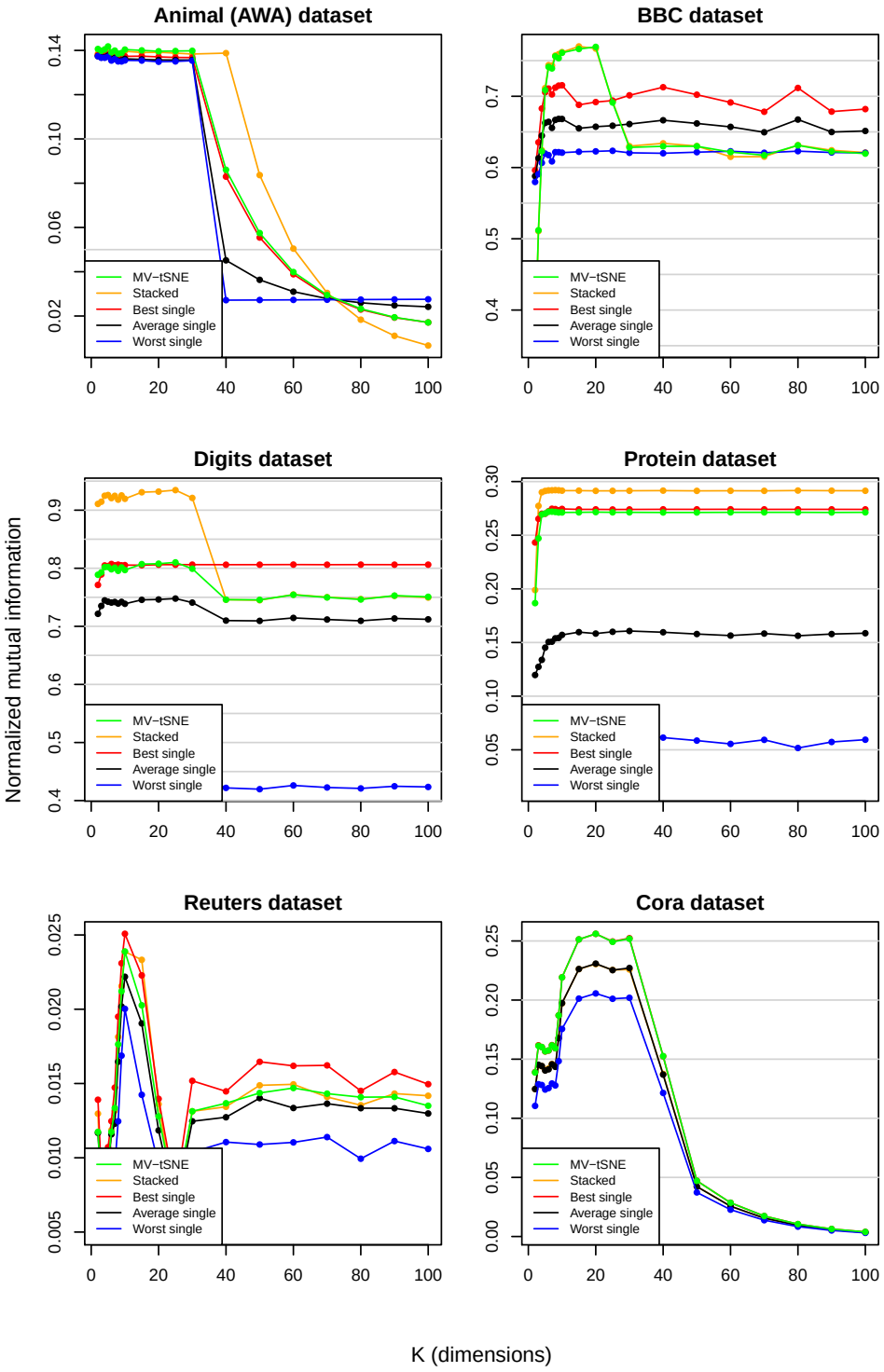


Figure 3.6: MV-tSNE clustering evaluation with clustering normalized mutual information.

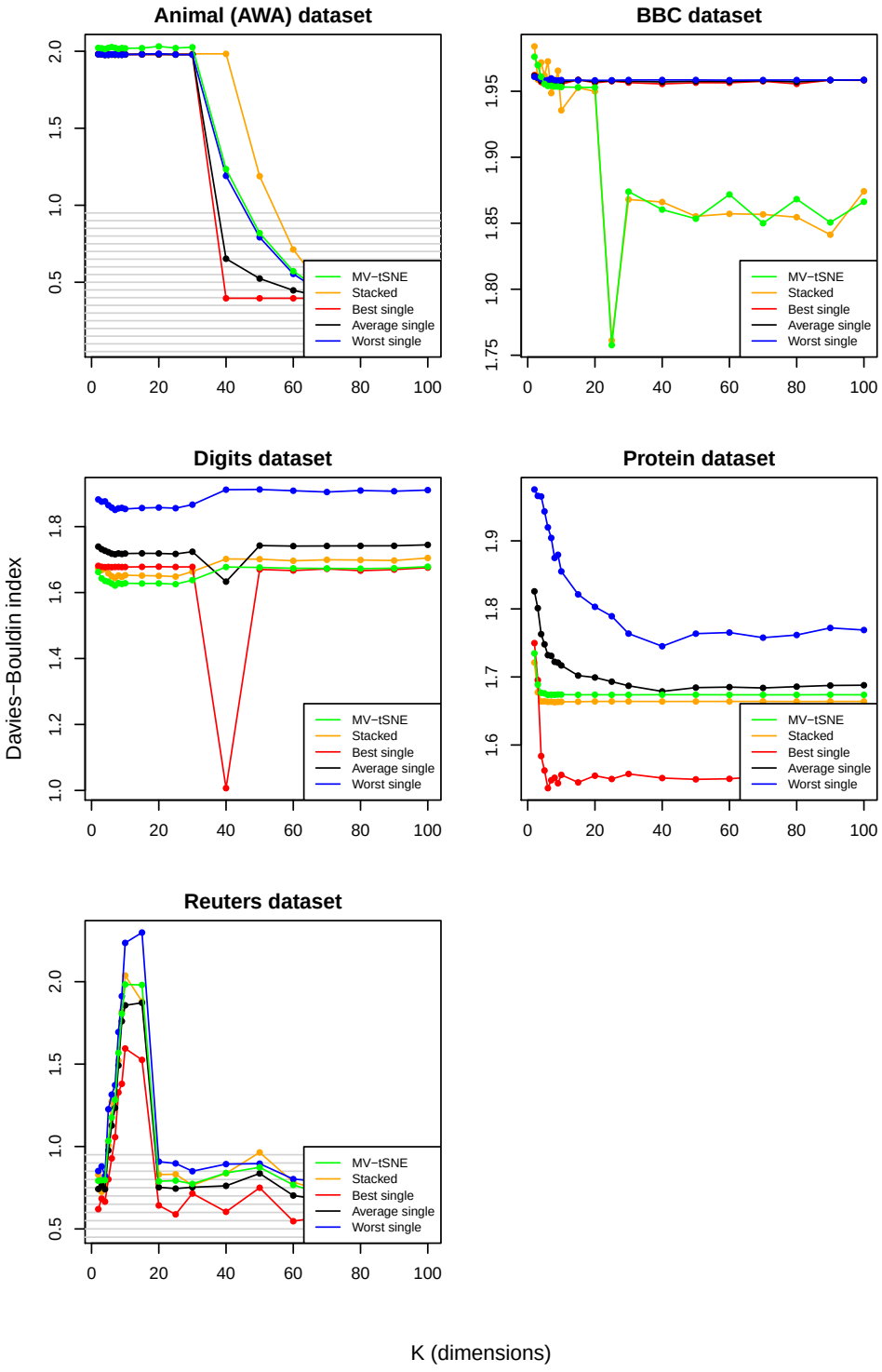


Figure 3.7: MV-tSNE clustering evaluation with the Davies-Bouldin index (average on all input views). Less is better.

3.4.2 MV-tSNE with respect to the state of the art

Table 3.1 shows the most relevant clustering purity results in the state of the art and compares them to the results of MV-tSNE on the same datasets and configuration. The clustering purity measures the faithfulness of the produced clustering assignments to the reference class assignments in the dataset. In other words, higher values mean that the clustering assignment is more similar to the class assignment, with a value of 1 meaning that both are identical.

There is no single method that clearly performs better than the others according to the clustering purity on these datasets. Although in most cases MV-tSNE is relatively close to the highest scores, it is never ranked in the first position.

Method	Digits	BBC	Reuters	AWA
CoregSC	0.822	0.887	0.552	0.580
MMSC	0.758	NA	0.390	0.585
MVC-SS	NA	NA	0.531	**0.629**
MV-KMeans	0.825	NA	NA	0.114
CoKmLDA	0.819	**0.914**	NA	NA
MFSC-MO	0.800	NA	NA	NA
MVC-PSS	**0.862**	NA	NA	0.325
MVSC-BG	0.844	NA	**0.577**	NA
MV-tSNE	0.789	0.902	0.295	0.607

"NA" means there are no available results of the method on the data set.

Table 3.1: Clustering purity wrt. the state of the art.

Table 3.2 shows the normalized mutual information (NMI) between the clustering assignments produced by the multiview clustering methods in the state of the art and the reference class labels in the dataset. A higher value implies a more similar assignment, with NMI=1 meaning a perfect match. However NMI measures the coincidence of the cluster assignments differently from purity, as it uses the mutual information between both assignments.

As in the case of the clustering purity, there is no method clearly superior to the others. Again, MV-tSNE scores near the best results in some datasets (digits and BBC), although its score is lower on the others and does not produce the best NMI on any dataset.

3.5 Discussion

When compared to the baseline methods on both the dimensionality reduction and the clustering tasks, MV-tSNE shows an average performance. In many cases its results overlap with the results of the stacked views configurations. In

Method	Digits	BBC	Reuters	AWA
CoregSC	**0.836**	0.769	0.326	0.695
MMSC	0.792	NA	0.134	0.698
MVC-SS	NA	NA	NA	**0.751**
MV-KMeans	0.807	NA	NA	0.117
CoKmLDA	0.818	**0.796**	NA	NA
MFSC-MO	0.785	NA	NA	NA
MVC-PSS	0.833	NA	NA	0.213
MVSC-BG	0.832	NA	**0.357**	NA
MV-tSNE	0.797	0.756	0.024	0.621

"NA" means there are no available results of the method on the data set.

Table 3.2: Clustering NMI wrt. the state of the art.

some cases, it overlaps with the best single view. Otherwise, its performance tends to be around the average single view, and rarely dominates the results on any of the experiments.

Compared with other multiview clustering methods in the state of the art, the method described in this chapter (MV-tSNE) shows an average performance, although it never ranks in the first positions. According to these results, MV-tSNE is not an improvement over other methods in the state of the art.

Chapter 4

Multiview multidimensional scaling

4.1 Motivation

The multidimensional scaling algorithm (MDS) [23] is one of the standard tools for dimensionality reduction and data visualization, where a high-dimensional data matrix is transformed into a low-dimensional projection matrix, usually in two or three dimensions in order to produce a graphical display of the data.

Reducing the dimensionality of multiview datasets poses new challenges, as the task becomes twofold: (1) combining the information in the different data views in an appropriate way, and (2) reducing the dimensionality of the common information found. In turn, efficient multiview dimensionality reduction methods may prove useful tools that let their users synthesize extremely complex data into a single representation while keeping the essential properties of the data. For instance, being able to visualize in a 2D plot the structure of data as complex as some of the multiview datasets described in Chapter 1.

For this reason, the development of a multiview equivalent of the MDS algorithm has been deemed as a potential and useful addition to the arsenal of dimensionality reduction methods. Multiview MDS (MV-MDS) produces a single, low-dimensional representation of the multiview, high-dimensional input data while keeping the essential information. This allows to use the low-dimensional matrix as input to a standard clustering method, thus creating a potentially useful multiview clustering method.

This Chapter explains some prior works related with the method, then describes MV-MDS itself, and finally presents and analyzes the results of several experiments on different multiview datasets.

4.2 Related work

4.2.1 Multidimensional Scaling

Multidimensional Scaling (MDS), proposed in [23], is one of the dimensionality reduction methods with more widespread use both for data dimensionality reduction and for data visualization. Although there are several revisions of MDS, such as Metric MDS ([60]) and Non-metric MDS ([61]), the original version of MDS is presented here, often referred to as Classical MDS.

Given an high-dimensional input space $\mathcal{X} \subseteq \mathbb{R}^p \times \mathbb{R}^n$, the main idea behind MDS is to obtain a low-dimensional representation of the p points in $\mathcal{X}$, $\mathcal{Y} \subseteq \mathbb{R}^p \times \mathbb{R}^m$, with $m < n$, so that each point $x_i \in \mathcal{X}$ has a representation in a new point $y_i \in \mathcal{Y}$ and the distance between points is preserved as much as possible. The correspondence of the distances between points in $\mathcal{X}$ and in $\mathcal{Y}$ is named **stress**, and it is defined as follows:

$$Stress(x_1, x_2, \ldots, x_p) = \sqrt{\frac{\sum_{ij}(\hat{d}_{i,j} - d_{i,j})^2}{\sum_{ij} \hat{d}_{i,j}^2}} \tag{4.1}$$

$\forall i, j = 1, 2, \ldots, p$, where d_{ij} is the Euclidean distance between points x_i and x_j, i.e. the original or input points, and $\hat{d}_{ij}$ is the Euclidean distance between points y_i and y_j, i.e. the projection of points x_i and x_j in the low-dimensional space $\mathcal{Y}$.

The dimensionality reduction problem can be viewed as an optimization problem where the stress function stated above is the *cost* of a given solution. The problem with this specific cost function can be solved by finding the eigenvectors of matrix $B = \mathcal{X}\mathcal{X}^T$, as described in Algorithm 5.

4.2.2 Stepwise common principal components

Common principal components (CPC) analysis, first proposed by [37], is a statistical method of simultaneously diagonalizing a set of positive-definite symmetric matrices. This method, also known as *joint diagonalization*, attempts to diagonalize the input matrices under the hypothesis of common components H, which states that there exists an orthogonal matrix W such that the C input matrices have the same diagonal form, as formulated in Equation 4.2.

$$H : L'_k = W^T \; L_k \; W, \; c = 1, 2, ..., C \tag{4.2}$$

where L_c is the positive-definite symmetric matrix of input matrix c, and L'_c is its diagonalized form, obtained from the linear transformation defined in matrix W. Note that the resulting eigenvectors (columns of W) are common

Algorithm 5 . Classical Multidimensional Scaling

Input: dataset $\mathcal{X} = \{x_1, x_2, \ldots, x_p\}$, with $x_i \in \mathbb{R}^n$,
desired number of dimensions in the output space: m

Output: low-dimensional data representation $\mathcal{Y} = \{y_1, y_2, \ldots, y_p\}$, with $y_i \in \mathbb{R}^m$.

function MDS($\mathcal{X}, m$)

Compute the squared distance matrix: $D^{(2)} \leftarrow [d^2(x_i, x_j)]$, with $i, j = 1, 2, \ldots, p$, where d is the Euclidean distance function

Compute the centering matrix: $J \leftarrow I - \frac{1}{p}\mathbb{O}$, where I is the identity matrix and $\mathbb{O}$ is a $p \times p$ matrix of ones

Apply double centering to the squared distance matrix: $B \leftarrow -\frac{1}{2}JD^{(2)}J$

Compute the m largest eigenvalues of B and their associated eigenvectors $e_1, e_2, \ldots, e_m$.

$\mathcal{Y} \leftarrow \{e_1, e_2, \ldots, e_m\}^T$, with $\mathcal{Y} \in \mathbb{R}^p \times \mathbb{R}^m$

end function

for all the input matrices, while the eigenvalues are specific to each input matrix. In other words, the input matrices are projected into the same subspace defined by the eigenvectors, with the relative weight of each subspace axis for each input matrix given by the associate eigenvalue.

The stepwise common principal components algorithm (referred here as S-CPC), described in [100], finds an approximate solution to this problem in an incremental manner. More specifically, it first computes the common components (common eigenvectors) with highest eigenvalues. Therefore it can stop after computing a given number of common principal components. This has a dramatic impact on the performance of the method presented in this paper, as explained in Section 4.3. S-CPC can be applied to any number C of input matrices.

4.3 MV-MDS description

The method presented here, called Multiview Multidimensional Scaling (MV-MDS), is an extension of classical MDS designed to work with multiview datasets. Therefore, instead of a single input space $\mathcal{X}$, the p input examples are sampled from v different input spaces, thus having as net input to the method a set of data matrices $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_v$. However, MV-MDS is expected to produce as its output a single projection space $\mathcal{Y}$. The goal of MV-MDS is to find such space $\mathcal{Y}$ so that the distances between points in it are as similar as possible to the distances between points in *all* v input spaces. Obviously this may not always be possible, for the simple reason that the distances in one

input space $\mathcal{X}_i$ may be in contradiction with the distances in another input space $\mathcal{X}_j$. However, MV-MDS is designed to try and produce the projection space that more faithfully preserves the distance relations common to all the input spaces.

Note that the input spaces may have different dimensionality.

Given the fact that there exist v input spaces, the stress or cost function for MV-MDS, adapted from the stress definition in Equation 4.1, becomes the stress of the low-dimensional space $\mathcal{Y}$ with respect to all the input, high-dimensional spaces $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_v$, and it is characterized by the following equation:

$$Stress_{MV}(x_1^{(k)}, x_2^{(k)}, \ldots, x_p^{(k)}) = \frac{1}{v} \sum_{k=1}^{v} \sqrt{\frac{\sum_{ij} (\hat{d}_{i,j} - d_{i,j}^{(k)})^2}{\sum_{ij} \hat{d}_{i,j}^2}} \tag{4.3}$$

$\forall i, j = 1, 2, \ldots, p$, $\forall k = 1, 2, \ldots, v$, where $x_i^{(k)}$ is the vector with the coordinates of point i in input space k, $d_{ij}^{(k)}$ is the Euclidean distance between points $x_i^{(k)}$ and $x_j^{(k)}$, i.e. the distance in input space k, and $\hat{d}_{ij}$ is the Euclidean distance between points y_i and y_j, i.e. the projection of points x_i and x_j in the low-dimensional space $\mathcal{Y}$.

Given that there are v input matrices $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_v$, now the dimensionality reduction problem characterized by Equation 4.3 requires finding the eigenvectors of the matrices $B^{(k)} = \mathcal{X}^{(k)} \mathcal{X}^{(k)T}, k = 1, 2, \ldots, v$. More specifically, if the desired projection space $\mathcal{Y}$ should be m-dimensional, then the m largest eigenvalues and their corresponding eigenvectors need to be determined.

Solving this problem directly would produce a set of m eigenvectors for each input matrix, invalidating the whole procedure. However, if the stepwise common principal components method (S-CPC, see Section 4.2.2) is applied on $B^{(k)} = \mathcal{X}^{(k)} \mathcal{X}^{(k)T}, k = 1, 2, \ldots, v$ then the m largest *common* eigenvalues and their eigenvectors can be computed.

For each input matrix, S-CPC computes an associated eigenvalue. However, the eigenvectors are unique, common to all input matrices. S-CPC guarantees that the eigenvectors produced are ordered by highest sum of the corresponding eigenvalues (i.e. the eigenvalues associated to all the input matrices). This satisfies the condition of MDS, where the highest eigenvalues must be computed and their eigenvectors returned as the resulting projection space.

The resulting MV-MDS algorithm is presented in Algorithm 6.

Algorithm 6 . Multiview Multidimensional Scaling

Input: v datasets $\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \ldots, \mathcal{X}^{(v)}, with \mathcal{X}^{(k)} \subseteq \mathbb{R}^p \times \mathbb{R}^{n_k}$,
desired number of dimensions in the output space: m
Output: low-dimensional data representation $\mathcal{Y} \subseteq \mathbb{R}^p \times \mathbb{R}^m$.
function MDS($\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \ldots, \mathcal{X}^{(v)}, m$)
Compute the centering matrix: $J \leftarrow I - \frac{1}{p}\mathbb{O}$, where I is the identity matrix and $\mathbb{O}$ is a $p \times p$ matrix of ones
for $k \leftarrow 1$ **to** v **do**
Compute the squared distance matrix: $D^{(k)(2)} \leftarrow [d^2(x_i^{(k)}, x_j^{(k)})]$, with $i, j = 1, 2, \ldots, p$, where d is the Euclidean distance function
Apply double centering to the squared distance matrix: $B^{(k)} \leftarrow -\frac{1}{2} J D^{(k)(2)} J$
end for
Compute the m largest common eigenvalues of $\{B^{(1)}, B^{(2)}, \ldots, ^{(v)}\}$ and their associated eigenvectors $e_1, e_2, \ldots, e_m$ using S-CPC.
$\mathcal{Y} \leftarrow \{e_1, e_2, \ldots, e_m\}^T$, with $\mathcal{Y} \in \mathbb{R}^n \times \mathbb{R}^m$
end function

4.4 Results

4.4.1 MV-MDS with respect to SC baseline

The first block of experiments compare the proposed MV-MDS method with the single-view MDS baseline method, either applied to each input view independently, or applied to all the input views stacked into a single data matrix. The goal of these experiments is to assess the advantages of the multiview method proposed with respect to the original, single view approach.

In order to synthesize the numerous results (different metrics, datasets and embedding dimensionalities), the results of each evaluation metric are summarized in a set of graph plots, one for each dataset in the experiments. Therefore, six graphs per evaluation metric are produced (five on the unsupervised metrics, as the Cora dataset has a graph space view that is not compatible with these metrics). The dimensionality of the results (x axis on the plots) will be referred as K for simplicity.

Due to their considerable extension, the detailed numerical results are given in Appendix B, with one table for each combination of evaluation metric and dataset, in a total of 33 tables to evaluate the MV-MDS method.

4.4.1.1 Dimensionality reduction evaluation

There are three evaluation metrics for the dimensionality reduction task: SVM classification, cophenetic correlation (average on all the input views) and area under the curve of the R_{NX} value (average on all input views).

Figure 4.1 shows the two dimensional projections of two example datasets using MV-MDS.

SVM classification (Figure 4.2). The results on the animal with attributes (AWA) dataset are low in general, given the specific difficulty of the task (there are 50 classes in this dataset). There is an initial peak and a posterior descent around K=10, probably induced by added information that is not related to the classification task that misguides the algorithms. There are considerable differences among single views. MV-MDS produces the absolute maximum on K=40, and is consistently better than the stacked views configuration.

The BBC segmented news dataset has two views that actually are two text segments of the same document, and as a consequence both views are mainly equivalent. This is reflected in the SVM results for BBC, where MV-MDS and stacked-MDS results overlap as there is no practical difference between both approaches in this dataset. MV-MDS (and stacked-MDS) perform consistently better than single-view MDS, showing the usefulness of adding more text content to the method.

On the handwritten digits dataset, MV-MDS performs slightly better than stacked-MDS on most dimensionalities. In turn, they perform slightly better than the best single view, and clearly better than the average of single views.

The performance of the different configurations with the Berkeley protein dataset shows high variation along the values of K, with some single view configurations dropping while others improving with K. MV-MDS shows the highest SVM accuracy, and is consistently higher than stacked view for all K's.

On the Reuters dataset there is a single view that is clearly better than the other configurations. This means that one of the input languages (Reuters has news documents on five languages) is more suited for document classification according to the topic given in the dataset. MV-MDS has a better than average performance.

Finally, on the Cora dataset MV-MDS is clearly superior to the baseline options, achieving the best results with a relatively low dimensionality (around 25). Probably the information added with further dimensions is noise from the classification point of view. It is interesting to observe how, in this case, the stacked-MDS is attracted by the worst single view, while the MV-MDS seems more robust and generates better results.

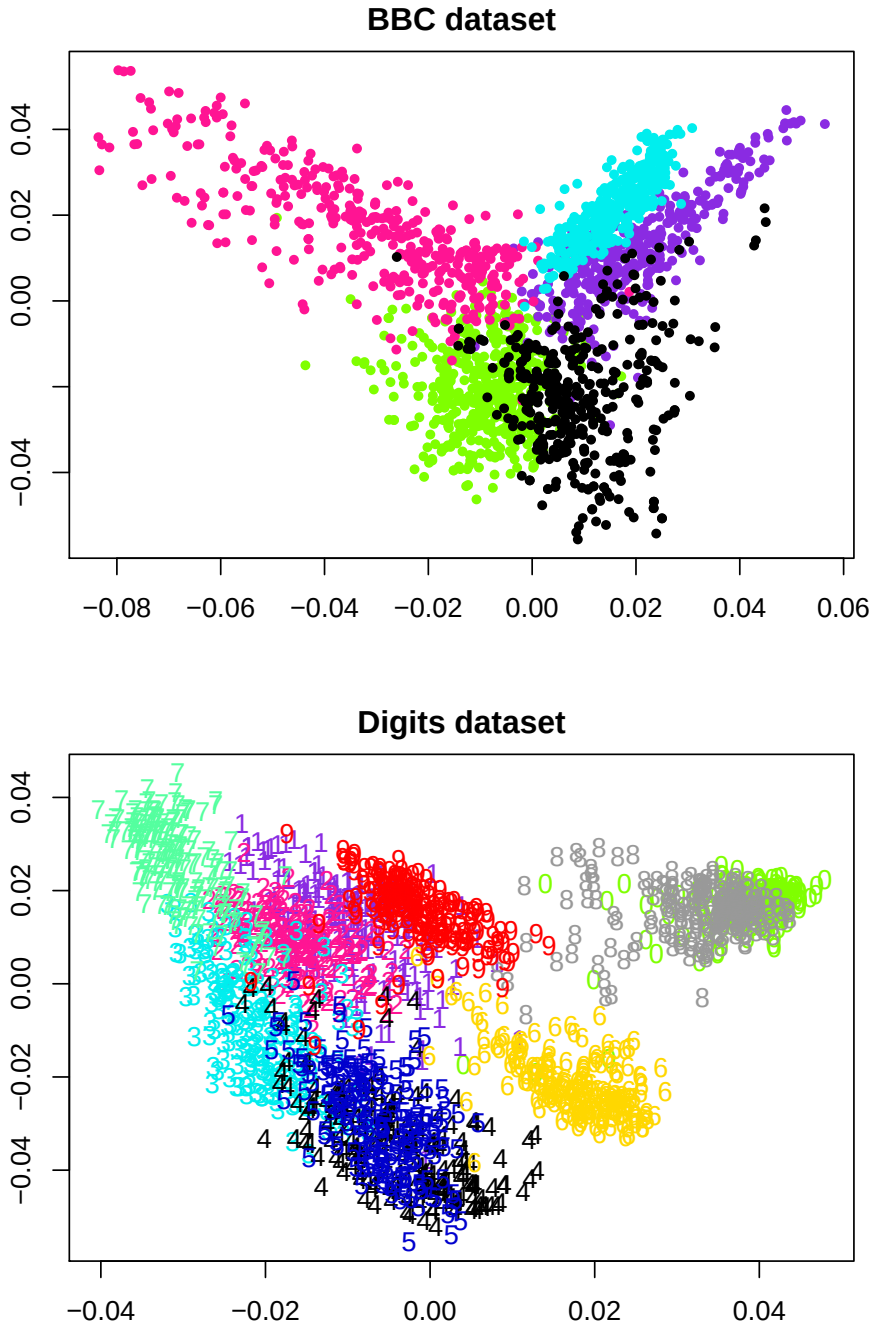


Figure 4.1: MV-MDS projection of two example datasets.

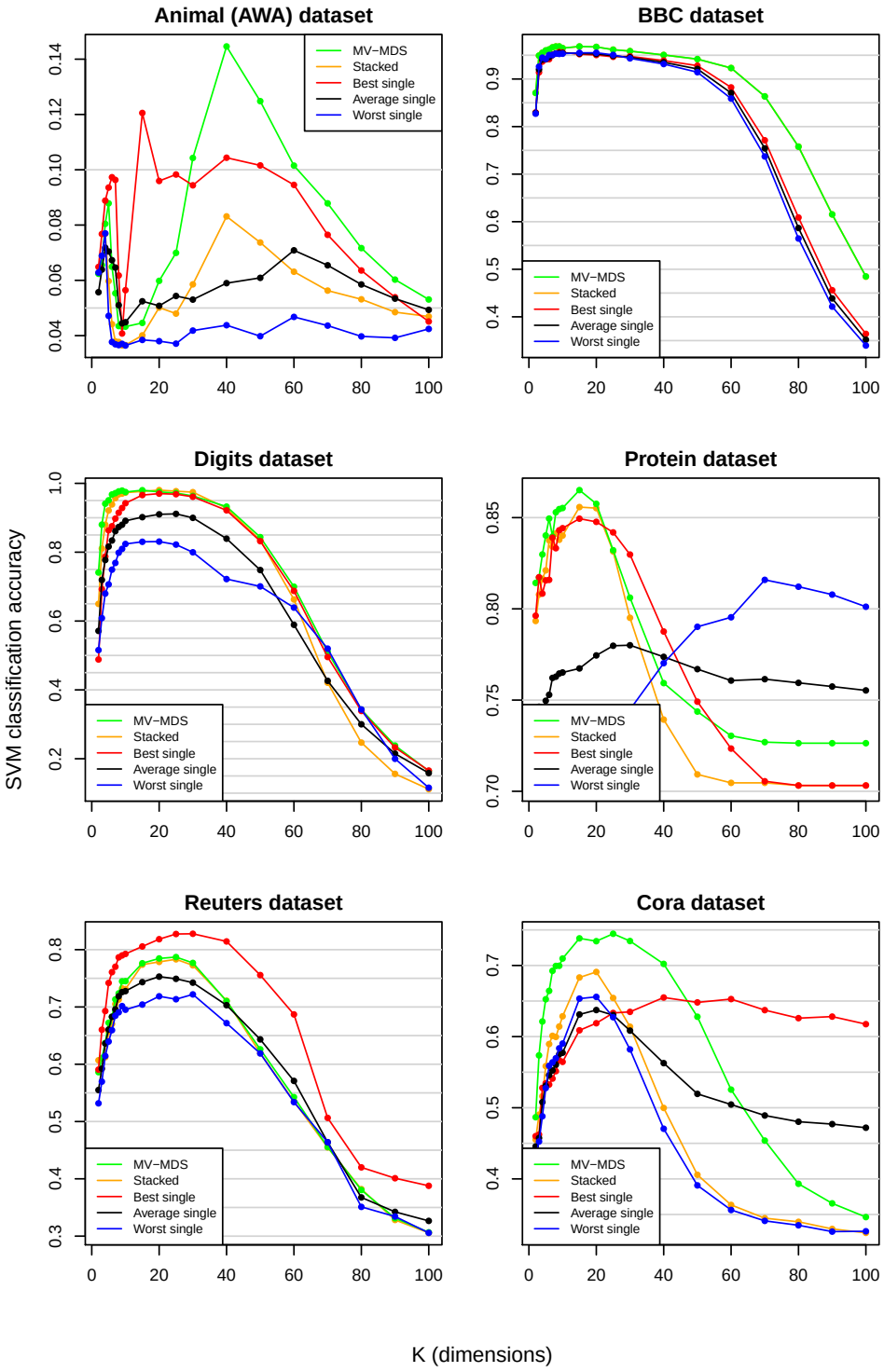


Figure 4.2: MV-MDS dimensionality reduction evaluation with SVM classification.

Cophenetic correlation (Figure 4.3). The cophenetic correlation measures the similarity of the distances of the new space with respect to the original input spaces. The average cophenetic correlation on all input views is provided. On the AWA dataset, MV-MDS shows clearly higher values. Note the considerable difference between the worst and best single views.

On the BBC dataset there is an initial peak and a posterior descent. Although MV-MDS performs well with low dimensions, its cophenetic correlation drops with more dimensions. Apparently it is producing a different distance structure than those present in the original data views. A similar behaviour appears with the digits dataset.

On the digits, protein and Reuters datasets, MV-MDS shows a lower cophenetic correlation than the other configurations, generally starting with average results, then dropping to lowest values.

Area under the R_{NX} curve (Figure 4.4). The area under the R_{NX} curve (AUC-RNX) also measures the similarity of the sample neighbourhoods in the projected space with respect to the original input view. Here, the average AUC-RNX over all input views is given.

On the animal and BBC datasets MV-MDS shows AUC-RNX values above the average. However, in digits, protein and Reuters MV-MDS starts on average values but then drops to lowest values.

4.4.1.2 Clustering evaluation

There are three evaluation metrics for the clustering task: the clustering purity, the clustering normalized mutual information, and the Davies-Bouldin index (average on all input views).

Clustering purity (Figure 4.5). On the AWA dataset the results are low in general, due to the high number of classes (50). However, MV-MDS performs better that the baseline MDS configurations. Both in AWA and digits, MV-MDS performance is intertwined with best single and stacked configurations.

On the BBC, protein and Cora datasets MV-MDS shows an irregular performance, although in all these cases it achieves the best absolute results.

On the Reuters dataset, both MV-MDS and stacked perform worse than single views in general.

Clustering normalized mutual information (Figure 4.6). On the digits dataset, the clustering normalized mutual information (NMI) of MV-MDS is clearly superior to the other methods. On the AWA, BBC and protein datasets it also achieves the best overall performance at specific K's.

On the Reuters and Cora datasets, MV-MDS performs around or below average.

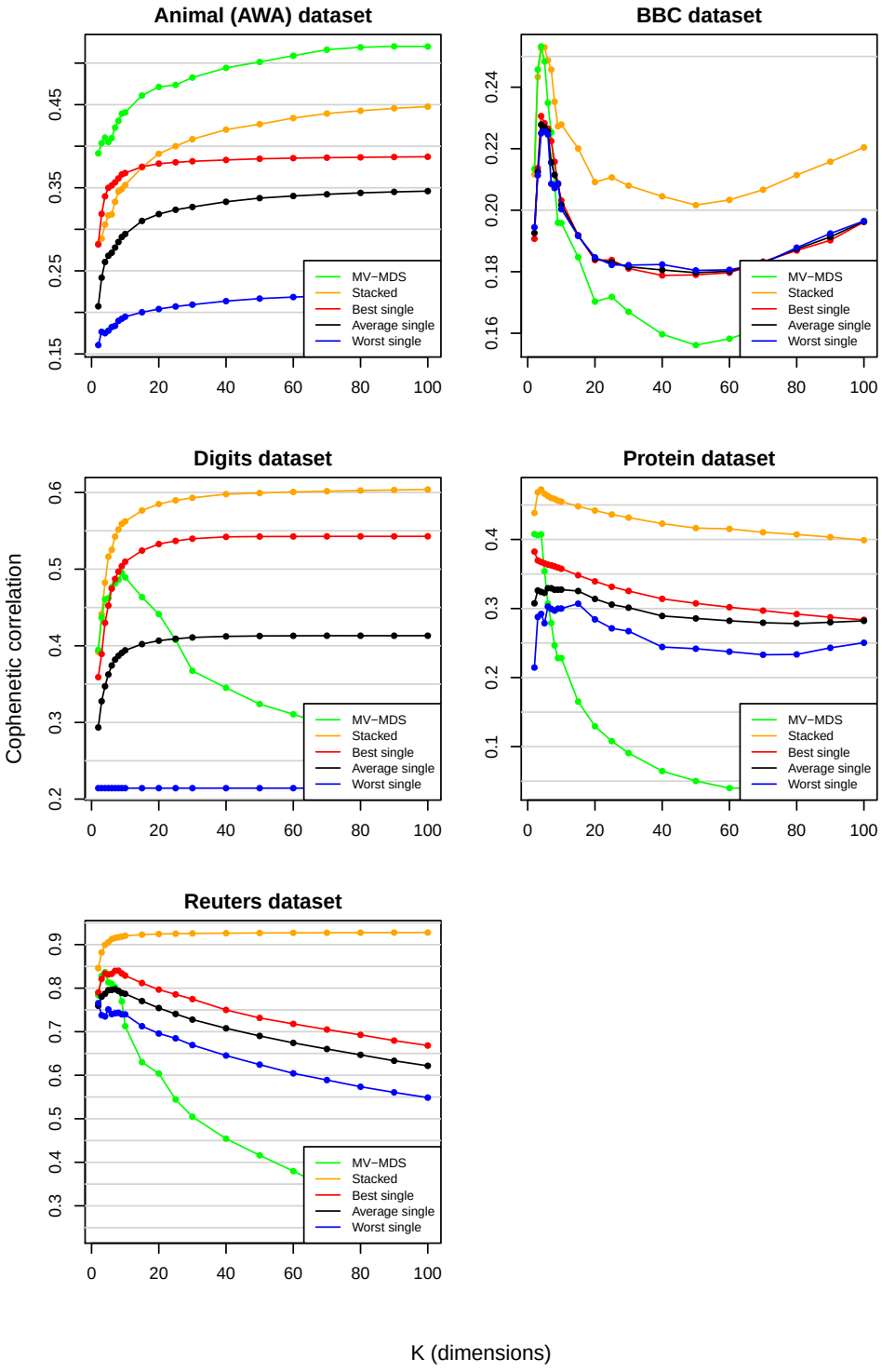


Figure 4.3: MV-MDS dimensionality reduction evaluation with cophenetic correlation (average on all input views).

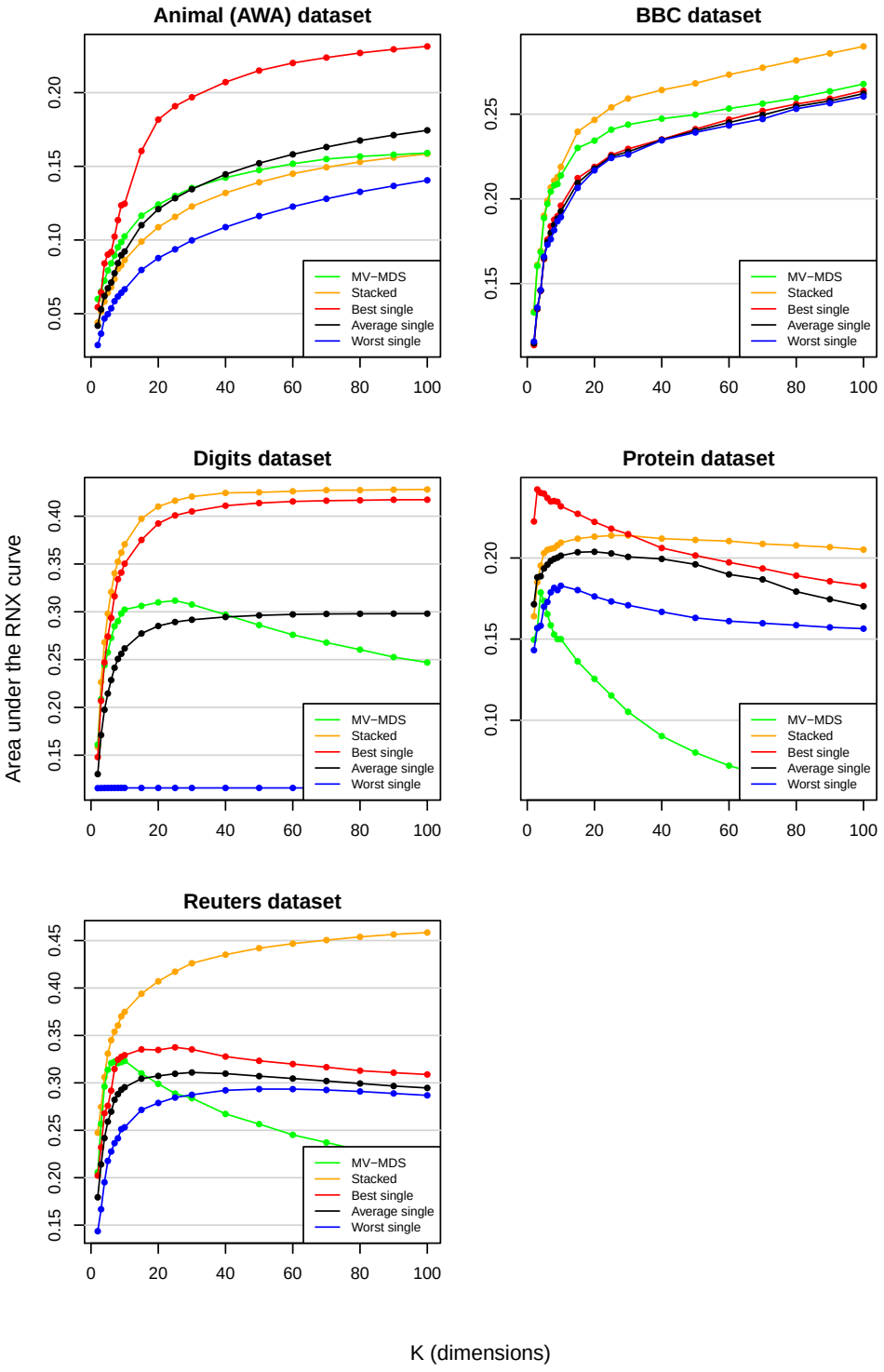


Figure 4.4: MV-MDS dimensionality reduction evaluation with area under the R_{NX} curve (average on all input views).

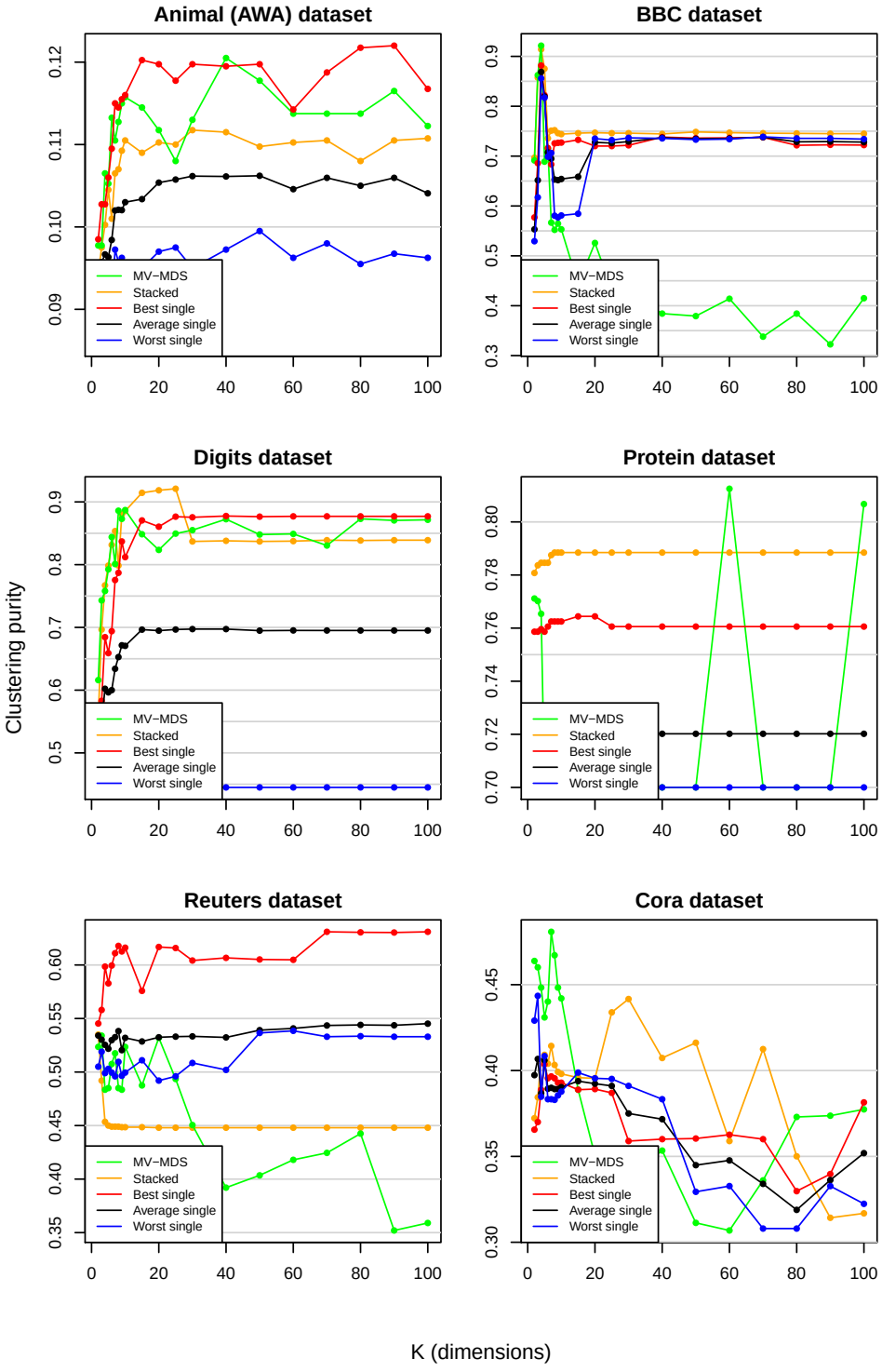


Figure 4.5: MV-MDS clustering evaluation with clustering purity.

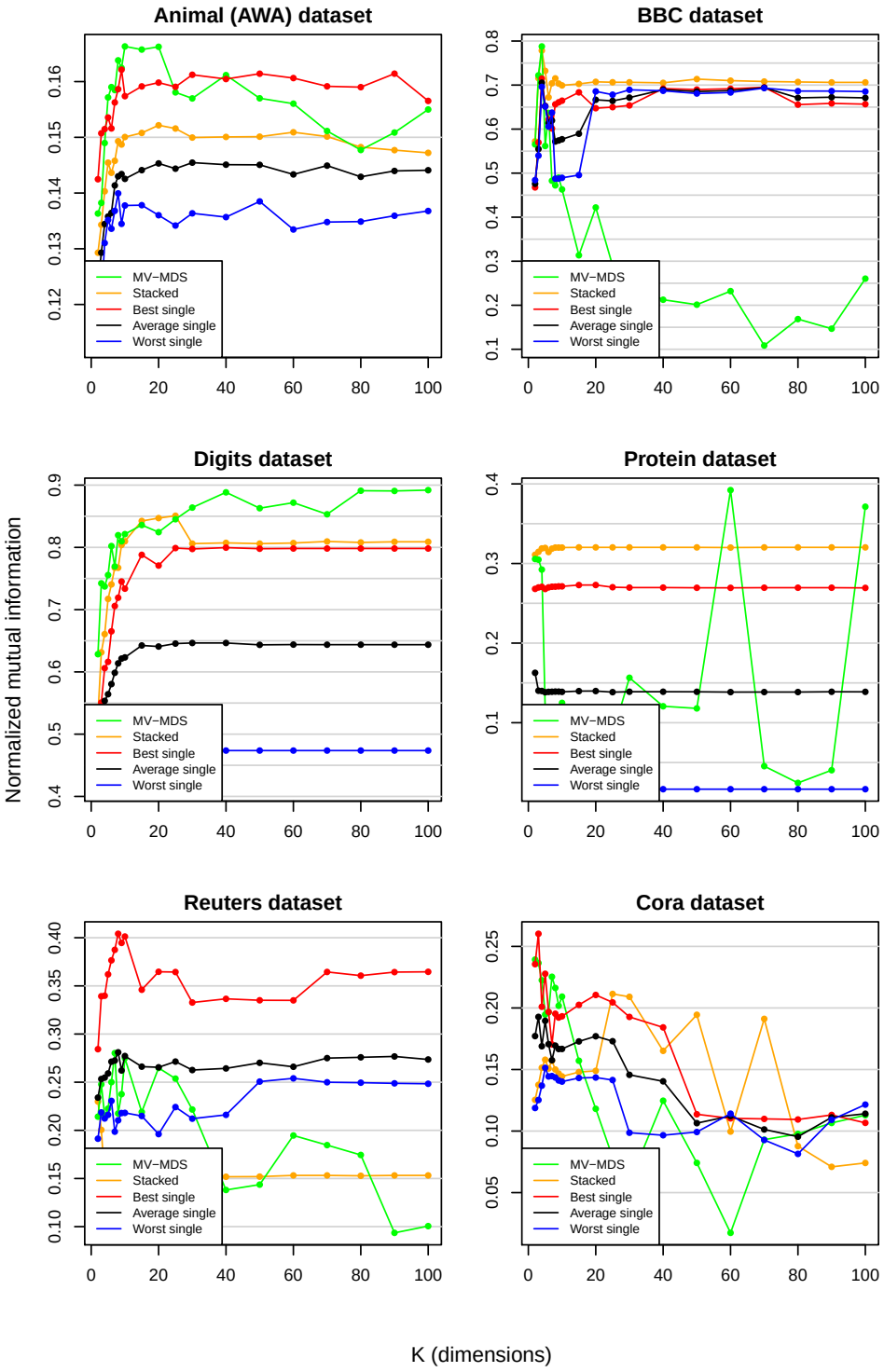


Figure 4.6: MV-MDS clustering evaluation with clustering normalized mutual information.

Davies-Bouldin index (Figure 4.7). The Davies-Bouldin index (DBI) measures the internal properties of the clusters with respect to other clusters, considering the distances in the input spaces. Here, the average Davies-Bouldin index over all input views is given. For DBI values, less is better (less distance).

On the AWA, BBC and Reuters datasets, MV-MDS produces the lowest DBI value, although its behaviour is quite irregular.

On the digits dataset, best single, stacked and MV-MDS show an equivalent performance, alternating with K. Finally on protein dataset, stacked and MV-MDS have higher (worse) values than average single view configurations.

4.4.2 MV-MDS with respect to the state of the art

Table 4.1 shows the most relevant clustering purity results in the state of the art and compares them to the results of MV-MDS on the same datasets and configuration. The clustering purity measures the faithfulness of the produced clustering assignments to the reference class assignments in the dataset. In other words, higher values mean that the clustering assignment is more similar to the class assignment, with a value of 1 meaning that both are identical.

The clustering of the handwritten digits dataset shows high clustering purity (> 0.8) on most methods, with MV-MDS producing a higher value than the other methods. The BBC segmented news dataset is only used on two papers in the state of the art, with quite high values in all cases. Here, MV-MDS also produces a better clustering than the other methods. On the Reuters multilingual dataset, MVSC-BG shows the highest clustering purity value. Finally, the animal with attributes dataset shows a high variability in the results in the state of the art, with MV-KMeans probably using a different preprocessing (it is not specified in the paper). Again, MV-MDS produces the best clustering assignment according to the purity result.

Table 4.2 shows the normalized mutual information (NMI) between the clustering assignments produced by the multiview clustering methods in the state of the art and the reference class labels in the dataset. A higher value implies a more similar assignment, with NMI=1 meaning a perfect match. However NMI measures the coincidence of the cluster assignments differently from purity, as it uses the mutual information between both assignments.

The clustering of the handwritten digits dataset has, in general, high NMI values. MVC-PSS and MVSC-BG show the highest NMI result on this dataset, with MV-MDS slightly below them. All NMI results on the BBC segmented news dataset are relatively similar, with CoKmLDA slightly higher. On the Reuters multilingual dataset, MVSC-BG shows higher NMI value than the other methods. Finally, on the animal with attributes dataset the results are mostly around 0.7 and above, with MV-MDS having NMI=0.843, clearly above the other methods.

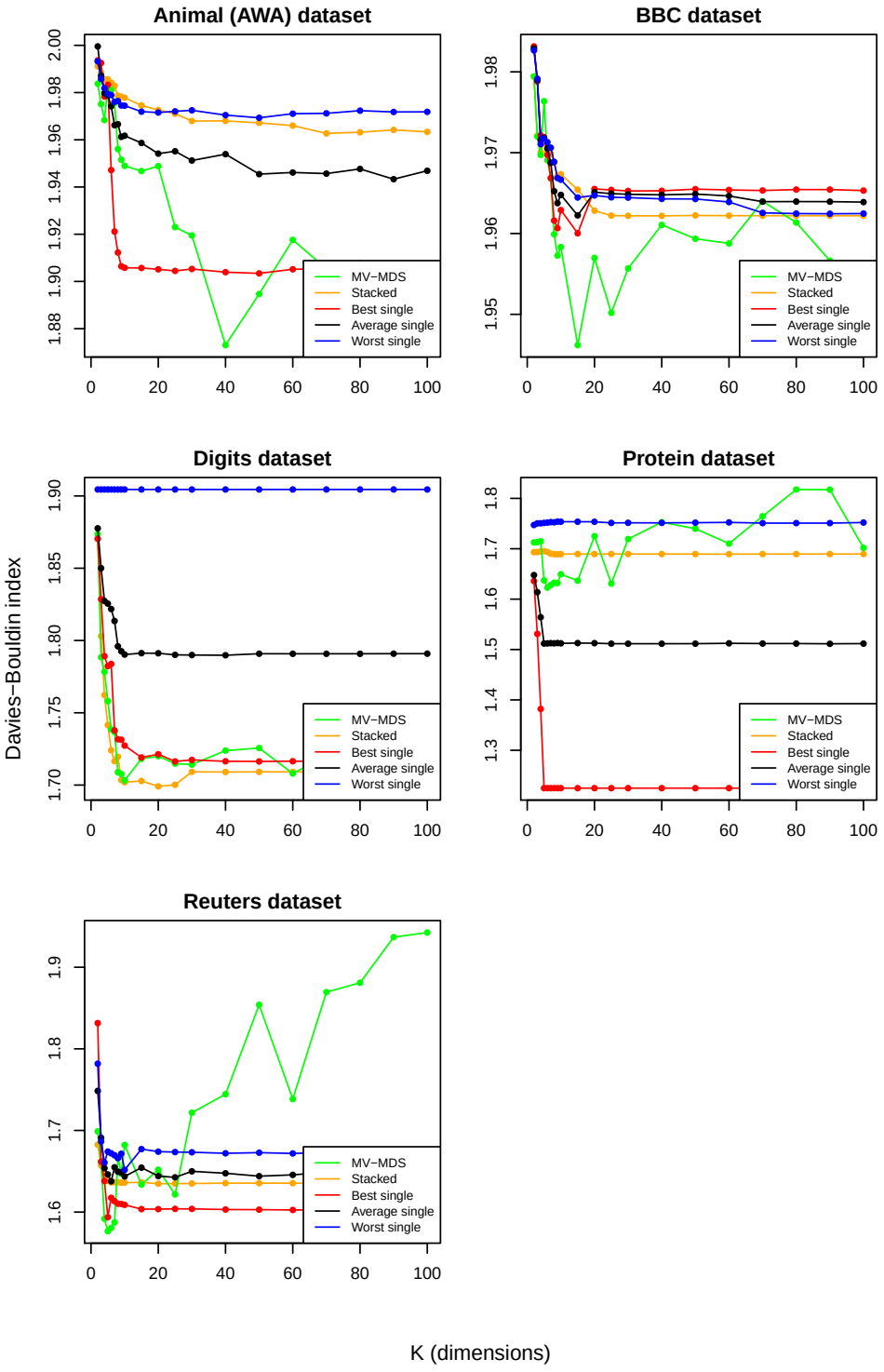


Figure 4.7: MV-MDS clustering evaluation with the Davies-Bouldin index (average on all input views). Less is better.

Method	Digits	BBC	Reuters	AWA
CoregSC	0.822	0.887	0.552	0.580
MMSC	0.758	NA	0.390	0.585
MVC-SS	NA	NA	0.531	0.629
MV-KMeans	0.825	NA	NA	0.114
CoKmLDA	0.819	0.914	NA	NA
MFSC-MO	0.800	NA	NA	NA
MVC-PSS	0.862	NA	NA	0.325
MVSC-BG	0.844	NA	**0.577**	NA
MV-MDS	**0.887**	**0.921**	0.534	**0.775**

"NA" means there are no available results of the method on the data set.

Table 4.1: Clustering purity wrt. the state of the art.

Method	Digits	BBC	Reuters	AWA
CoregSC	0.836	0.769	0.326	0.695
MMSC	0.792	NA	0.134	0.698
MVC-SS	NA	NA	NA	0.751
MV-KMeans	0.807	NA	NA	0.117
CoKmLDA	0.818	**0.796**	NA	NA
MFSC-MO	0.785	NA	NA	NA
MVC-PSS	**0.833**	NA	NA	0.213
MVSC-BG	**0.832**	NA	**0.357**	NA
MV-MDS	0.821	0.788	0.280	**0.843**

"NA" means there are no available results of the method on the data set.

Table 4.2: Clustering NMI wrt. the state of the art.

4.5 Discussion

First, MV-MDS is compared with the baseline methods (single view and stacked views MDS). In the dimensionality reduction task, MV-MDS shows an overall better performance than the baseline methods, reflected in the highest absolute SVM score on five datasets as well as a superior performance on the majority of dimensionalities and datasets. In general it performs better than the average single-view and the stacked-view MDS configurations, which are more realistic than the best single-view, as in a real task it probably will not be known which is the single best view in a multiview dataset.

The unsupervised metrics for dimensionality reduction evaluation (cophenetic correlation and AUC-RNX) give an average or lower than average score to MV-MDS. These metrics measure the similarity of the low-dimensional

space generated by the method with the original input space; in the case of multiview datasets, the average measures over all the input views has been computed. It is expected that the projections using the original views (one by one or stacked) resemble the original data very closely. The fact that MV-MDS scores in middle or lower positions in these metrics suggests that the projection it generates has an internal structure that differs from the structure of the original input views. This is reasonable, as MV-MDS has to include information from all the views and as a consequence it has to produce an innovative data structure.

Regarding the clustering performance of MV-MDS, its performance is on the top on some of the datasets (3 or 4 depending on the metric considered). However, its average performance is not clearly superior to the average single-view or stacked views configurations. This is to be expected, as neither MDS nor MV-MDS are specifically designed as clustering methods, but as dimensionality reduction methods.

Compared with other multiview clustering methods in the state of the art, the method proposed in this thesis (MV-MDS) shows better overall clustering results when measured with clustering purity in three datasets, while its results with NMI are only slightly below the best ones in two datasets and clearly above the other methods in AWA dataset. This makes MV-MDS a reasonable alternative for multiview clustering, comparable or slightly better than other methods in the state of the art. This result is specially remarkable given the fact that neither MDS or MV-MDS are initially conceived as clustering methods, but as dimensionality reduction methods instead.

A possible explanation for these results is that the dimensionality reduction produced by MV-MDS condenses the most relevant information of the multiview datasets, while discarding noise or contradictory data. As a consequence, the posterior clustering performed on the low-dimensional projection generated by MV-MDS is of high quality.

MV-MDS computes the common eigenvectors of the normalized distance matrices of the input data views. The distance matrices of the different views of a multiview dataset are supposed to contain a considerable amount of shared information, as the distances between samples in different views are likely to be coherent in general. The probable cause of the good results of MV-MDS as a multiview clustering method may be the fact that, when computing the common eigenvectors, the redundant information from the distance matrices is (1) retrieved first as it has a stronger signal, and (2) is only retrieved once, leaving room for other useful information in the common subspace matrix generated. This behaviour may be making MV-MDS better at synthesizing the information contained in the multiple views, therefore feeding richer information into the clustering stage.

On the other hand, the qualitative advantages of MV-MDS over other

multiview clustering methods in the state of the art are the following.

- Some methods, like CoregSC [62] and MMSC [12], can only work with two input data views. MV-MDS can work with any number of input views.
- CoregSC [62], MV-KMeans [11] and MVC-SS [106] can only work with feature space input data. MV-MDS can work with either feature or graph space input data, or with any combination of both.
- CoKmLDA [117] requires all data views to have the same dimensionality, while MV-MDS can work with data matrices of different dimensionality.

Chapter 5

Multiview spectral clustering and Laplacian eigenmaps

5.1 Motivation

Spectral clustering (SC) [92] is a well-known clustering algorithm that is based on spectral graph theory [96]. One of its distinctive features is that it clusters samples by connectivity, not merely by distance, and therefore it allows the clustering of data sets with concave or nested clusters. It is one of the most frequently used clustering methods when such complex data has to be processed and clustered.

Laplacian eigenmaps [7] is a dimensionality reduction algorithm that is very closely related to spectral clustering, in fact in some formulations the Laplacian eigenmaps algorithm is included in the SC algorithm. As a consequence, Laplacian eigenmaps show most of the properties of SC, as the fact that it groups points by connectivity, being able to identify sets of contiguous samples.

This Chapter presents a novel multiview spectral clustering algorithm, MVSC-CEV (multiview spectral clustering by common eigenvectors) that extends the features of spectral clustering to multiview datasets. The experiments on standard multiview data sets show that the MVSC-CEV algorithm has a better overall clustering performance than existing multiview clustering algorithms.

The development of MVSC-CEV also allows to develop a multiview version of Laplacian eigenmaps. Thus, a new multiview dimensionality reduction method is proposed, multiview Laplacian eigenmaps. However, for the sake of simplicity both algorithms will be referred simply as MVSC-CEV, either applied to clustering or to dimensionality reduction tasks.

5.2 Related work

The MVSC-CEV algorithm is mainly based on two well known algorithms: the Ng, Jordan and Weiss (NJW) spectral clustering algorithm [84] and stepwise common principal components method [100]. In this Section, we outline both algorithms in order to provide a theoretical background for the MVSC-CEV algorithm described in Section 5.3.

5.2.1 Spectral clustering

Spectral graph theory gives the conditions under which a graph can be partitioned into non-connected subgraphs. The spectral clustering algorithm [92] is an application of spectral graph theory to the task of graph clustering, which can be further applied to any data expressed as a matrix of similarities between samples.

There are several variants of the spectral clustering algorithm. The method presented in this paper is based on the variant described in [84], as it is a well accepted and proven variant and it allows a reduction in the computational cost of the MVSC-CEV algorithm, as will be explained in Section 5.3. As a reference for the definition of MSCV-CEV on Section 5.3, the NJW spectral clustering algorithm is shown in Algorithm 7.

Algorithm 7 Spectral Clustering - NJW variant

Input: a set of data samples $P = \{p_1, \ldots, p_n\}$ and the number k of desired clusters.

1. Build a similarity matrix $S \in \mathbb{R}^{n \times n}$ from P using the Gaussian similarity function $G_{ij} = exp(-\|p_i - p_j\|^2/2\sigma^2)$.
2. Construct the normalized symmetrical Laplacian matrix $L = D^{-1/2}SD^{-1/2}$, where D is the diagonal matrix whose (i,i) element is the sum of the i-th row of S.
3. Create a matrix $X \in \mathbb{R}^{n \times k}$ with the k largest eigenvectors of L disposed in columns.
4. Normalize X so that each row has unit length, obtaining $Y \in \mathbb{R}^{n \times k}$.
5. Apply K-means or another clustering algorithm to Y.

Output: a clustering assignment in k clusters of the n samples in P.

The only parameter besides the desired number of clusters k is σ, which controls the influence of the distance between two points p_i and p_j on their similarity s_{ij}. By default, the heuristic proposed in [87] is used to choose

σ. However, the similarity matrix S built in Step 1 can be obtained using a different similarity function.

Note that the algorithm is formulated to receive data in feature space (observations/features matrix) as input. However, it can also be applied to data in graph space (adjacency or similarity matrix).

An important feature of spectral clustering is that it groups points by connectivity, not merely by distance, i.e. if point a is connected to b and b to c then a and c will be assigned to the same cluster, even if the distance between a and c is relatively large. This allows clusterings on data sets with concave groups of points to be found, identifying clusters similarly to how humans identify connected shapes.

5.2.2 Laplacian Eigenmaps

Laplacian Eigenmaps is a method of dimensionality reduction described in [7] that tries to find a low-dimensional manifold supposedly embedded in a high-dimensional space. Laplacian Eigenmaps algorithm comprises three main stages: (1) computing an adjacency matrix between the input points, (2) computing the Laplacian matrix of the adjacency matrix, and (3) computing the eigenvectors of the Laplacian matrix. The detailed Laplacian Eigenmaps method is presented in Algorithm 8. Although the original formulation by [7] proposes the use of a heat kernel to compute the adjacency matrix, other adjacency functions have been proposed for this method, for example Gaussian radial basis functions.

Algorithm 8 . Laplacian Eigenmaps

Input: dataset $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$,

Output: low-dimensional data representation $\mathcal{Y} = \{y_1, y_2, \ldots, y_n\}, y_i \in \mathbb{R}^k$

function LAPLACIANEIGENMAPS($\mathcal{X}$)

compute adjacency matrix A using the Gaussian similarity function, so that $A_{ij} = exp(-\|p_i - p_j\|^2/2\sigma^2)$.

compute the normalized Laplacian matrix $L = D^{-1/2}SD^{-1/2}$, where D is the diagonal matrix whose (i, i) element is the sum of the i-th row of S.

let $\mathcal{Y}$ be the first k eigenvectors of L

end function

Laplacian Eigenmaps are very closely related to Spectral Clustering. According to their authors, the application of a standard clustering algorithm (as K-means) to the space generated by Laplacian Eigenmaps would produce the same clustering as in the Spectral Clustering algorithm. This is specially clear if a Gaussian similarity function is used to compute the adjacencies in the Laplacian Eigenmaps algorithm.

5.3 MVSC-CEV description

5.3.1 Description of the algorithm

The multiview spectral clustering algorithm by common eigenvectors presented in this paper (MVSC-CEV), detailed in Algorithm 9, has an structure that resembles that of the NJW spectral clustering algorithm (described in Section 5.2.1). The main difference lies in the fact that MVSC-CEV replaces the single input data matrix with a set of C input data views $\bar{V} = \{V_1, V_2, \dots V_C\}$. In turn, for each input matrix in $\bar{V}$, a similarity matrix and its corresponding Laplacian matrix are computed. Then the eigenvectors common to all C Laplacian matrices are computed and used to obtain the clustering.

Algorithm 9 Multiview spectral clustering by common eigenvectors (MVSC-CEV)

Input: C view matrices $\bar{V} = \{V_1, V_2, \dots V_C\}$ of the data (with n samples each) and the number k of desired clusters.

1. For each data view $V_c \in \bar{V}$, compute a similarity matrix $S_c \in \mathbb{R}^{n \times n}$ using the Gaussian similarity function $G_{ij} = exp(-\|p_i - p_j\|^2/2\sigma^2)$. The final result is a set of similarity matrices $\bar{S} = \{S_1, S_2, \dots S_C\}$

2. For each similarity matrix $S_c \in \bar{S}$ construct the normalized symmetrical Laplacian matrix $L_c = D^{-1/2} S D^{-1/2}$, where D is the diagonal matrix whose (i,i) element is the sum of S_c's i-th row. The result is a set of Laplacian matrices $\bar{L} = \{L_1, L_2, \dots L_C\}$

3. Create a matrix $X \in \mathbb{R}^{n \times k}$ with the k largest common eigenvectors of the matrices in $\bar{L}$, computed using the S-CPC algorithm (Section 4.2.2).

4. Normalize X so that each row has unit length, obtaining $Y \in \mathbb{R}^{n \times k}$.

5. Apply K-means or another clustering algorithm to Y.

Output: a single clustering assignment of the n input data samples in k clusters.

A consequence of using the NJW spectral clustering formulation is that only the k largest eigenvectors have to be computed. S-CPC algorithm makes this possible in an efficient way, thus reducing the computational complexity of Step 3 of Algorithm 9 from $O(Cn^2)$ to $O(Ckn)$, with $k \ll n$ in the vast majority of cases.

MVSC-CEV can operate on both feature space and graph space input views, and on any combination of both. Input matrices in feature space can have any dimension, possibly differing across the different matrices.

5.3.2 Ideal clustering case

To show why Algorithm 9 works as expected, let us consider first an ideal case, where there are $k = 3$ perfectly separated clusters, i.e. the points in different clusters are infinitely far apart from each other. Moreover, all C data views have the same clustering structure. In order to simplify the discussion, let us also assume that the points in the data views are ordered according to the cluster they belong to, so points belonging to cluster 1 appear first, then points of cluster 2 and finally the points of cluster 3.

Consequently, the similarity matrices of this example will be block-diagonal: $\forall S \in \{S_1, S_2, \dots S_C\}$, $S_{ij} = 0$ if data points i and j do not belong to the same cluster, or greater than zero otherwise. Representing each non-zero subblock of the similarity matrices with a parenthesized superscript:

$$S_c = \begin{bmatrix} S^{(1)} & 0 & 0 \\ 0 & S^{(2)} & 0 \\ 0 & 0 & S^{(3)} \end{bmatrix} \quad \forall S_c \in \bar{S} \tag{5.1}$$

On the next step of the algorithm, for each similarity matrix S_c a Laplacian matrix L_c is computed, whose block-diagonal structure will be the same:

$$L_c = \begin{bmatrix} L^{(1)} & 0 & 0 \\ 0 & L^{(2)} & 0 \\ 0 & 0 & L^{(3)} \end{bmatrix} \quad \forall L_c \in \bar{L} \tag{5.2}$$

In step 3 of Algorithm 9, the set of Laplacian matrices $\bar{L}$ defined in (5.2) is passed to the S-CPC algorithm along with k in order to compute their common eigenvectors. According to [100], S-CPC finds the k eigenvectors whose sum of eigenvalues is highest. Each such eigenvector is located on the $\mathbb{R}^n$ sphere, where n is the number of samples in the input data. The common eigenvectors are mutually orthogonal. Given the common structure of the Laplacian matrices in $L_c \in \bar{L}$, they have the same eigenvectors, which therefore are the common eigenvectors computed by S-CPC. Following [84], the k largest eigenvectors of the Laplacian matrices in $L_c \in \bar{L}$ are the first eigenvectors (i.e. those with largest eigenvalue) $x_1^{(i)}$ of each submatrix $L_c^{(i)}$, properly padded with zeros to complete the missing elements:

$$X = \begin{bmatrix} x_1^{(1)} & \vec{0} & \vec{0} \\ \vec{0} & x_1^{(2)} & \vec{0} \\ \vec{0} & \vec{0} & x_1^{(3)} \end{bmatrix} \in \mathbb{R}^{n \times k} \tag{5.3}$$

Finally, the normalization of the rows of X to make them of unit length results in a matrix Y, of the form:

$$Y = \begin{bmatrix} y^{(1)} & \vec{0} & \vec{0} \\ \vec{0} & y^{(2)} & \vec{0} \\ \vec{0} & \vec{0} & y^{(3)} \end{bmatrix} \in \mathbb{R}^{n \times k} \tag{5.4}$$

where $y^{(i)}$ is a vector of ones with as many values as the number of elements in cluster i. Applying K-Means to Y produces the clustering assignment of the input data samples common to the C input views.

5.3.3 Deviations from the ideal case

On multiview clustering problems there are two possible sources of deviation from the ideal case discussed in Section 5.3.2.

The first possible situation occurs when the off-diagonal blocks in the similarity matrices $S \in \bar{S}$ are not zero, i.e. the clusters are not perfectly separated from each other. This case is discussed in [84] for a single similarity matrix, but its extension to several similarity matrices with the same structure is straightforward.

The second possible deviation from the ideal case stems from structural discrepancies across data views, where not all similarity matrices share the same structure. Obviously this second deviation implies the former one, as at least some of the views cannot exhibit a perfect block-diagonal structure if there are differences between them.

In this case, the eigenvalues associated to each of the different Laplacian matrices in $\bar{L}$ may not decrease simultaneously; but even in such a case, S-CPC guarantees that the **sum of the eigenvalues** associated to each successive common eigenvector is decreasing. Let $\lambda_c^{(i)}$ be the eigenvalue associated to Laplacian matrix L_c obtained on iteration i of S-CPC, i.e. associated with the i-th eigenvector. Therefore, the following relation holds:

$$\sum_{c=1}^{C} \lambda_c^{(i)} \geq \sum_{c=1}^{C} \lambda_c^{(i+1)} \quad \forall i = 1, 2, \dots k \tag{5.5}$$

in other words, the *eigengaps* (difference between consecutive eigenvalues) are conserved:

$$\delta^{(i)} = \sum_{c=1}^{C} \lambda_c^{(i)} - \sum_{c=1}^{C} \lambda_c^{(i+1)} \geq 0 \quad \forall i = 1, 2, \dots k \tag{5.6}$$

This satisfies the matrix perturbation theory condition [96], that guarantees the stability of the *subspace* defined by the first k eigenvectors a matrix as long as the eigengaps are conserved.

5.3.4 Multiview Laplacian Eigenmaps

Given the close correspondence between Spectral Clustering and Laplacian Eigenmaps described in Section 5.2.2, an equivalent correspondence can be established between the Multiview Spectral Clustering algorithm presented here (Section 5.3) and a multiview extension to Laplacian Eigenmaps. The Multiview Laplacian Eigenmaps algorithm (MV-LE) proposed is a variation of Algorithm 9, where the resulting low-dimensional projection is the normalized common eigenvector matrix Y. MV-LE is detailed in Algorithm 10.

Algorithm 10 Multiview Laplacian Eigenmaps

Input: C view matrices $\bar{V} = \{V_1, V_2, \dots V_C\}$ of the data (with n samples each) and the number k of desired dimensions of the projection.

1. For each data view $V_c \in \bar{V}$, compute a similarity matrix $S_c \in \mathbb{R}^{n \times n}$ using the Gaussian similarity function $G_{ij} = exp(-\|p_i - p_j\|^2 / 2\sigma^2)$. The final result is a set of similarity matrices $\bar{S} = \{S_1, S_2, \dots S_C\}$

2. For each similarity matrix $S_c \in \bar{S}$ construct the normalized symmetrical Laplacian matrix $L_c = D^{-1/2} S D^{-1/2}$, where D is the diagonal matrix whose (i, i) element is the sum of S_c's i-th row. The result is a set of Laplacian matrices $\bar{L} = \{L_1, L_2, \dots L_C\}$

3. Create a matrix $X \in \mathbb{R}^{n \times k}$ with the k largest common eigenvectors of the matrices in $\bar{L}$, computed using the S-CPC algorithm (Section 4.2.2).

4. Normalize X so that each row has unit length, obtaining $Y \in \mathbb{R}^{n \times k}$.

Output: a single k-dimensional projection (Y) of the multiview input samples.

5.4 Results

Strictly speaking, MVSC-CEV is a multiview clustering algorithm and MV-LE is a multiview dimensionality reduction algorithm. However, given that they are so closely related (their code is almost identical) and for the sake of simplicity, the results from MV-LE when applied to dimensionality reduction maps will simply be referred to as results for MVSC-CEV.

5.4.1 MVSC-CEV with respect to SC baseline

The first block of experiments compare the proposed MVSC-CEV method with the spectral clustering baseline, either applied to each input view independently, or applied to all the input views stacked into a single data matrix.

The goal of these experiments is to assess the advantages of the multiview method proposed with respect to single view approaches.

In order to synthesize the numerous results (different metrics, datasets and embedding dimensionalities), the results of each evaluation metric are summarized in a set of graph plots, one for each dataset in the experiments. Therefore, six graphs per evaluation metric are produced (five on the unsupervised metrics, as the Cora dataset has a graph space view that is not compatible with these metrics).

However, the detailed numerical results are given in Appendix C, with one table for each combination of evaluation metric and dataset, in a total of 33 tables to evaluate the MVSC-CEV method.

5.4.1.1 Dimensionality reduction evaluation

There are three evaluation metrics for the dimensionality reduction task: SVM classification, cophenetic correlation (average on all the input views) and area under the curve of the R_{NX} value (average on all input views).

Figure 5.1 shows the two dimensional projections of two example datasets using MVSC-CEV.

SVM classification (Figure 5.2). The results on the animal with attributes (AWA) dataset are low and irregular in general, given the specific difficulty of the task (there are 50 classes in this dataset). There is an initial peak and a posterior descent around K=10, probably induced by added information that is not related to the classification task and misguides the algorithm. Then the results rise again around K=50, which happens to be the number of classes in the dataset and as such a recommended dimensionality for this dataset. Apparently one of the single views shows a better performance than the other configurations.

The BBC segmented news dataset has two views that actually are two text segments of the same document, and as a consequence both views are mainly equivalent. This is reflected in the SVM results for BBC, where MVSC-CEV and stacked-SC results overlap as there is no practical difference between both approaches in this dataset. MVSC-CEV (and stacked-SC) perform consistently better than single-view SC, showing the usefulness of adding more text segments to the dataset.

On the handwritten digits dataset, MVSC-CEV performs slightly better than stacked-SC on most dimensionalities. Here there is a considerable difference of performance between single views, what leads to thinking that some views are not so useful in the classification task.

On the Berkeley protein dataset, MVSC-CEV clearly performs above the baseline methods for all dimensionalities ≤ 30. This is specially interesting as it is the range where most information compression is achieved. Interestingly,

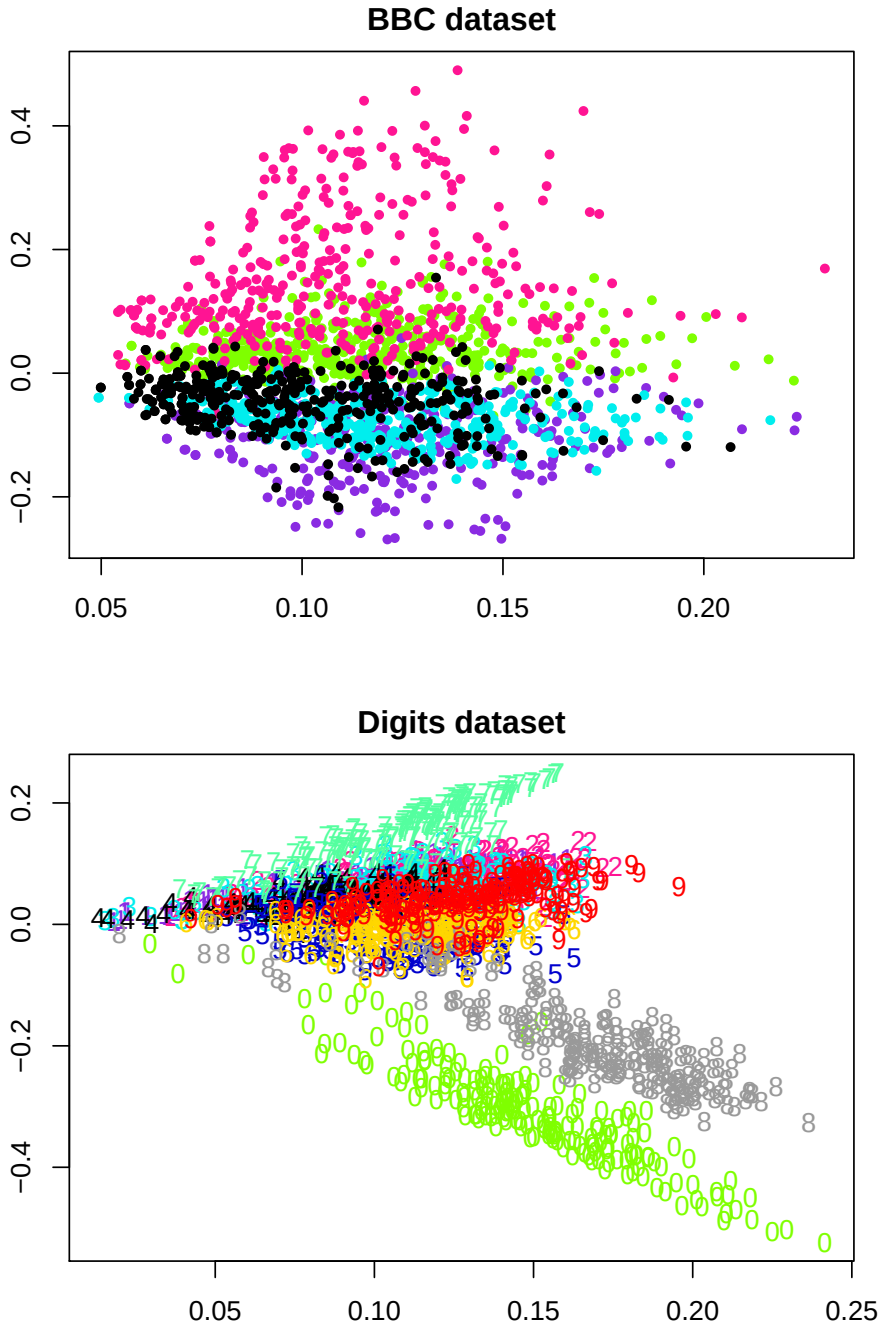


Figure 5.1: MVSC-CEV projection of two example datasets.

the worst single-view configuration ends performing the best in very high dimensionalities (≥ 70).

On the Reuters dataset, the best single view SC performs better than the other methods. MVSC-CEV, however, performs clearly better than stacked and average single view configurations on most K's.

Finally, on the Cora dataset the stacked-SC performs better for low dimensionalities, achieving the absolute maximum. The best single-SC performs better at higher dimenensionalities. In this case, the worst single view seems to attract both MVSC-CEV and stacked-SC, as they closely follow its curve of results.

Cophenetic correlation (Figure 5.3). The cophenetic correlation measures the similarity of the distances of the new space with respect to the original input spaces. The average cophenetic correlation on all input views is provided. On the AWA and digits datasets, MVSC-CEV shows a performance around the average single view SC.

On the BBC dataset, its cophenetic correlation is consistently above single view SC, although overlapped with stacked-SC due to the design of this dataset.

On the Berkeley protein dataset, MVSC-CEV shows the higher correlation on low dimensionalities, although then it drops with dimensionalities ≥ 20.

Finally, on the Reuters dataset all methods start with a high value to drop with K. MVSC-CEV and stacked rank below average single views.

Area under the R_{NX} curve (Figure 5.4). The area under the R_{NX} curve (AUC-RNX) also measures the similarity of the sample neighbourhoods in the projected space with respect to the original input view. Here, the average AUC-RNX over all input views is given.

In the AWA and BBC datasets, MVSC-CEV shows the highest AUC-RNX values. On the other hand, it shows values around the average on the digits and protein datasets.

On the Reuters dataset, MVSC-CEV and stacked show intertwined results, with MVSC-CEV in general slightly above and showing the absolute maximum AUC-RNX values.

5.4.1.2 Clustering evaluation

There are three evaluation metrics for the clustering task: the clustering purity, the clustering normalized mutual information, and the Davies-Bouldin index (average on all input views).

Clustering purity (Figure 5.5). MVSC-CEV consistently shows the highest clustering purity on the animal dataset, and also on most dimensionalities on the Cora dataset.

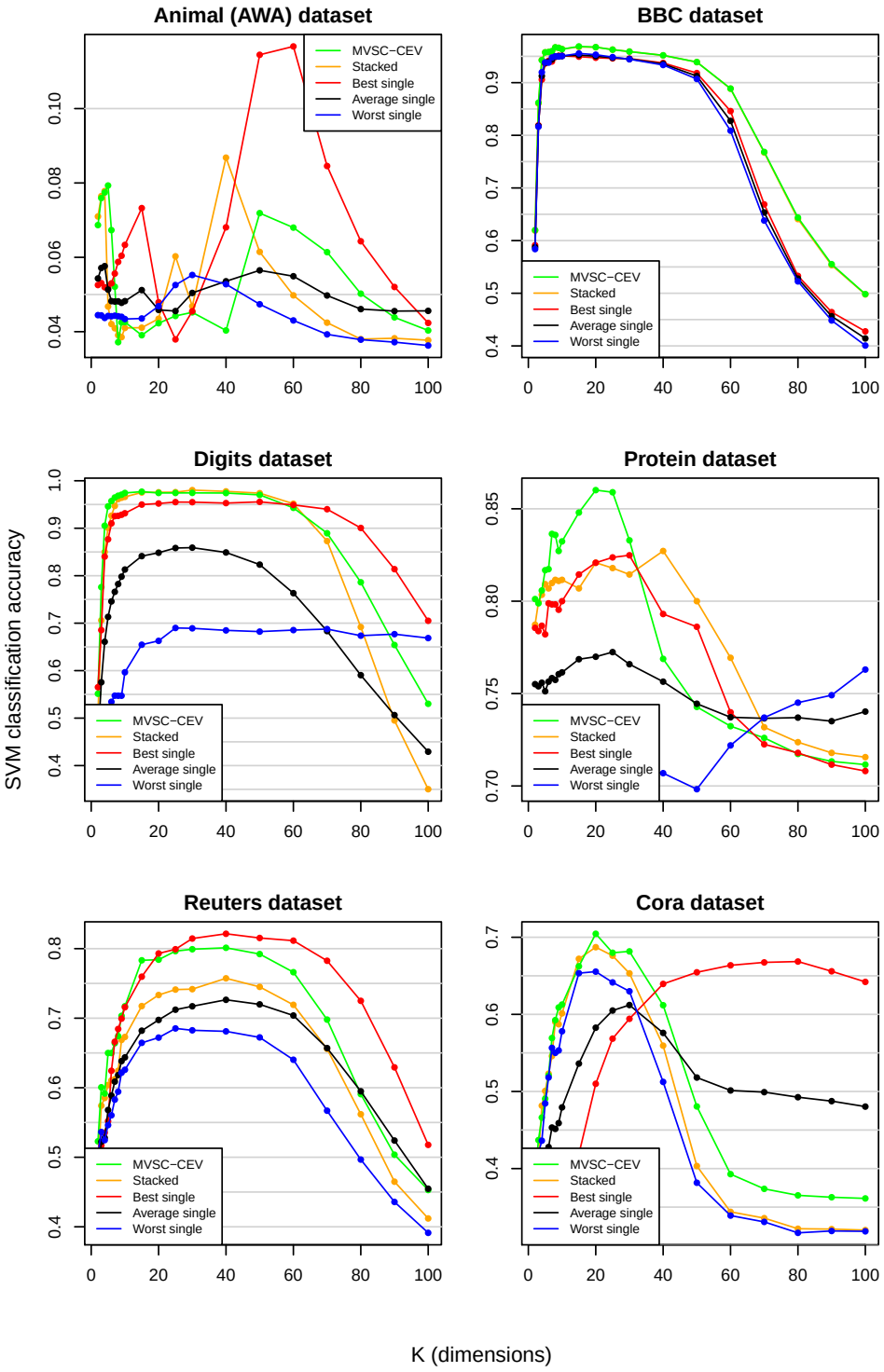


Figure 5.2: MVSC-CEV dimensionality reduction evaluation with SVM classification.

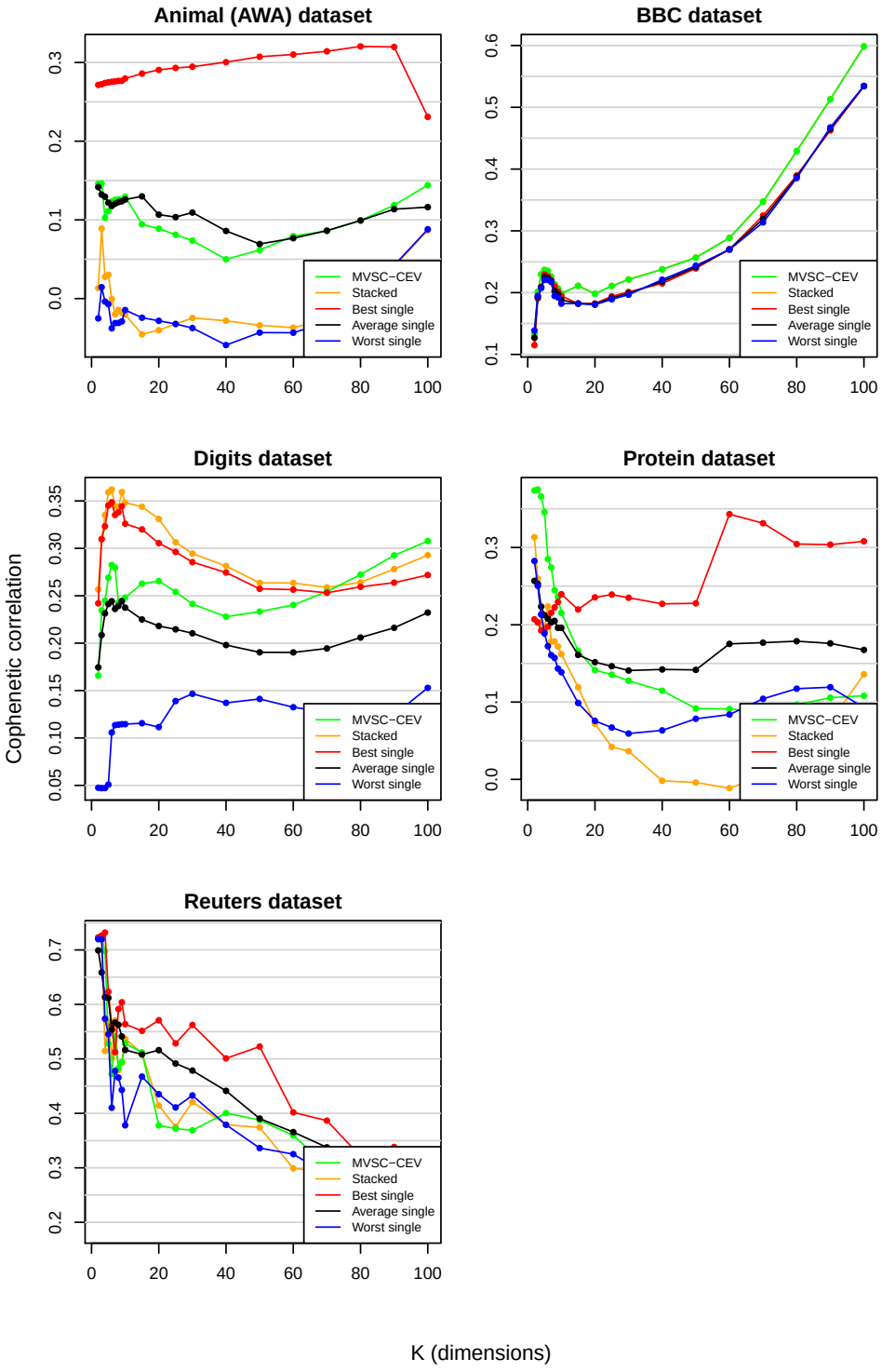


Figure 5.3: MVSC-CEV dimensionality reduction evaluation with cophenetic correlation (average on all input views).

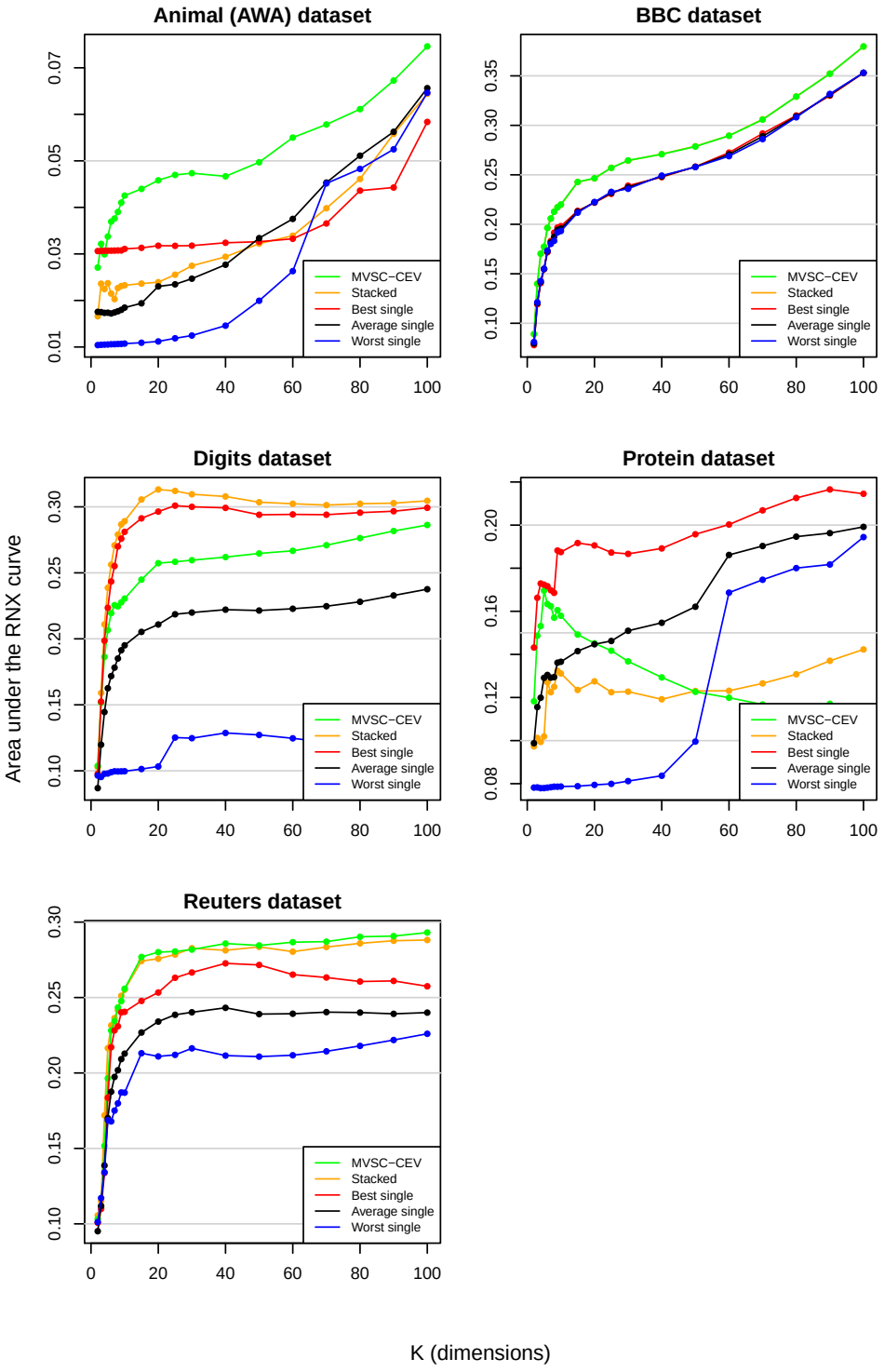


Figure 5.4: MVSC-CEV dimensionality reduction evaluation with area under the R_{NX} curve (average on all input views).

On specific dimensionalities, MVSC-CEV produces the highest purity values on the BBC, digits and proteins datasets. The first and last of these datasets show quite irregular results, with a tendency on the protein dataset of attracting the purity values to 0.70.

Finally, on the Reuters dataset the best single view shows the best performance. MVSC-CEV is clearly better than stacked and average single views.

Clustering normalized mutual information (Figure 5.6). The clustering NMI results are quite similar to the clustering purity results, with MVSC-CEV showing the best results on animal, BBC, digits, protein and Cora datasets.

On the Reuters dataset, the best single view configuration performs better than the other methods. MVSC-CEV, however, performs better than stacked and average single views.

Davies-Bouldin index (Figure 5.7). The Davies-Bouldin index (DBI) measures the internal properties of the clusters with respect to other clusters, considering the distances in the input spaces. Here, the average Davies-Bouldin index over all input views is given. For DBI, less is better.

On the animal and protein datasets, there is a wide difference between the best single view SC and the other configurations. Both MVSC-CEV and stacked show a DBI around that of the worst single view.

On the BBC dataset, MVSC-CEV and stacked DBI results overlap on most K's and show the best results. On the digits dataset, the best single view overlaps and finally improves over MVSC-CEV and stacked views.

Finally, on the Reuters dataset all results are mixed and irregular, with best single view showing the absolute minimum DBI value.

5.4.2 MVSC-CEV with respect to the state of the art

Table 5.1 shows the most relevant clustering purity results in the state of the art and compares them to the results of MVSC-CEV on the same datasets and configuration. The clustering purity measures the faithfulness of the produced clustering assignments to the reference class assignments in the dataset. In other words, higher values mean that the clustering assignment is more similar to the class assignment, with a value of 1 meaning that both are identical.

The clustering of the handwritten digits dataset shows high clustering purity (> 0.8) on most methods, with MVSC-CEV producing a clearly higher value than the other methods. The BBC segmented news dataset is only used on two papers in the state of the art, with quite high values in all cases. Here, MVSC-CEV also produces a better clustering than the other methods. The Reuters multilingual dataset renders lower purity values in general, as it appears to be a harder task. In any case, MVSC-CEV still holds the best purity result. Finally, the animal with attributes dataset shows a high variability in

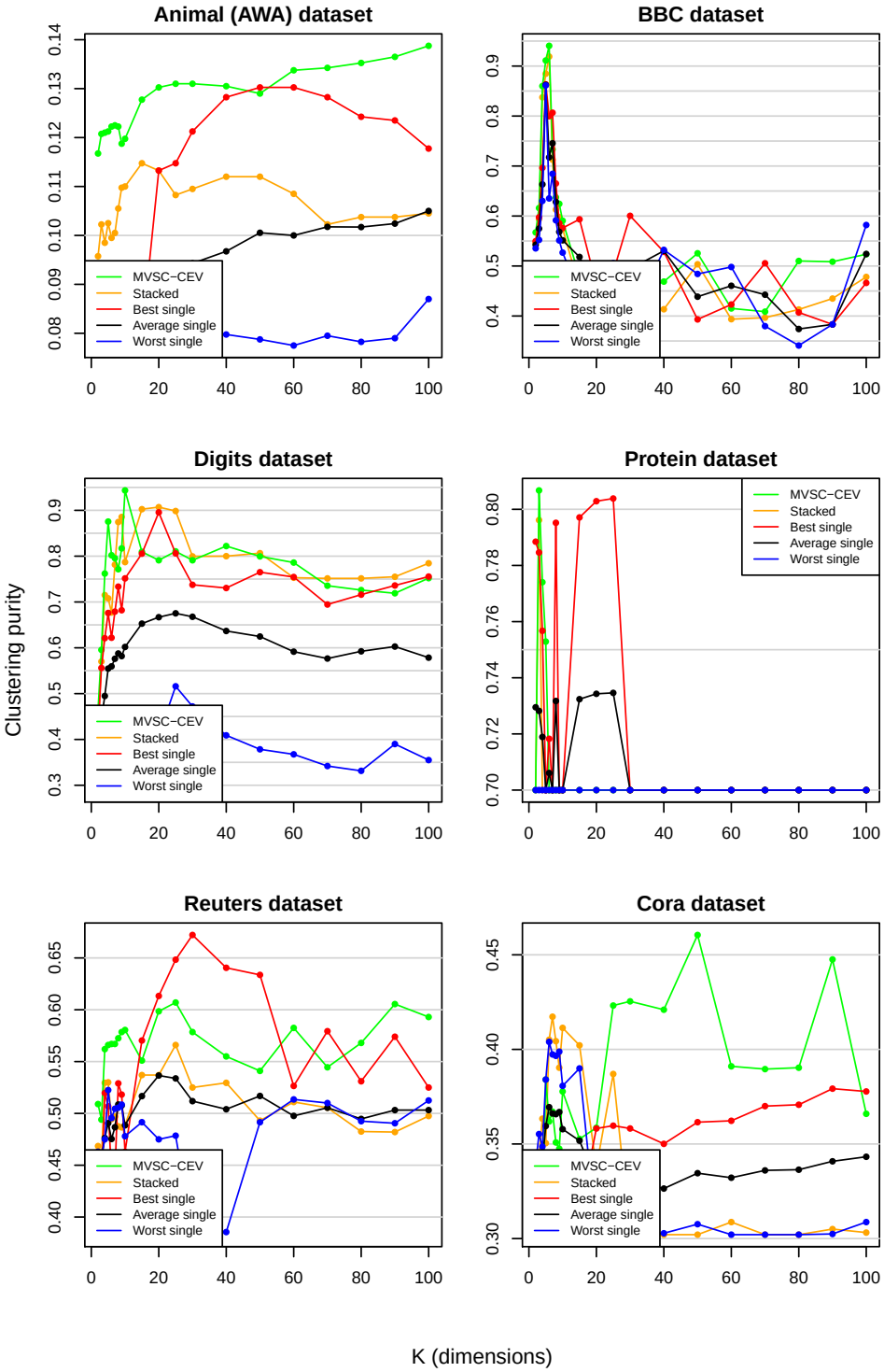


Figure 5.5: MVSC-CEV clustering evaluation with clustering purity.

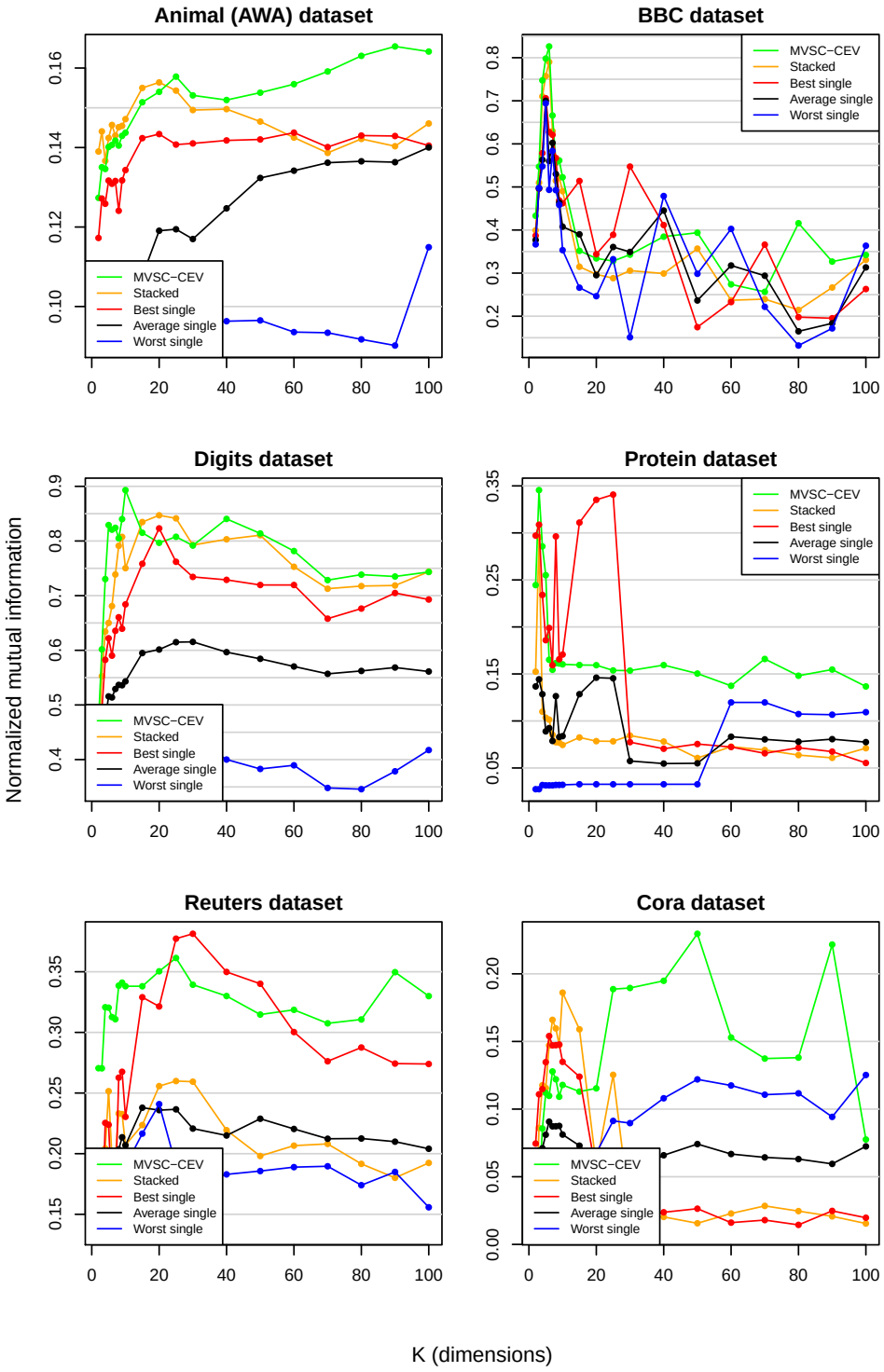


Figure 5.6: MVSC-CEV clustering evaluation with clustering normalized mutual information.

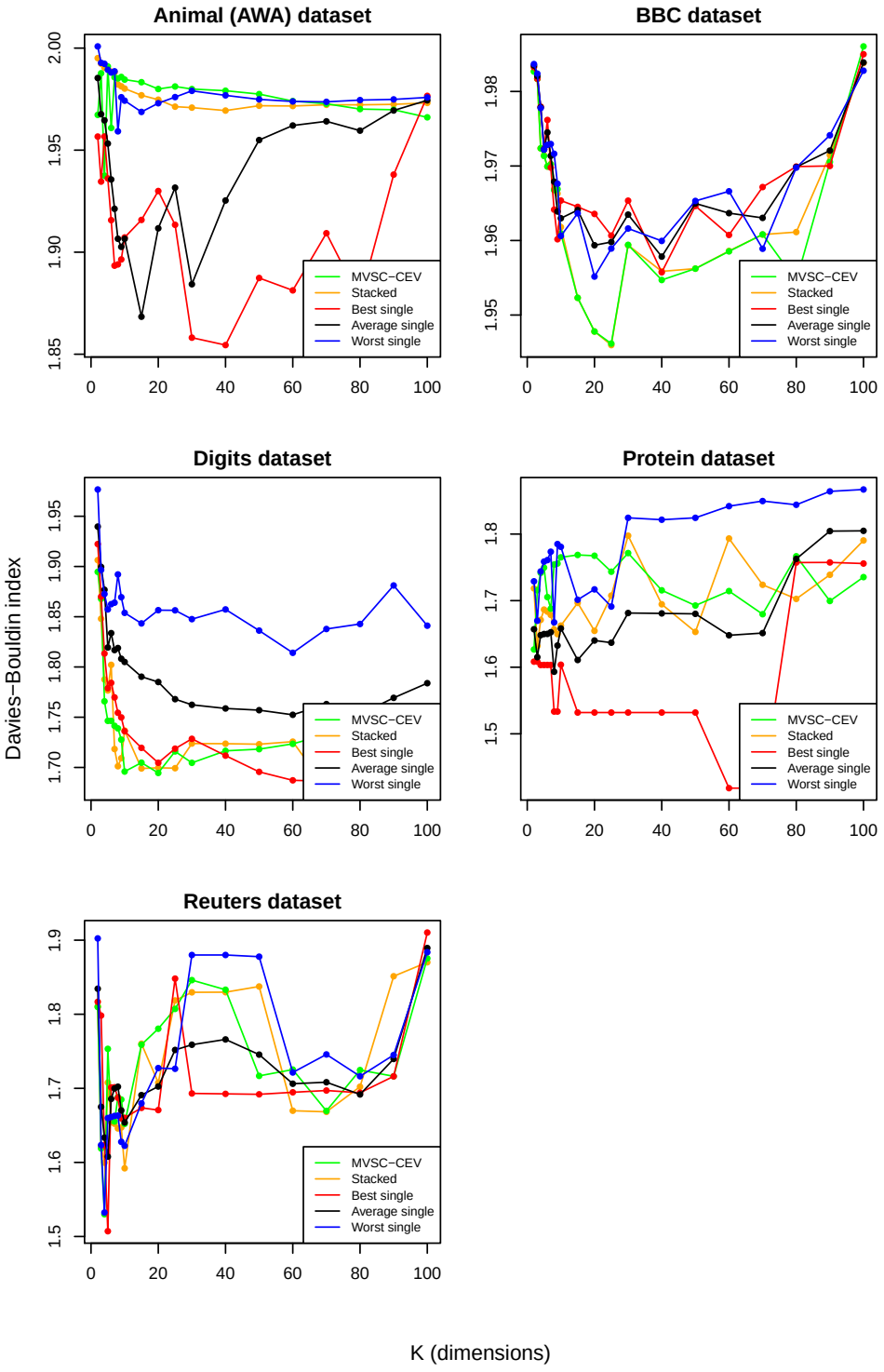


Figure 5.7: MVSC-CEV clustering evaluation with the Davies-Bouldin index (average on all input views). Less is better.

the results in the state of the art, with MV-KMeans probably using a different preprocessing (it is not specified in the paper). Again, MVSC-CEV produces the best clustering assignment according to the purity result.

Method	Digits	BBC	Reuters	AWA
CoregSC	0.822	0.887	0.552	0.580
MMSC	0.758	NA	0.390	0.585
MVC-SS	NA	NA	0.531	0.629
MV-KMeans	0.825	NA	NA	0.114
CoKmLDA	0.819	0.914	NA	NA
MFSC-MO	0.800	NA	NA	NA
MVC-PSS	0.862	NA	NA	0.325
MVSC-BG	0.844	NA	0.577	NA
MVSC-CEV	**0.946**	**0.940**	**0.619**	**0.795**

"NA" means there are no available results of the method on the data set.

Table 5.1: Clustering purity wrt. the state of the art.

Table 5.2 shows the normalized mutual information (NMI) between the clustering assignments produced by the multiview clustering methods in the state of the art and the reference class labels in the dataset. A higher value implies a more similar assignment, with NMI=1 meaning a perfect match. However NMI measures the coincidence of the cluster assignments differently from purity, as it uses the mutual information between both assignments.

The clustering of the handwritten digits dataset has, in general, high NMI values. MVSC-CEV shows the highest NMI result on this dataset. The two results on the BBC segmented news dataset are relatively similar, while the NMI result of MVSC-CEV is slightly higher. MVSC-BG [70] shows the highest results on the Reuters multilingual dataset, although the values are in general lower than in other datasets. Finally, on the animal with attributes dataset the results are mostly around 0.7 and above, with MVSC-CEV having NMI=0.833, clearly above the other methods.

5.5 Discussion

When compared to the baseline methods on the dimensionality reduction task, MVSC-CEV shows the best performance (SVM classification accuracy) on four out of six datasets. On the Reuters dataset, it performs better than stacked or average single views, which are a more realistic reference than the best single view, which probably is not known beforehand. Finally on the animal dataset the overall results of all methods are low and irregular so they may not be of much relevance to evaluate this task. On the unsupervised dimensionality

Method	Digits	BBC	Reuters	AWA
CoregSC	0.836	0.769	0.326	0.695
MMSC	0.792	NA	0.134	0.698
MVC-SS	NA	NA	NA	0.751
MV-KMeans	0.807	NA	NA	0.117
CoKmLDA	0.818	0.796	NA	NA
MFSC-MO	0.785	NA	NA	NA
MVC-PSS	0.833	NA	NA	0.213
MVSC-BG	0.832	NA	**0.357**	NA
MVSC-CEV	**0.892**	**0.826**	0.341	**0.833**

"NA" means there are no available results of the method on the data set.

Table 5.2: Clustering NMI wrt. the state of the art.

reduction evaluation metrics, MVSC-CEV shows the best performance on half the datasets and an average perfomance on the other three. This is likely caused by the fact that these metrics measure the likeliness of the low-dimensional projection to the original input data, and MVSC-CEV probably has to perform deeper transformations of the data in order to embed the different input views into a single projection space. Each single view SC only has to embed one input space, and the stacked view configuration simply adds the different input spaces into a large data matrix, without a profound change in its structure.

Regarding the clustering task, MVSC-CEV achieves the best performance on five out of six datasets, both on clustering purity and NMI. This evidences its usefulness as multiview clustering method with respect to using single view or stacked view solutions on standard spectral clustering. On the remaining dataset (Reuters), it shows better performance than both the stacked views and the average single view, which as it has been argued before is a realistic assessment of its practical advantages when processing new multiview datasets, where it is unknown if one of the views is better than the others and as a consequence all views or a concatenation of all views has to be used instead.

Compared with other multiview clustering methods in the state of the art, the method proposed in this thesis (MVSC-CEV) shows better overall clustering results. This has been tested with four well known multiview datasets and two supervised clustering evaluation metrics, where MVSC-CEV produces better clusterings in all cases except one. This suggests that MVSC-CEV may be a better multiview clustering method than the other methods in the state of the art.

MVSC-CEV computes the common eigenvectors of the Laplacian matrices,

computed in turn from the similarity matrices of the input data views. The similarity matrices of the different views of a multiview dataset are supposed to contain a considerable amount of shared information, as the similarities between samples in different views are likely to be coherent in general. The probable cause of the advantage of MVSC-CEV over other multiview clustering methods may be the fact that, when computing the common eigenvectors, this redundant information is (1) retrieved first as it has a stronger signal, and (2) is only retrieved once, leaving room for other useful information in the common subspace matrix generated. This behaviour may be making MVSC-CEV better at synthesizing the information contained in the multiple views, therefore feeding richer information into the clustering stage.

On the other hand, the qualitative advantages of MVSC-CEV over other multiview clustering methods in the state of the art are the following.

- Some methods, like CoregSC [62] and MMSC [12], can only work with two input data views. MVSC-CEV can work with any number of input views.
- CoregSC [62], MV-KMeans [11] and MVC-SS [106] can only work with feature space input data. MVSC-CEV can work with either feature or graph space input data, or with any combination of both.
- CoKmLDA [117] requires all input data to have the same dimensionality, while MVSC-CEV can work with input matrices of different dimensionality.

Chapter 6

Method comparison

6.1 Motivation

The previous Chapters included the description of the three methods proposed in this thesis, as well as the exposition of results and analysis of the experiments comparing them with their single-view counterparts and equivalent methods in the state of the art. However, no direct comparison between the three proposed methods has been done yet. The goal of this Chapter is to directly compare the three proposed methods, analyze their respective results and, if possible, suggest which one to use given a certain task.

6.2 Multiview dimensionality reduction

SVM classification (Figure 6.1). On the AWA, protein and Cora datasets, MV-MDS shows better results than the other methods. On the BBC, digits and Reuters datasets, MV-MDS and MVSC-CEV are fairly balanced, although on the Reuters dataset MVSC-CEV seems to dominate. MV-tSNE remains below on all the datasets, although its behaviour is more stable in some datasets.

Cophenetic correlation (Figure 6.2). MV-MDS shows better results on the AWA, digits, protein and Reuters datasets, in all or most dimensionalities. On the BBC dataset MV-tSNE begins with the better cophenetic correlation, although with higher K's MVSC-CEV has better results.

Area under the R_{NX} curve (Figure 6.3). Again MV-MDS shows better AUC-RNX values on the AWA, digits, protein and Reuters datasets. As in the case of the cophenetic correlation, on the BBC dataset MV-tSNE has the best results at low K's but afterwards MVSC-CEV increases its AUC-RNX and surpasses it.

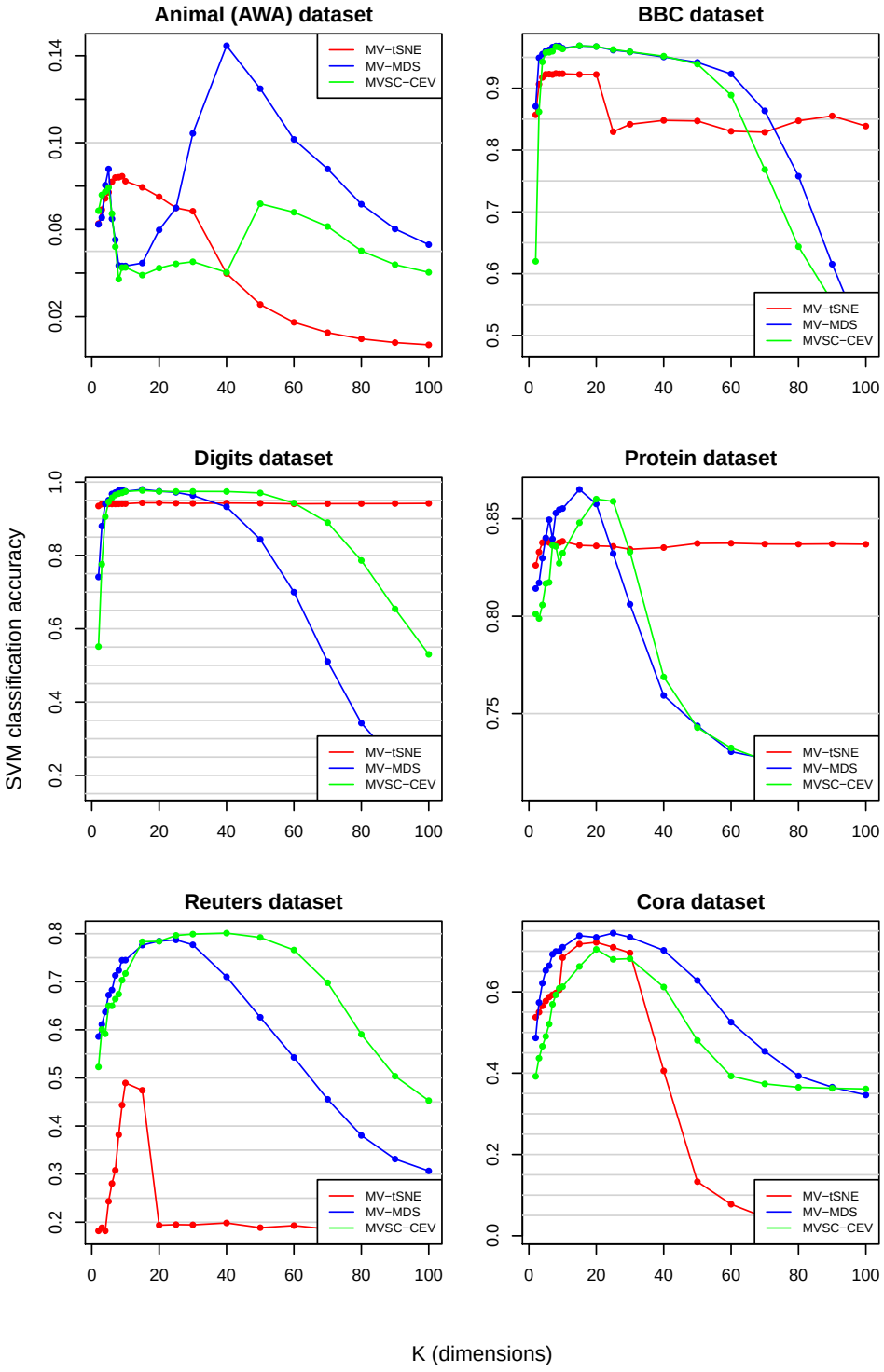


Figure 6.1: Dimensionality reduction evaluation with SVM classification.

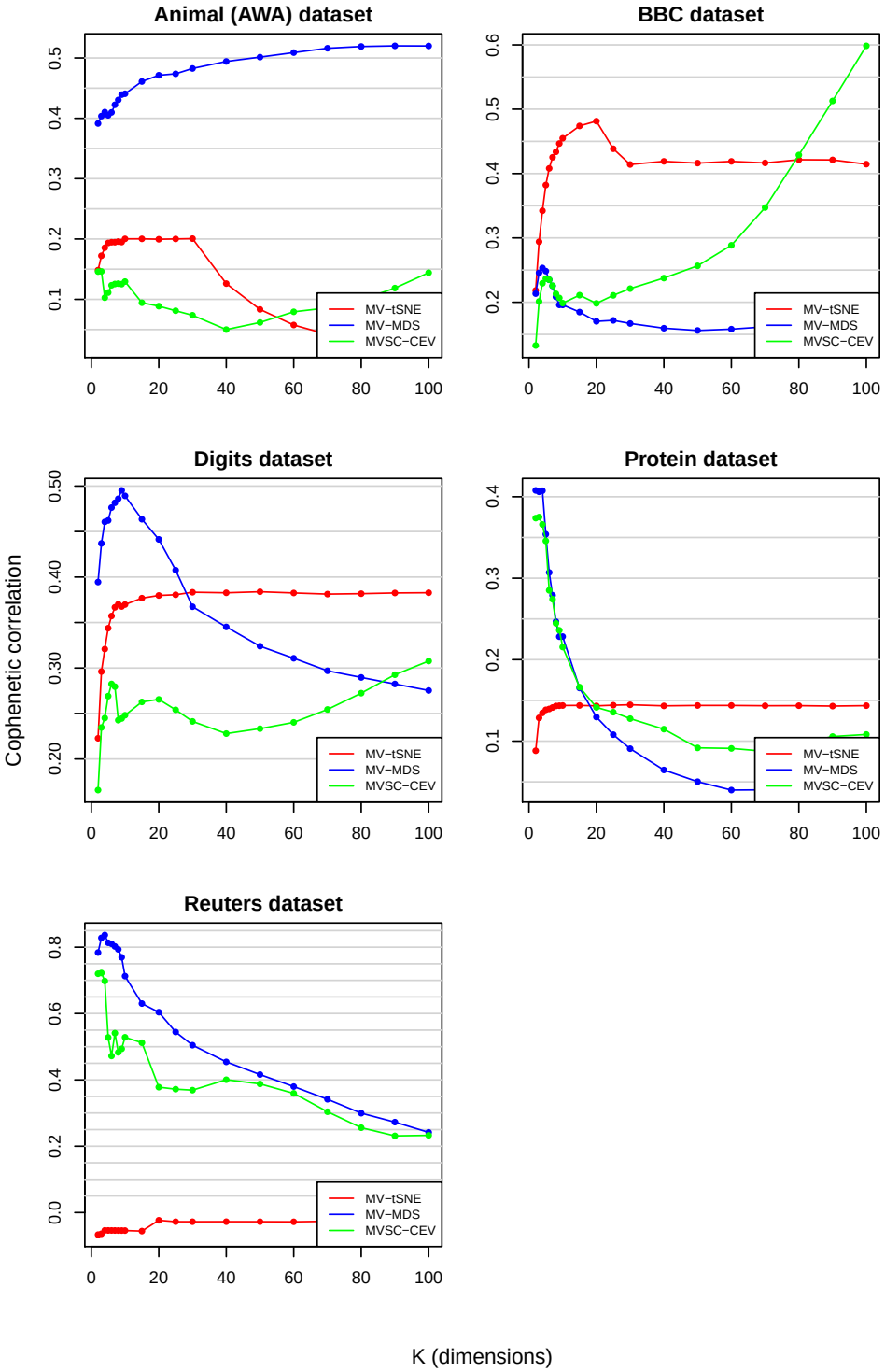


Figure 6.2: Dimensionality reduction evaluation with cophenetic correlation.

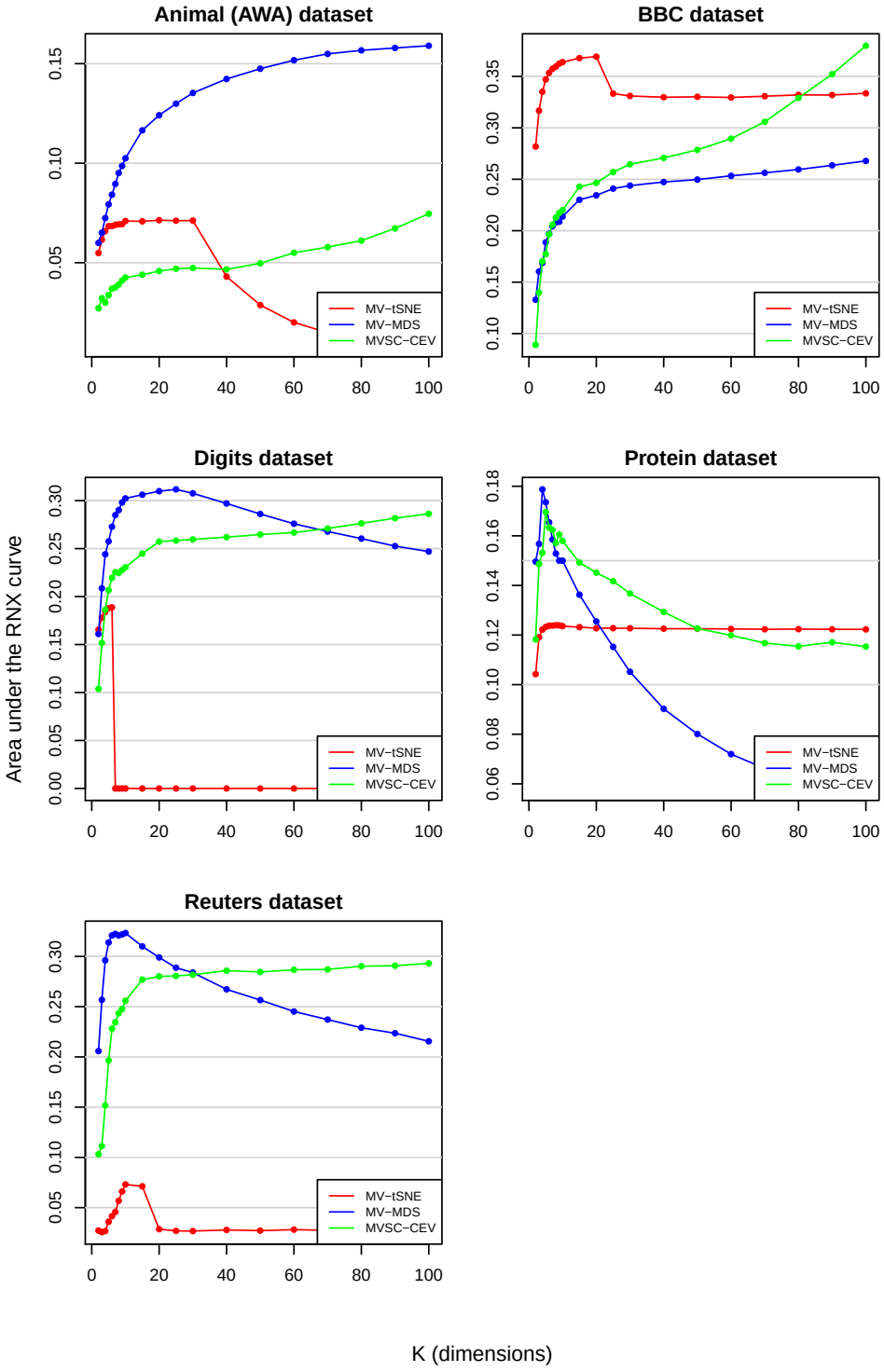


Figure 6.3: Dimensionality reduction evaluation with AUC-RNX.

6.3 Multiview clustering

Clustering purity (Figure 6.4). MVSC-CEV shows the highest purity values on AWA, BBC, digits and Reuters, although in some cases only at specific dimensionalities. On the protein dataset MV-MDS shows the best purity result. Finally, on the Cora dataset, MV-tSNE has the highest purity.

Normalized mutual information (Figure 6.5). MV-MDS and MVSC-CEV are quite balanced on AWA and digits datasets. MVSC-CEV has a slightly higher value on the BBC dataset, and clearly dominates on the Reuters dataset. The protein dataset maximum is achieved by MV-MDS, while the best NMI results on Cora are obtained by MV-tSNE.

Davies-Bouldin index (Figure 6.6). MV-tSNE gives the lowest (best) DBI values on AWA, BBC, Digits and Reuters datasets. On the protein dataset, the minimum is achieved by MV-MDS.

6.4 Discussion

According to the evaluation metrics used in the experiments, MV-MDS seems the best option for dimensionality reduction tasks, as it produces the best results on most metrics and datasets. However, it is important to analyze the behaviour of the different methods along the number of dimensions of the output space, as the specific application may require a concrete dimensionality. For example, for data visualization the dimensionality of the output space should be 2 or 3. In general, both MV-MDS and MVSC-CEV seem to produce the best results with a number of dimensions between 10 and 30; adding more dimensions, in general, makes the quality indicators drop. This is probably caused by the fact that the intrinsic dimensionality of the manifold that is being captured is of that order, and adding more dimensions simply adds noise to the representation.

Regarding the multiview clustering task, in general MVSC-CEV shows better results than the other methods, although in some cases MV-MDS performs better. It is also relevant to recall the results of the methods proposed with respect to the multiview clustering methods in the state of the art, where MVSC-CEV showed the best results on most comparisons.

As a guideline for the potential users of the methods proposed in this thesis, MV-MDS seems the best option for dimensionality reduction tasks, while MVSC-CEV would be the first option on multiview clustering tasks.

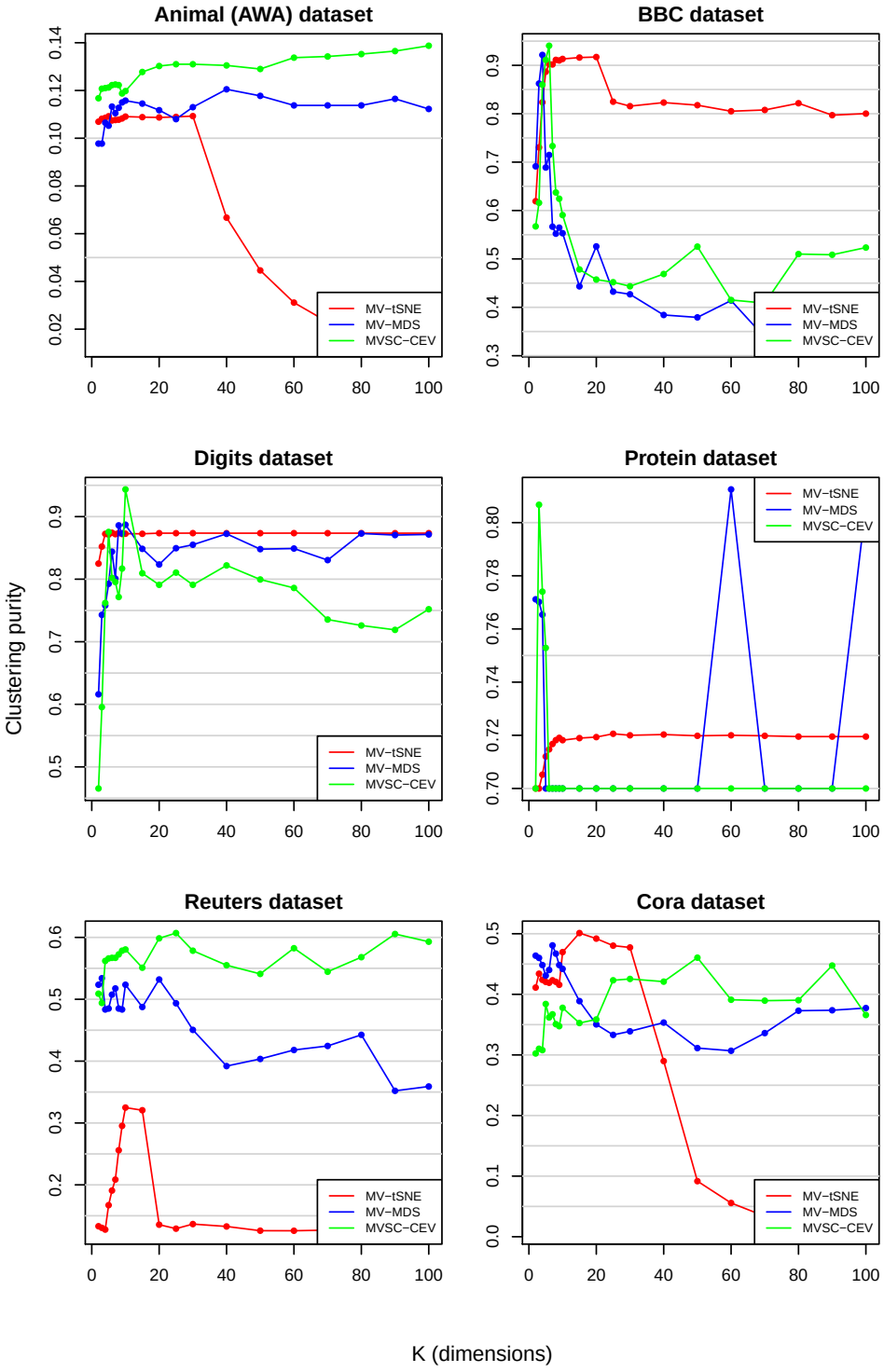


Figure 6.4: Clustering purity.

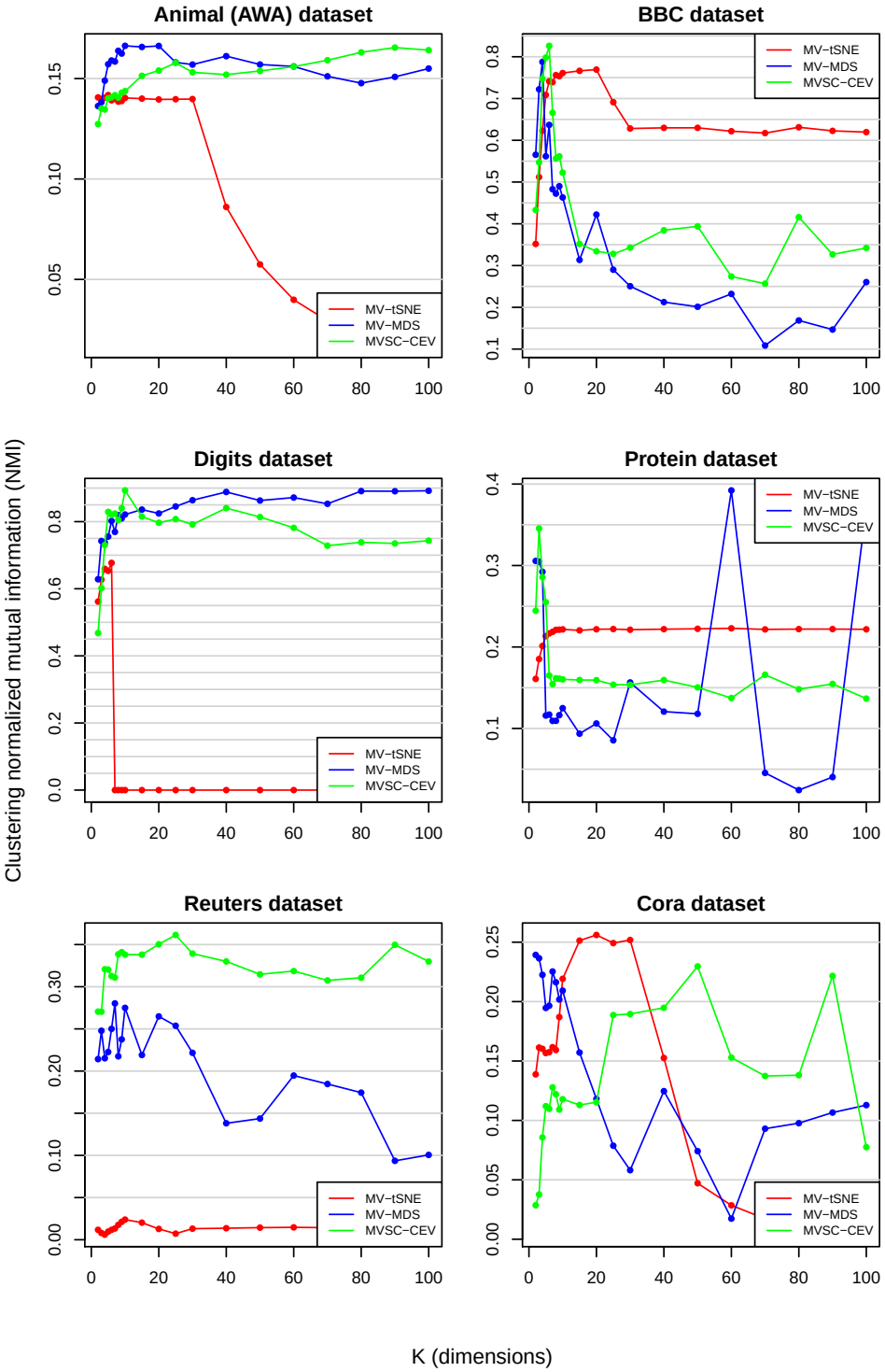


Figure 6.5: Clustering normalized mutual information.

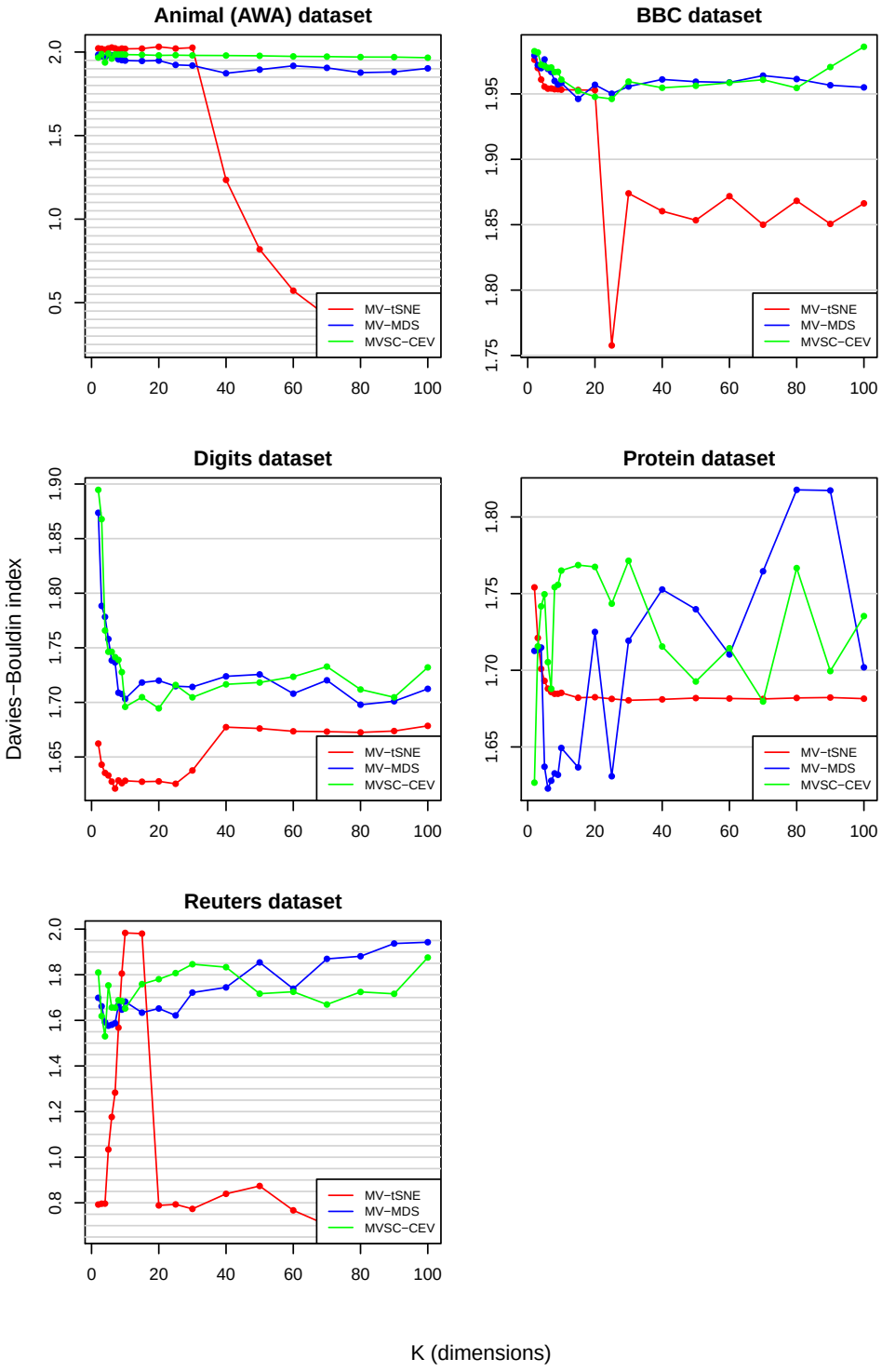


Figure 6.6: Clustering evaluation with Davies-Bouldin index (less is better).

Chapter 7

Multiview software package

7.1 Motivation

Currently there are no multiview dimensionality reduction or clustering methods freely available to the community. This fact has conditioned the experiments presented in this thesis, as the comparisons have only been possible with respect to the published results in the state of the art.

The reasons for creating a software package with the methods presented in this thesis are:

- Given that the proposed methods have been found better alternatives than their single view counterparts and that other methods in the state of the art, providing these methods to the community seems a desirable contribution.
- Let fellow researchers reproduce the experiments presented here or run their own experiments.
- Provide with sample data and documentation to let users new to multiview methods learn to use them and take advantage when possible.
- Make multiview data and methods knowledgeable to the community, so that their use spreads and the area evolves accordingly.

7.2 Package "multiview"

A software package named "multiview" [57] has been created to satisfy the objectives listed above. The language of choice is the R language for statistical computing [89], given its widespread use in the scientific community, and specially in the bioengineering and bioinformatics areas.

Package "multiview" includes the following features:

- An implementation of MV-tSNE, that accepts multiple data views and produces a single, low-dimensional projection of the original data.
- An implementation of MV-MDS, that accepts multiple data views and produces a single, low-dimensional projection of the original data.
- An implementation of MVSC-CEV, that accepts multiple data views and produces both a single, low-dimensional projection of the original data and a single clustering assignment for the input data samples.
- Synthetic multiview datasets that allow the users to get familiar with the different methods in the package.
- A user's manual describing the different methods provided, their parameters, and their return values, along with examples with the datasets provided.

Chapter 8

Conclusions and main developments

A multiview extension of the t-distributed stochastic neighbour embedding algorithm (t-SNE) has been developed (MV-tSNE). This algorithm is based on the generation of a common matrix of distance probabilities using expert consensus theory. Afterwards, the embedding optimization of t-SNE can be applied to the common matrix, therefore taking advantage of the different optimization strategies included in t-SNE. MV-tSNE shows an average performance both on dimensionality reduction tasks and on clustering tasks.

A multiview extension of the multidimensional scaling algorithm (MDS) has been developed (MV-MDS). This algorithm computes the common eigenvectors of the normalized distance matrices of the input views, therefore prioritizing the extraction of information common to all input views. In turn this is expected to produce a single high-quality, low-dimensional representation of the multiview, high-dimensional input data. MV-MDS shows better performance than the baseline methods (single view MDS) on dimensionality reduction tasks. MV-MDS also shows performance above the average on clustering tasks, with better overall results than other methods in the state of the art.

A mutiview extension of the spectral clustering algorithm has been developed (MVSC-CEV). This algorithm computes the common eigenvectors of the normalized Laplacian matrices of the similarity input matrices. This is expected to condense the most relevant bonds among samples over the different input spaces in a single representation that is both useful as a reduced dimensionality projection and as a space to apply clustering on it. Actually this algorithm also includes a multiview extension of the dimensionality reduction method Laplacian eigenmaps. MVSC-CEV produces the best clustering assignments on the vast majority of experiments run, both with respect to baseline single-view spectral clustering configurations as well as to other multiview clustering methods in the state of the art. MVSC-CEV also shows the

best performance on most datasets on dimensionality reduction tasks.

A thorough set of experiments has been designed and run on all three methods proposed in this thesis. For each of the tasks analyzed (dimensionality reduction and clustering), three evaluation metrics have been carefully chosen, giving special relevance to the fact that both supervised and unsupervised metrics be used on each task. From the extense variety of multiview datasets, six of them have been selected according to their relevance in the state of the art and their different multiview design, as well as their different area of interest.

The proposed methods have been compared with their single-view counterparts, with different configurations: either processing each view independently, either concatenating all views into a single matrix and using it as a single input view. The goal of these experiments has been to assess the usefulness of multiview methods in front of single view methods. MV-tSNE does not show any improvements over baseline methods. MVSC-CEV has proved clearly a better clustering method than the single view baseline methods. MV-MDS has proved a better dimensionality method than the single view baseline methods.

Also, the proposed methods have been compared with all the available results in the state of the art. There is no software available that implements neither the multiview dimensionality reduction methods nor the multiview clustering methods in the state of the art. For this reason, the comparison with the methods in the state of the art has been based on the published results. The dataset and experimental setups have been carefully chosen in order to reproduce as faithfully as possible those described in the literature. MVSC-CEV has consistently shown better multiview clustering performance than the methods in the state of the art.

The comparison of the three proposed methods among them confirms that MVSC-CEV is the most suited method for multiview clustering tasks, while MV-MDS is the most suited method for multiview dimensionality reduction tasks.

Finally, a software package that includes all three proposed methods along with some test data and a proper documentation has been created and released to the public. Given the good results obtained by the methods presented, making them publicly available may help the research community, specially given the lack of any public multiview methods. The language of choice is the R language, as it is one of the most relevant languages in the biomedical engineering and bioinformatics fields.

As a final conclusion, the methods and results exposed in this work show that using multiview data along with multiview methods in general improves the results in dimensionality reduction and data clustering tasks. As a consequence, using the methods proposed here can be useful when dealing with such tasks on multiview data.

Articles and conferences

Samir Kanaan-Izquierdo, Andrey Ziyatdinov, Raimon Massanet, and Alexandre Perera. Multiview approach to spectral clustering. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 1254–1257, 2012

Samir Kanaan-Izquierdo, Andrey Ziyatdinov, and Alexandre Perera-Lluna. Multiview and multifeature spectral clustering using common eigenvectors. *Pattern Recognition Letters*, 2017 *(Under review)*

Samir Kanaan-Izquierdo, Andrey Ziyatdinov, and Alexandre Perera-Lluna. Multiview: an R package for multiview pattern recognition. Technical report, B2SLAB - CREB - UPC, 2017

Samir Kanaan-Izquierdo and Alexandre Perera-Lluna. Multiview t-distributed stochastic neighbour embedding. Technical report, B2SLAB - CREB - UPC, 2017

Samir Kanaan-Izquierdo and Alexandre Perera-Lluna. Multiview multidimensional scaling. Technical report, B2SLAB - CREB - UPC, 2017

Appendix A

Results of MV-tSNE experiments

In this appendix, the detailed results of the experiments with the MV-tSNE method are presented. There is a results table for each combination of dataset and evaluation method, yielding a total of 36 tables. The methods are compared with the counterpart single view method, either applied to single views individually or to all views stacked on a single matrix. For single views, the worst, average and best views (on average) are given.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.488 ± 0.002	0.835 ± 0.409	**0.972** ± 0.002	**0.978** ± 0.002	0.967 ± 0.002
3	0.490 ± 0.001	0.838 ± 0.410	**0.973** ± 0.001	**0.978** ± 0.001	**0.970** ± 0.001
4	0.492 ± 0.001	0.838 ± 0.406	**0.971** ± 0.001	**0.976** ± 0.001	**0.972** ± 0.001
5	0.490 ± 0.001	0.839 ± 0.409	**0.971** ± 0.001	**0.980** ± 0.001	**0.973** ± 0.001
6	0.490 ± 0.001	0.838 ± 0.409	**0.971** ± 0.001	**0.978** ± 0.001	**0.973** ± 0.001
7	0.492 ± 0.001	0.839 ± 0.408	**0.971** ± 0.001	**0.979** ± 0.001	**0.973** ± 0.001
8	0.493 ± 0.001	0.839 ± 0.406	**0.972** ± 0.001	**0.978** ± 0.001	**0.973** ± 0.001
9	0.492 ± 0.001	0.839 ± 0.408	**0.971** ± 0.001	**0.978** ± 0.001	**0.972** ± 0.001
10	0.492 ± 0.001	0.839 ± 0.409	**0.972** ± 0.001	**0.981** ± 0.001	**0.973** ± 0.001
15	0.491 ± 0.000	0.840 ± 0.410	**0.972** ± 0.000	**0.982** ± 0.000	**0.974** ± 0.000
20	0.493 ± 0.001	0.840 ± 0.407	**0.973** ± 0.001	**0.980** ± 0.001	**0.973** ± 0.001
25	0.491 ± 0.001	0.840 ± 0.410	0.973 ± 0.001	**0.982** ± 0.001	**0.974** ± 0.001
30	0.493 ± 0.000	0.840 ± 0.407	**0.972** ± 0.000	**0.981** ± 0.000	**0.973** ± 0.000
40	0.491 ± 0.001	0.840 ± 0.409	**0.972** ± 0.001	**0.979** ± 0.001	**0.972** ± 0.001
50	0.492 ± 0.001	0.840 ± 0.409	**0.973** ± 0.001	**0.980** ± 0.001	**0.973** ± 0.001
60	0.492 ± 0.000	0.839 ± 0.409	**0.973** ± 0.000	**0.978** ± 0.000	**0.973** ± 0.000
70	0.492 ± 0.000	0.839 ± 0.408	**0.973** ± 0.000	**0.979** ± 0.000	**0.973** ± 0.000
80	0.492 ± 0.000	0.839 ± 0.408	**0.972** ± 0.000	**0.980** ± 0.000	**0.972** ± 0.000
90	0.490 ± 0.000	0.838 ± 0.409	**0.972** ± 0.000	**0.978** ± 0.000	**0.972** ± 0.000
100	0.489 ± 0.000	0.839 ± 0.411	**0.973** ± 0.000	**0.981** ± 0.000	**0.972** ± 0.000

Table A.1: One-vs-one SVM classification accuracy on the digits dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.206 ± 0.109	0.236 ± 0.248	**0.321** ± 0.136	0.223 ± 0.075	0.223 ± 0.075
3	0.260 ± 0.115	0.304 ± 0.268	**0.383** ± 0.148	0.296 ± 0.084	0.296 ± 0.084
4	0.280 ± 0.123	0.329 ± 0.284	**0.412** ± 0.148	0.321 ± 0.093	0.321 ± 0.093
5	0.282 ± 0.121	0.347 ± 0.303	**0.424** ± 0.155	0.344 ± 0.104	0.344 ± 0.104
6	0.283 ± 0.121	0.356 ± 0.307	**0.426** ± 0.154	0.357 ± 0.107	0.357 ± 0.107
7	0.291 ± 0.123	0.364 ± 0.310	**0.427** ± 0.154	0.367 ± 0.109	0.367 ± 0.109
8	0.293 ± 0.123	0.367 ± 0.312	**0.427** ± 0.153	0.370 ± 0.112	0.370 ± 0.112
9	0.291 ± 0.124	0.365 ± 0.314	**0.429** ± 0.156	0.368 ± 0.111	0.368 ± 0.111
10	0.292 ± 0.122	0.367 ± 0.309	**0.429** ± 0.155	0.370 ± 0.109	0.370 ± 0.109
15	0.292 ± 0.122	0.371 ± 0.320	**0.429** ± 0.156	0.376 ± 0.116	0.377 ± 0.116
20	0.295 ± 0.122	0.374 ± 0.316	**0.428** ± 0.156	0.380 ± 0.113	0.380 ± 0.113
25	0.293 ± 0.122	0.374 ± 0.321	**0.429** ± 0.156	0.381 ± 0.116	0.381 ± 0.116
30	0.293 ± 0.123	0.376 ± 0.320	**0.429** ± 0.156	0.382 ± 0.116	0.383 ± 0.115
40	0.293 ± 0.122	0.376 ± 0.318	**0.429** ± 0.155	0.382 ± 0.115	0.383 ± 0.115
50	0.293 ± 0.123	0.376 ± 0.319	**0.429** ± 0.155	0.384 ± 0.115	0.384 ± 0.115
60	0.294 ± 0.123	0.376 ± 0.320	**0.429** ± 0.156	0.383 ± 0.115	0.382 ± 0.115
70	0.293 ± 0.123	0.375 ± 0.320	**0.429** ± 0.156	0.381 ± 0.115	0.381 ± 0.115
80	0.293 ± 0.122	0.375 ± 0.319	**0.429** ± 0.156	0.382 ± 0.115	0.382 ± 0.115
90	0.293 ± 0.123	0.375 ± 0.319	**0.429** ± 0.156	0.382 ± 0.115	0.382 ± 0.115
100	0.294 ± 0.123	0.376 ± 0.319	**0.429** ± 0.156	0.383 ± 0.115	0.383 ± 0.115

Table A.2: Cophenetic correlation on the digits dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.202 ± 0.088	0.237 ± 0.183	**0.270** ± 0.118	0.238 ± 0.049	0.238 ± 0.048
3	0.213 ± 0.095	0.251 ± 0.194	**0.283** ± 0.125	0.254 ± 0.051	0.253 ± 0.051
4	0.216 ± 0.097	0.255 ± 0.198	**0.289** ± 0.127	0.257 ± 0.052	0.256 ± 0.052
5	0.215 ± 0.097	0.258 ± 0.203	**0.292** ± 0.130	0.260 ± 0.055	0.260 ± 0.055
6	0.214 ± 0.097	0.258 ± 0.204	**0.292** ± 0.130	0.260 ± 0.055	0.261 ± 0.055
7	0.216 ± 0.097	0.260 ± 0.205	**0.292** ± 0.130	0.263 ± 0.056	0.263 ± 0.056
8	0.216 ± 0.097	0.260 ± 0.206	**0.292** ± 0.130	0.263 ± 0.057	0.263 ± 0.057
9	0.216 ± 0.097	0.259 ± 0.206	**0.293** ± 0.131	0.261 ± 0.056	0.261 ± 0.056
10	0.216 ± 0.097	0.258 ± 0.204	**0.292** ± 0.131	0.261 ± 0.055	0.260 ± 0.055
15	0.216 ± 0.097	0.260 ± 0.207	**0.292** ± 0.131	0.263 ± 0.057	0.263 ± 0.057
20	0.217 ± 0.097	0.260 ± 0.207	**0.292** ± 0.131	0.263 ± 0.058	0.263 ± 0.058
25	0.216 ± 0.097	0.260 ± 0.207	**0.293** ± 0.131	0.263 ± 0.058	0.262 ± 0.058
30	0.216 ± 0.097	0.260 ± 0.208	**0.293** ± 0.131	0.263 ± 0.058	0.263 ± 0.058
40	0.216 ± 0.097	0.260 ± 0.207	**0.293** ± 0.131	0.263 ± 0.058	0.263 ± 0.058
50	0.216 ± 0.097	0.260 ± 0.207	**0.293** ± 0.131	0.263 ± 0.058	0.263 ± 0.058
60	0.216 ± 0.097	0.260 ± 0.208	**0.293** ± 0.131	0.263 ± 0.058	0.263 ± 0.058
70	0.216 ± 0.097	0.260 ± 0.208	**0.293** ± 0.131	0.263 ± 0.058	0.263 ± 0.058
80	0.216 ± 0.097	0.260 ± 0.208	**0.293** ± 0.131	0.263 ± 0.058	0.263 ± 0.058
90	0.216 ± 0.097	0.260 ± 0.207	**0.293** ± 0.131	0.263 ± 0.058	0.263 ± 0.058
100	0.216 ± 0.097	0.260 ± 0.208	**0.293** ± 0.131	0.263 ± 0.058	0.263 ± 0.058

Table A.3: Area under the curve of the R_{NX} index on the digits dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.442 ± 0.035	0.766 ± 0.373	0.825 ± 0.037	**0.929** ± 0.035	0.810 ± 0.035
3	0.458 ± 0.030	0.754 ± 0.334	0.852 ± 0.030	**0.923** ± 0.030	0.800 ± 0.030
4	0.472 ± 0.014	0.762 ± 0.324	0.872 ± 0.003	**0.926** ± 0.014	0.804 ± 0.014
5	0.453 ± 0.019	0.755 ± 0.338	0.871 ± 0.004	**0.917** ± 0.019	0.797 ± 0.019
6	0.451 ± 0.018	0.752 ± 0.340	0.874 ± 0.001	**0.912** ± 0.018	0.793 ± 0.018
7	0.452 ± 0.018	0.754 ± 0.340	0.872 ± 0.004	**0.915** ± 0.018	0.796 ± 0.018
8	0.449 ± 0.020	0.751 ± 0.340	0.874 ± 0.000	**0.908** ± 0.020	0.790 ± 0.020
9	0.453 ± 0.018	0.754 ± 0.339	0.872 ± 0.003	**0.916** ± 0.018	0.797 ± 0.018
10	0.449 ± 0.018	0.750 ± 0.340	0.873 ± 0.003	**0.907** ± 0.018	0.789 ± 0.018
15	0.456 ± 0.015	0.759 ± 0.338	0.873 ± 0.003	**0.926** ± 0.015	0.804 ± 0.015
20	0.457 ± 0.015	0.759 ± 0.337	0.874 ± 0.000	**0.925** ± 0.015	0.804 ± 0.015
25	0.458 ± 0.015	0.760 ± 0.337	0.874 ± 0.000	**0.927** ± 0.015	0.805 ± 0.015
30	0.450 ± 0.034	0.752 ± 0.345	0.874 ± 0.000	**0.911** ± 0.034	0.792 ± 0.034
40	0.412 ± 0.011	0.712 ± 0.356	**0.874** ± 0.000	0.725 ± 0.011	0.725 ± 0.011
50	0.412 ± 0.014	0.712 ± 0.357	**0.874** ± 0.000	0.725 ± 0.014	0.725 ± 0.014
60	0.416 ± 0.010	0.716 ± 0.354	**0.874** ± 0.000	0.732 ± 0.010	0.732 ± 0.010
70	0.414 ± 0.010	0.714 ± 0.355	**0.874** ± 0.000	0.727 ± 0.010	0.728 ± 0.010
80	0.411 ± 0.008	0.711 ± 0.357	**0.874** ± 0.000	0.724 ± 0.008	0.724 ± 0.008
90	0.416 ± 0.010	0.716 ± 0.354	**0.874** ± 0.000	0.731 ± 0.010	0.731 ± 0.010
100	0.412 ± 0.024	0.712 ± 0.358	**0.874** ± 0.000	0.726 ± 0.024	0.726 ± 0.024

Table A.4: Clustering purity on the digits dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.423 ± 0.017	0.722 ± 0.330	0.771 ± 0.021	**0.911** ± 0.017	0.789 ± 0.017
3	0.454 ± 0.021	0.735 ± 0.311	0.789 ± 0.017	**0.914** ± 0.021	0.793 ± 0.021
4	0.467 ± 0.011	0.745 ± 0.305	0.805 ± 0.003	**0.925** ± 0.011	0.802 ± 0.011
5	0.453 ± 0.015	0.742 ± 0.319	0.804 ± 0.004	**0.926** ± 0.015	0.802 ± 0.015
6	0.451 ± 0.013	0.741 ± 0.319	0.807 ± 0.002	**0.921** ± 0.013	0.798 ± 0.013
7	0.453 ± 0.011	0.742 ± 0.318	0.805 ± 0.004	**0.925** ± 0.011	0.801 ± 0.011
8	0.451 ± 0.016	0.739 ± 0.317	0.806 ± 0.000	**0.918** ± 0.016	0.796 ± 0.016
9	0.452 ± 0.014	0.742 ± 0.319	0.805 ± 0.003	**0.925** ± 0.014	0.801 ± 0.014
10	0.447 ± 0.014	0.739 ± 0.321	0.805 ± 0.003	**0.920** ± 0.014	0.797 ± 0.014
15	0.455 ± 0.014	0.746 ± 0.320	0.805 ± 0.003	**0.931** ± 0.013	0.807 ± 0.014
20	0.456 ± 0.011	0.746 ± 0.319	0.806 ± 0.000	**0.932** ± 0.011	0.808 ± 0.011
25	0.457 ± 0.011	0.748 ± 0.319	0.806 ± 0.000	**0.935** ± 0.011	0.810 ± 0.011
30	0.450 ± 0.026	0.741 ± 0.323	0.806 ± 0.000	**0.921** ± 0.026	0.799 ± 0.026
40	0.422 ± 0.013	0.710 ± 0.322	**0.806** ± 0.000	0.746 ± 0.013	0.746 ± 0.013
50	0.420 ± 0.017	0.709 ± 0.325	**0.806** ± 0.000	0.745 ± 0.017	0.746 ± 0.017
60	0.426 ± 0.010	0.715 ± 0.321	**0.806** ± 0.000	0.755 ± 0.010	0.754 ± 0.010
70	0.423 ± 0.010	0.712 ± 0.322	**0.806** ± 0.000	0.750 ± 0.010	0.750 ± 0.010
80	0.421 ± 0.008	0.710 ± 0.322	**0.806** ± 0.000	0.747 ± 0.008	0.746 ± 0.008
90	0.425 ± 0.010	0.714 ± 0.321	**0.806** ± 0.000	0.753 ± 0.010	0.753 ± 0.010
100	0.424 ± 0.016	0.712 ± 0.322	**0.806** ± 0.000	0.750 ± 0.016	0.751 ± 0.016

Table A.5: Clustering normalized mutual information on the digits dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	1.882 ± 0.209	1.739 ± 0.493	1.681 ± 0.209	**1.673** ± 0.209	**1.662** ± 0.209
3	1.876 ± 0.252	1.731 ± 0.569	1.678 ± 0.252	1.668 ± 0.252	**1.643** ± 0.252
4	1.876 ± 0.273	1.727 ± 0.610	1.677 ± 0.273	1.672 ± 0.273	**1.635** ± 0.273
5	1.865 ± 0.269	1.723 ± 0.600	1.678 ± 0.268	1.658 ± 0.268	**1.633** ± 0.269
6	1.858 ± 0.290	1.718 ± 0.639	1.677 ± 0.290	1.650 ± 0.290	**1.627** ± 0.290
7	1.851 ± 0.295	1.716 ± 0.646	1.678 ± 0.295	1.645 ± 0.295	**1.621** ± 0.295
8	1.855 ± 0.289	1.719 ± 0.636	1.678 ± 0.289	1.652 ± 0.288	**1.628** ± 0.289
9	1.857 ± 0.289	1.717 ± 0.635	1.677 ± 0.289	1.647 ± 0.288	**1.626** ± 0.288
10	1.853 ± 0.289	1.718 ± 0.636	1.677 ± 0.289	1.653 ± 0.289	**1.628** ± 0.290
15	1.856 ± 0.281	1.719 ± 0.622	1.678 ± 0.281	1.651 ± 0.282	**1.627** ± 0.281
20	1.858 ± 0.279	1.719 ± 0.619	1.678 ± 0.279	1.651 ± 0.280	**1.628** ± 0.280
25	1.856 ± 0.280	1.717 ± 0.621	1.678 ± 0.281	1.648 ± 0.281	**1.625** ± 0.280
30	1.867 ± 0.267	1.724 ± 0.598	1.678 ± 0.267	1.664 ± 0.267	**1.638** ± 0.267
40	1.912 ± 0.234	1.633 ± 0.876	**1.007** ± 0.233	1.702 ± 0.234	1.677 ± 0.234
50	1.913 ± 0.234	1.742 ± 0.551	**1.670** ± 0.234	1.701 ± 0.234	**1.676** ± 0.234
60	1.909 ± 0.231	1.741 ± 0.546	**1.667** ± 0.231	1.696 ± 0.231	**1.674** ± 0.231
70	1.905 ± 0.236	1.741 ± 0.552	**1.672** ± 0.236	1.699 ± 0.236	**1.673** ± 0.236
80	1.909 ± 0.240	1.742 ± 0.561	**1.666** ± 0.240	1.699 ± 0.240	**1.673** ± 0.240
90	1.907 ± 0.231	1.742 ± 0.545	**1.670** ± 0.231	1.697 ± 0.231	**1.674** ± 0.231
100	1.911 ± 0.229	1.745 ± 0.541	**1.676** ± 0.229	1.705 ± 0.228	**1.679** ± 0.229

Table A.6: Davies-Boulding index on the digits dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.165 ± 0.162	0.176 ± 0.467	**0.196** ± 0.228	0.180 ± 0.213	0.182 ± 0.223
3	0.162 ± 0.187	0.180 ± 0.501	**0.203** ± 0.260	0.172 ± 0.227	0.188 ± 0.231
4	0.147 ± 0.179	0.177 ± 0.495	**0.196** ± 0.253	0.190 ± 0.202	0.182 ± 0.223
5	0.179 ± 0.143	0.234 ± 0.395	**0.287** ± 0.195	0.263 ± 0.167	0.243 ± 0.180
6	0.214 ± 0.122	0.256 ± 0.331	**0.301** ± 0.184	0.284 ± 0.148	0.280 ± 0.155
7	0.242 ± 0.101	0.288 ± 0.292	**0.335** ± 0.153	0.304 ± 0.128	0.308 ± 0.136
8	0.311 ± 0.067	0.361 ± 0.208	**0.403** ± 0.097	0.381 ± 0.077	0.382 ± 0.084
9	0.399 ± 0.034	0.435 ± 0.116	**0.489** ± 0.047	0.453 ± 0.038	0.443 ± 0.042
10	0.455 ± 0.008	0.469 ± 0.085	**0.533** ± 0.011	0.472 ± 0.010	0.490 ± 0.010
15	0.355 ± 0.005	0.432 ± 0.113	0.479 ± 0.008	**0.494** ± 0.006	0.474 ± 0.007
20	0.166 ± 0.185	0.182 ± 0.527	**0.211** ± 0.264	0.199 ± 0.248	0.194 ± 0.237
25	0.148 ± 0.194	0.187 ± 0.507	**0.203** ± 0.278	0.191 ± 0.225	0.195 ± 0.239
30	0.165 ± 0.196	0.187 ± 0.505	**0.210** ± 0.265	0.188 ± 0.244	0.194 ± 0.238
40	0.166 ± 0.192	0.195 ± 0.511	**0.221** ± 0.259	0.178 ± 0.234	0.198 ± 0.245
50	0.140 ± 0.175	0.173 ± 0.519	**0.203** ± 0.277	0.188 ± 0.243	0.189 ± 0.239
60	0.142 ± 0.178	0.175 ± 0.465	0.190 ± 0.240	0.184 ± 0.213	**0.193** ± 0.220
70	0.147 ± 0.159	0.175 ± 0.460	**0.201** ± 0.244	0.190 ± 0.215	0.186 ± 0.218
80	0.163 ± 0.168	0.180 ± 0.464	**0.198** ± 0.244	0.190 ± 0.209	0.185 ± 0.226
90	0.139 ± 0.210	0.171 ± 0.511	0.185 ± 0.256	**0.193** ± 0.253	0.182 ± 0.229
100	0.130 ± 0.192	0.169 ± 0.512	**0.204** ± 0.261	0.186 ± 0.210	0.179 ± 0.232

Table A.7: One-vs-one SVM classification accuracy on the Reuters multilingual corpus dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	-0.075 ± 0.086	-0.066 ± 0.172	-0.055 ± 0.066	-0.064 ± 0.082	-0.067 ± 0.084
3	-0.069 ± 0.089	-0.062 ± 0.175	-0.054 ± 0.065	-0.058 ± 0.079	-0.064 ± 0.080
4	-0.057 ± 0.073	-0.049 ± 0.142	-0.039 ± 0.054	-0.056 ± 0.073	-0.054 ± 0.068
5	-0.061 ± 0.061	-0.054 ± 0.119	-0.047 ± 0.044	-0.059 ± 0.059	-0.054 ± 0.058
6	-0.062 ± 0.064	-0.052 ± 0.113	-0.042 ± 0.038	-0.053 ± 0.050	-0.054 ± 0.052
7	-0.063 ± 0.050	-0.051 ± 0.096	-0.045 ± 0.036	-0.055 ± 0.047	-0.054 ± 0.047
8	-0.059 ± 0.039	-0.052 ± 0.079	-0.046 ± 0.029	-0.053 ± 0.033	-0.054 ± 0.035
9	-0.063 ± 0.027	-0.053 ± 0.058	-0.044 ± 0.022	-0.055 ± 0.023	-0.054 ± 0.025
10	-0.064 ± 0.019	-0.052 ± 0.041	-0.047 ± 0.013	-0.052 ± 0.017	-0.054 ± 0.017
15	-0.064 ± 0.020	-0.054 ± 0.044	-0.040 ± 0.014	-0.059 ± 0.018	-0.056 ± 0.018
20	-0.027 ± 0.035	-0.021 ± 0.066	-0.017 ± 0.025	-0.024 ± 0.033	-0.023 ± 0.031
25	-0.028 ± 0.041	-0.026 ± 0.081	-0.023 ± 0.033	-0.030 ± 0.037	-0.028 ± 0.036
30	-0.032 ± 0.040	-0.026 ± 0.077	-0.022 ± 0.026	-0.026 ± 0.037	-0.028 ± 0.036
40	-0.029 ± 0.038	-0.026 ± 0.079	-0.020 ± 0.032	-0.027 ± 0.036	-0.027 ± 0.035
50	-0.031 ± 0.038	-0.027 ± 0.074	-0.024 ± 0.027	-0.028 ± 0.033	-0.028 ± 0.035
60	-0.032 ± 0.039	-0.028 ± 0.077	-0.022 ± 0.031	-0.027 ± 0.039	-0.028 ± 0.036
70	-0.029 ± 0.042	-0.025 ± 0.083	-0.020 ± 0.028	-0.026 ± 0.041	-0.026 ± 0.038
80	-0.032 ± 0.043	-0.026 ± 0.081	-0.023 ± 0.030	-0.026 ± 0.037	-0.027 ± 0.037
90	-0.031 ± 0.040	-0.025 ± 0.076	-0.020 ± 0.029	-0.025 ± 0.033	-0.027 ± 0.036
100	-0.027 ± 0.035	-0.024 ± 0.069	-0.021 ± 0.028	-0.026 ± 0.031	-0.026 ± 0.031

Table A.8: Cophenetic correlation on the Reuters multilingual corpus dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.022 ± 0.027	0.026 ± 0.073	**0.030** ± 0.036	0.028 ± 0.037	0.027 ± 0.033
3	0.021 ± 0.026	0.024 ± 0.069	**0.027** ± 0.036	0.024 ± 0.030	0.026 ± 0.032
4	0.023 ± 0.030	0.027 ± 0.073	**0.031** ± 0.038	0.026 ± 0.034	0.027 ± 0.033
5	0.030 ± 0.023	0.035 ± 0.058	**0.039** ± 0.030	0.037 ± 0.028	0.036 ± 0.027
6	0.036 ± 0.018	0.042 ± 0.051	**0.050** ± 0.026	0.042 ± 0.023	0.042 ± 0.023
7	0.036 ± 0.017	0.044 ± 0.046	0.048 ± 0.022	**0.049** ± 0.021	0.046 ± 0.021
8	0.050 ± 0.010	0.054 ± 0.032	**0.064** ± 0.017	0.054 ± 0.013	0.057 ± 0.013
9	0.053 ± 0.007	0.062 ± 0.025	**0.073** ± 0.009	0.064 ± 0.007	0.066 ± 0.008
10	0.061 ± 0.002	0.069 ± 0.018	**0.076** ± 0.003	0.071 ± 0.003	0.073 ± 0.003
15	0.056 ± 0.003	0.067 ± 0.019	**0.079** ± 0.004	0.070 ± 0.003	0.071 ± 0.003
20	0.023 ± 0.030	0.027 ± 0.077	**0.029** ± 0.037	0.027 ± 0.037	0.029 ± 0.035
25	0.021 ± 0.026	0.025 ± 0.068	**0.028** ± 0.034	**0.028** ± 0.033	0.027 ± 0.033
30	0.021 ± 0.026	0.025 ± 0.072	**0.028** ± 0.035	0.028 ± 0.033	0.027 ± 0.033
40	0.023 ± 0.029	0.026 ± 0.072	**0.030** ± 0.035	0.030 ± 0.034	0.028 ± 0.034
50	0.024 ± 0.025	0.026 ± 0.071	**0.030** ± 0.036	0.027 ± 0.036	0.027 ± 0.034
60	0.022 ± 0.030	0.026 ± 0.073	**0.032** ± 0.038	0.029 ± 0.035	0.028 ± 0.036
70	0.024 ± 0.027	0.026 ± 0.072	**0.031** ± 0.036	0.026 ± 0.034	0.027 ± 0.034
80	0.024 ± 0.027	0.028 ± 0.072	**0.034** ± 0.036	0.028 ± 0.034	0.030 ± 0.032
90	0.022 ± 0.023	0.028 ± 0.066	**0.035** ± 0.034	0.030 ± 0.028	0.029 ± 0.030
100	0.022 ± 0.023	0.026 ± 0.059	**0.031** ± 0.031	0.026 ± 0.025	0.028 ± 0.028

Table A.9: Area under the curve of the R_{NX} index on the Reuters multilingual corpus dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.102 ± 0.133	0.123 ± 0.345	0.135 ± 0.173	**0.138** ± 0.168	0.133 ± 0.163
3	0.105 ± 0.126	0.125 ± 0.347	**0.141** ± 0.171	0.124 ± 0.165	0.131 ± 0.160
4	0.103 ± 0.127	0.126 ± 0.343	**0.144** ± 0.175	0.129 ± 0.140	0.128 ± 0.156
5	0.135 ± 0.112	0.156 ± 0.282	**0.169** ± 0.141	0.155 ± 0.118	0.167 ± 0.129
6	0.160 ± 0.086	0.176 ± 0.236	**0.213** ± 0.111	0.190 ± 0.111	0.191 ± 0.112
7	0.163 ± 0.073	0.195 ± 0.217	**0.240** ± 0.111	0.225 ± 0.111	0.209 ± 0.099
8	0.213 ± 0.056	0.240 ± 0.150	**0.279** ± 0.069	0.268 ± 0.067	0.256 ± 0.066
9	0.255 ± 0.033	0.284 ± 0.111	**0.324** ± 0.042	0.281 ± 0.037	0.295 ± 0.038
10	0.242 ± 0.012	0.298 ± 0.111	**0.384** ± 0.020	0.305 ± 0.017	0.325 ± 0.017
15	0.268 ± 0.005	0.306 ± 0.066	0.326 ± 0.007	**0.340** ± 0.007	0.321 ± 0.007
20	0.109 ± 0.151	0.129 ± 0.384	**0.150** ± 0.201	0.140 ± 0.177	0.136 ± 0.166
25	0.097 ± 0.126	0.119 ± 0.343	0.132 ± 0.173	**0.142** ± 0.162	0.129 ± 0.158
30	0.109 ± 0.130	0.127 ± 0.365	**0.155** ± 0.190	0.124 ± 0.155	0.137 ± 0.167
40	0.114 ± 0.153	0.135 ± 0.408	**0.155** ± 0.208	0.129 ± 0.203	0.133 ± 0.190
50	0.116 ± 0.148	0.118 ± 0.416	0.121 ± 0.225	**0.130** ± 0.183	0.126 ± 0.199
60	0.099 ± 0.158	0.119 ± 0.437	**0.138** ± 0.202	0.135 ± 0.218	0.126 ± 0.207
70	0.095 ± 0.180	0.120 ± 0.469	**0.137** ± 0.239	0.126 ± 0.229	0.127 ± 0.216
80	0.099 ± 0.156	0.116 ± 0.453	**0.138** ± 0.240	0.118 ± 0.214	0.126 ± 0.214
90	0.091 ± 0.165	0.119 ± 0.472	**0.138** ± 0.231	0.112 ± 0.226	0.124 ± 0.212
100	0.115 ± 0.165	0.129 ± 0.454	**0.154** ± 0.233	0.140 ± 0.219	0.134 ± 0.211

Table A.10: Clustering purity on the Reuters multilingual corpus dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.010 ± 0.011	0.012 ± 0.031	**0.014** ± 0.016	0.013 ± 0.014	0.012 ± 0.015
3	0.007 ± 0.008	0.008 ± 0.021	**0.009** ± 0.010	0.007 ± 0.010	0.008 ± 0.010
4	0.005 ± 0.007	0.006 ± 0.017	**0.006** ± 0.008	0.006 ± 0.007	0.006 ± 0.008
5	0.008 ± 0.006	0.009 ± 0.015	**0.011** ± 0.007	0.009 ± 0.007	0.010 ± 0.007
6	0.010 ± 0.005	0.012 ± 0.015	**0.012** ± 0.008	0.012 ± 0.007	0.012 ± 0.007
7	0.009 ± 0.005	0.012 ± 0.015	**0.015** ± 0.008	0.013 ± 0.007	0.013 ± 0.007
8	0.012 ± 0.005	0.016 ± 0.014	**0.020** ± 0.007	0.018 ± 0.006	0.018 ± 0.006
9	0.017 ± 0.004	0.020 ± 0.013	**0.023** ± 0.006	0.022 ± 0.006	0.021 ± 0.006
10	0.020 ± 0.004	0.022 ± 0.012	**0.025** ± 0.006	0.024 ± 0.005	0.024 ± 0.005
15	0.014 ± 0.004	0.019 ± 0.012	0.022 ± 0.006	**0.023** ± 0.005	0.020 ± 0.005
20	0.010 ± 0.013	0.012 ± 0.034	**0.014** ± 0.016	0.014 ± 0.016	0.013 ± 0.016
25	0.006 ± 0.008	0.007 ± 0.020	**0.008** ± 0.010	0.007 ± 0.010	0.007 ± 0.009
30	0.010 ± 0.011	0.012 ± 0.034	**0.015** ± 0.018	0.013 ± 0.016	0.013 ± 0.016
40	0.011 ± 0.012	0.013 ± 0.033	**0.014** ± 0.017	0.013 ± 0.017	0.014 ± 0.016
50	0.011 ± 0.013	0.014 ± 0.033	**0.016** ± 0.017	0.015 ± 0.015	0.014 ± 0.015
60	0.011 ± 0.014	0.013 ± 0.035	**0.016** ± 0.017	0.015 ± 0.016	0.015 ± 0.016
70	0.011 ± 0.011	0.014 ± 0.035	**0.016** ± 0.018	0.014 ± 0.016	0.014 ± 0.016
80	0.010 ± 0.013	0.013 ± 0.036	**0.015** ± 0.018	0.014 ± 0.016	0.014 ± 0.017
90	0.011 ± 0.014	0.013 ± 0.039	**0.016** ± 0.020	0.014 ± 0.019	0.014 ± 0.018
100	0.011 ± 0.013	0.013 ± 0.035	**0.015** ± 0.019	0.014 ± 0.019	0.014 ± 0.017

Table A.11: Clustering normalized mutual information on the Reuters multilingual corpus dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.851 ± 0.745	0.742 ± 2.056	**0.620** ± 0.996	0.829 ± 0.973	0.793 ± 0.971
3	0.879 ± 0.757	0.764 ± 2.095	**0.684** ± 1.113	0.725 ± 1.002	0.796 ± 0.975
4	0.821 ± 0.771	0.741 ± 2.045	**0.666** ± 1.070	0.775 ± 0.994	0.797 ± 0.976
5	1.226 ± 0.622	0.977 ± 1.737	**0.801** ± 0.853	1.007 ± 0.804	1.034 ± 0.781
6	1.315 ± 0.542	1.127 ± 1.475	**0.928** ± 0.755	1.273 ± 0.650	1.176 ± 0.665
7	1.373 ± 0.491	1.235 ± 1.345	**1.057** ± 0.659	1.263 ± 0.597	1.283 ± 0.578
8	1.694 ± 0.263	1.492 ± 0.781	**1.328** ± 0.388	1.522 ± 0.334	1.568 ± 0.345
9	1.912 ± 0.121	1.760 ± 0.691	**1.380** ± 0.171	1.819 ± 0.163	1.805 ± 0.151
10	2.235 ± 0.004	1.857 ± 0.566	**1.595** ± 0.006	2.038 ± 0.005	1.983 ± 0.005
15	2.298 ± 0.004	1.872 ± 0.684	**1.526** ± 0.005	1.880 ± 0.005	1.980 ± 0.005
20	0.908 ± 0.739	0.753 ± 2.089	**0.642** ± 1.094	0.830 ± 0.946	0.789 ± 0.966
25	0.898 ± 0.846	0.745 ± 2.057	**0.588** ± 1.078	0.831 ± 0.968	0.793 ± 0.971
30	0.850 ± 0.780	0.753 ± 2.039	**0.715** ± 1.022	0.765 ± 0.956	0.773 ± 0.947
40	0.893 ± 0.774	0.761 ± 2.202	**0.604** ± 1.178	0.837 ± 1.025	0.839 ± 1.007
50	0.896 ± 0.741	0.837 ± 2.062	**0.750** ± 1.099	0.964 ± 0.985	0.874 ± 0.968
60	0.802 ± 0.738	0.703 ± 2.024	**0.547** ± 1.149	0.784 ± 0.959	0.767 ± 0.956
70	0.785 ± 0.674	0.671 ± 1.986	**0.578** ± 1.045	0.721 ± 0.894	0.703 ± 0.950
80	0.711 ± 0.859	0.621 ± 2.075	**0.486** ± 0.976	0.713 ± 0.953	0.674 ± 0.967
90	0.798 ± 0.854	0.708 ± 2.088	**0.604** ± 1.047	0.746 ± 0.992	0.737 ± 0.963
100	0.872 ± 0.708	0.741 ± 1.962	**0.584** ± 0.976	0.781 ± 0.878	0.791 ± 0.918

Table A.12: Davies-Boulding index on the Reuters multilingual corpus dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.875 ± 0.005	0.890 ± 0.021	**0.904** ± 0.004	0.859 ± 0.014	0.857 ± 0.014
3	0.871 ± 0.005	0.887 ± 0.023	**0.903** ± 0.004	**0.905** ± 0.007	**0.907** ± 0.007
4	0.872 ± 0.002	0.889 ± 0.024	0.905 ± 0.002	**0.920** ± 0.004	**0.917** ± 0.004
5	0.871 ± 0.002	0.889 ± 0.026	0.907 ± 0.001	**0.929** ± 0.002	**0.922** ± 0.002
6	0.869 ± 0.003	0.888 ± 0.027	0.907 ± 0.001	**0.921** ± 0.002	**0.923** ± 0.002
7	0.869 ± 0.004	0.887 ± 0.027	0.906 ± 0.002	**0.918** ± 0.002	**0.922** ± 0.002
8	0.868 ± 0.003	0.887 ± 0.027	0.906 ± 0.001	**0.931** ± 0.002	**0.924** ± 0.002
9	0.869 ± 0.002	0.888 ± 0.026	0.906 ± 0.001	**0.925** ± 0.002	**0.923** ± 0.002
10	0.868 ± 0.002	0.887 ± 0.028	0.907 ± 0.001	**0.925** ± 0.002	**0.923** ± 0.002
15	0.873 ± 0.002	0.890 ± 0.025	0.907 ± 0.001	**0.924** ± 0.002	**0.922** ± 0.002
20	0.874 ± 0.001	0.890 ± 0.023	0.906 ± 0.002	**0.924** ± 0.001	**0.922** ± 0.001
25	0.874 ± 0.003	0.891 ± 0.023	**0.907** ± 0.001	0.828 ± 0.277	0.830 ± 0.277
30	0.874 ± 0.001	0.890 ± 0.023	**0.906** ± 0.001	0.837 ± 0.001	0.842 ± 0.001
40	0.875 ± 0.002	0.890 ± 0.022	**0.905** ± 0.001	0.848 ± 0.001	0.848 ± 0.001
50	0.875 ± 0.001	0.889 ± 0.021	**0.904** ± 0.001	0.848 ± 0.001	0.847 ± 0.001
60	0.874 ± 0.001	0.888 ± 0.021	**0.903** ± 0.001	0.831 ± 0.001	0.830 ± 0.001
70	0.874 ± 0.001	0.888 ± 0.021	**0.903** ± 0.001	0.830 ± 0.001	0.829 ± 0.001
80	0.875 ± 0.001	0.889 ± 0.020	**0.902** ± 0.001	0.848 ± 0.001	0.847 ± 0.001
90	0.875 ± 0.001	0.888 ± 0.019	**0.901** ± 0.001	0.854 ± 0.001	0.855 ± 0.001
100	0.876 ± 0.001	0.888 ± 0.018	**0.901** ± 0.001	0.833 ± 0.001	0.839 ± 0.001

Table A.13: One-vs-one SVM classification accuracy on the BBC segmented news dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.288 ± 0.010	**0.290** ± 0.012	**0.292** ± 0.006	0.218 ± 0.019	0.218 ± 0.019
3	0.327 ± 0.013	0.331 ± 0.015	**0.336** ± 0.003	0.293 ± 0.012	0.294 ± 0.012
4	0.356 ± 0.016	0.364 ± 0.020	**0.372** ± 0.001	0.342 ± 0.010	0.342 ± 0.010
5	0.376 ± 0.019	0.386 ± 0.024	**0.397** ± 0.002	0.384 ± 0.004	0.382 ± 0.004
6	0.390 ± 0.020	0.402 ± 0.026	**0.413** ± 0.002	**0.409** ± 0.005	0.408 ± 0.005
7	0.399 ± 0.021	0.410 ± 0.026	0.422 ± 0.001	**0.428** ± 0.004	**0.425** ± 0.004
8	0.406 ± 0.022	0.417 ± 0.027	0.428 ± 0.002	**0.435** ± 0.005	**0.434** ± 0.005
9	0.409 ± 0.022	0.421 ± 0.027	0.432 ± 0.001	**0.445** ± 0.004	**0.447** ± 0.004
10	0.412 ± 0.022	0.423 ± 0.027	0.434 ± 0.001	**0.454** ± 0.003	**0.455** ± 0.003
15	0.421 ± 0.023	0.431 ± 0.027	0.441 ± 0.001	**0.470** ± 0.003	**0.474** ± 0.003
20	0.425 ± 0.023	0.434 ± 0.027	0.444 ± 0.001	**0.479** ± 0.003	**0.481** ± 0.003
25	0.426 ± 0.023	0.435 ± 0.027	**0.445** ± 0.001	0.436 ± 0.147	0.439 ± 0.146
30	0.427 ± 0.023	0.437 ± 0.027	**0.446** ± 0.001	0.415 ± 0.012	0.414 ± 0.012
40	0.429 ± 0.023	0.438 ± 0.027	**0.447** ± 0.001	0.420 ± 0.012	0.419 ± 0.012
50	0.430 ± 0.024	0.439 ± 0.027	**0.448** ± 0.001	0.419 ± 0.011	0.416 ± 0.011
60	0.430 ± 0.024	0.439 ± 0.027	**0.448** ± 0.001	0.417 ± 0.012	0.419 ± 0.012
70	0.431 ± 0.024	0.439 ± 0.027	**0.448** ± 0.001	0.416 ± 0.012	0.417 ± 0.012
80	0.431 ± 0.024	0.440 ± 0.027	**0.448** ± 0.001	0.425 ± 0.012	0.422 ± 0.012
90	0.431 ± 0.024	0.440 ± 0.027	**0.449** ± 0.001	0.422 ± 0.012	0.421 ± 0.012
100	0.432 ± 0.024	0.440 ± 0.027	**0.449** ± 0.001	0.415 ± 0.012	0.415 ± 0.012

Table A.14: Cophenetic correlation on the BBC segmented news dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.309 ± 0.085	0.315 ± 0.085	**0.322** ± 0.003	0.283 ± 0.006	0.282 ± 0.006
3	0.323 ± 0.090	0.330 ± 0.091	**0.337** ± 0.002	0.317 ± 0.005	0.317 ± 0.005
4	0.330 ± 0.092	0.337 ± 0.093	**0.344** ± 0.001	0.334 ± 0.004	0.335 ± 0.004
5	0.333 ± 0.093	0.341 ± 0.094	**0.348** ± 0.001	**0.347** ± 0.002	**0.347** ± 0.002
6	0.335 ± 0.094	0.343 ± 0.095	**0.351** ± 0.001	**0.353** ± 0.002	**0.353** ± 0.002
7	0.337 ± 0.094	0.344 ± 0.095	0.352 ± 0.001	**0.357** ± 0.001	**0.357** ± 0.001
8	0.338 ± 0.094	0.345 ± 0.095	0.353 ± 0.001	**0.361** ± 0.002	**0.359** ± 0.002
9	0.338 ± 0.095	0.346 ± 0.095	0.354 ± 0.001	**0.365** ± 0.001	**0.362** ± 0.001
10	0.339 ± 0.095	0.346 ± 0.095	0.354 ± 0.001	**0.361** ± 0.001	**0.364** ± 0.001
15	0.340 ± 0.095	0.348 ± 0.095	0.355 ± 0.001	**0.367** ± 0.001	**0.368** ± 0.001
20	0.341 ± 0.095	0.348 ± 0.095	0.356 ± 0.001	**0.372** ± 0.001	**0.369** ± 0.001
25	0.341 ± 0.095	0.348 ± 0.096	**0.356** ± 0.001	0.334 ± 0.112	0.333 ± 0.111
30	0.341 ± 0.095	0.348 ± 0.096	**0.356** ± 0.001	0.329 ± 0.046	0.331 ± 0.046
40	0.341 ± 0.095	0.349 ± 0.096	**0.356** ± 0.001	0.332 ± 0.045	0.330 ± 0.045
50	0.341 ± 0.095	0.349 ± 0.096	**0.356** ± 0.001	0.330 ± 0.046	0.330 ± 0.046
60	0.342 ± 0.095	0.349 ± 0.096	**0.356** ± 0.001	0.331 ± 0.046	0.329 ± 0.046
70	0.342 ± 0.095	0.349 ± 0.095	**0.356** ± 0.001	0.331 ± 0.046	0.331 ± 0.046
80	0.342 ± 0.095	0.349 ± 0.096	**0.356** ± 0.001	0.332 ± 0.045	0.332 ± 0.046
90	0.342 ± 0.095	0.349 ± 0.096	**0.356** ± 0.001	0.334 ± 0.045	0.332 ± 0.046
100	0.342 ± 0.095	0.349 ± 0.096	**0.356** ± 0.001	0.329 ± 0.046	0.334 ± 0.046

Table A.15: Area under the curve of the R_{NX} index on the BBC segmented news dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	**0.812** ± 0.021	**0.810** ± 0.041	**0.808** ± 0.035	0.616 ± 0.078	0.619 ± 0.078
3	**0.818** ± 0.063	**0.820** ± 0.102	**0.822** ± 0.081	0.733 ± 0.058	0.730 ± 0.058
4	0.834 ± 0.029	0.847 ± 0.085	**0.859** ± 0.078	0.825 ± 0.058	0.824 ± 0.058
5	0.845 ± 0.004	0.866 ± 0.030	**0.887** ± 0.006	**0.886** ± 0.010	**0.887** ± 0.010
6	0.844 ± 0.007	0.867 ± 0.033	0.889 ± 0.004	**0.900** ± 0.012	**0.902** ± 0.012
7	0.832 ± 0.043	0.854 ± 0.070	0.875 ± 0.047	**0.901** ± 0.007	**0.902** ± 0.007
8	0.847 ± 0.004	0.869 ± 0.031	0.891 ± 0.002	**0.914** ± 0.005	**0.911** ± 0.005
9	0.846 ± 0.002	0.869 ± 0.032	0.891 ± 0.004	**0.911** ± 0.004	**0.910** ± 0.004
10	0.846 ± 0.002	0.869 ± 0.032	0.892 ± 0.001	**0.912** ± 0.003	**0.913** ± 0.003
15	0.847 ± 0.003	0.847 ± 0.070	0.846 ± 0.070	**0.914** ± 0.002	**0.916** ± 0.002
20	0.847 ± 0.002	0.853 ± 0.064	0.859 ± 0.063	**0.919** ± 0.002	**0.917** ± 0.002
25	0.847 ± 0.002	**0.854** ± 0.062	**0.860** ± 0.062	0.832 ± 0.276	0.825 ± 0.275
30	0.846 ± 0.002	0.860 ± 0.051	**0.875** ± 0.047	0.817 ± 0.023	0.816 ± 0.023
40	0.846 ± 0.001	0.868 ± 0.032	**0.891** ± 0.001	0.812 ± 0.001	0.823 ± 0.001
50	0.847 ± 0.001	0.861 ± 0.050	**0.875** ± 0.046	0.813 ± 0.022	0.818 ± 0.022
60	0.847 ± 0.001	**0.853** ± 0.064	**0.858** ± 0.064	0.805 ± 0.031	0.805 ± 0.031
70	**0.846** ± 0.001	**0.844** ± 0.072	**0.842** ± 0.072	0.805 ± 0.035	0.808 ± 0.035
80	0.848 ± 0.001	0.869 ± 0.030	**0.890** ± 0.002	0.826 ± 0.001	0.822 ± 0.001
90	**0.846** ± 0.001	**0.845** ± 0.072	**0.843** ± 0.072	0.798 ± 0.035	0.797 ± 0.034
100	**0.846** ± 0.002	**0.845** ± 0.070	**0.844** ± 0.070	0.804 ± 0.034	0.800 ± 0.034

Table A.16: Clustering purity on the BBC segmented news dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.580 ± 0.028	0.588 ± 0.056	**0.596** ± 0.047	0.350 ± 0.067	0.352 ± 0.068
3	0.591 ± 0.057	0.613 ± 0.085	**0.635** ± 0.055	0.511 ± 0.060	0.512 ± 0.060
4	0.607 ± 0.028	0.645 ± 0.087	**0.683** ± 0.062	0.624 ± 0.068	0.623 ± 0.068
5	0.620 ± 0.007	0.663 ± 0.062	**0.706** ± 0.012	**0.711** ± 0.018	**0.709** ± 0.018
6	0.617 ± 0.013	0.664 ± 0.068	0.711 ± 0.007	**0.744** ± 0.017	**0.741** ± 0.018
7	0.609 ± 0.036	0.656 ± 0.083	0.703 ± 0.033	**0.743** ± 0.011	**0.739** ± 0.011
8	0.622 ± 0.007	0.667 ± 0.065	0.712 ± 0.005	**0.757** ± 0.011	**0.756** ± 0.011
9	0.622 ± 0.004	0.668 ± 0.067	0.715 ± 0.008	**0.759** ± 0.007	**0.753** ± 0.007
10	0.621 ± 0.004	0.668 ± 0.067	0.715 ± 0.003	**0.762** ± 0.007	**0.761** ± 0.007
15	0.622 ± 0.006	0.655 ± 0.063	0.688 ± 0.042	**0.770** ± 0.005	**0.766** ± 0.005
20	0.623 ± 0.004	0.657 ± 0.065	0.692 ± 0.042	**0.767** ± 0.006	**0.769** ± 0.006
25	0.623 ± 0.004	0.659 ± 0.064	**0.694** ± 0.040	**0.693** ± 0.230	**0.691** ± 0.230
30	0.621 ± 0.003	0.661 ± 0.066	**0.701** ± 0.033	0.630 ± 0.017	0.628 ± 0.017
40	0.620 ± 0.002	0.666 ± 0.066	**0.713** ± 0.002	0.634 ± 0.002	0.630 ± 0.002
50	0.622 ± 0.003	0.662 ± 0.065	**0.702** ± 0.031	0.630 ± 0.016	0.630 ± 0.016
60	0.623 ± 0.002	0.657 ± 0.064	**0.691** ± 0.042	0.615 ± 0.021	0.622 ± 0.021
70	0.621 ± 0.002	0.649 ± 0.064	**0.678** ± 0.049	0.615 ± 0.025	0.617 ± 0.025
80	0.623 ± 0.002	0.667 ± 0.063	**0.712** ± 0.003	0.631 ± 0.002	0.631 ± 0.002
90	0.621 ± 0.002	0.650 ± 0.065	**0.678** ± 0.051	0.624 ± 0.025	0.623 ± 0.026
100	0.621 ± 0.003	0.651 ± 0.063	**0.682** ± 0.046	0.621 ± 0.023	0.620 ± 0.023

Table A.17: Clustering normalized mutual information on the BBC segmented news dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	**1.961** ± 0.004	**1.962** ± 0.005	**1.962** ± 0.002	1.984 ± 0.004	**1.976** ± 0.004
3	**1.960** ± 0.005	**1.960** ± 0.007	**1.960** ± 0.005	**1.959** ± 0.005	**1.970** ± 0.005
4	**1.959** ± 0.005	**1.958** ± 0.006	**1.957** ± 0.003	**1.972** ± 0.005	**1.961** ± 0.005
5	**1.958** ± 0.001	**1.957** ± 0.002	**1.956** ± 0.001	**1.962** ± 0.001	**1.956** ± 0.001
6	**1.958** ± 0.000	**1.957** ± 0.002	**1.956** ± 0.001	**1.973** ± 0.001	**1.954** ± 0.001
7	**1.960** ± 0.005	**1.958** ± 0.007	**1.957** ± 0.004	**1.949** ± 0.001	**1.954** ± 0.001
8	**1.958** ± 0.000	**1.957** ± 0.002	**1.955** ± 0.001	**1.955** ± 0.001	**1.954** ± 0.001
9	**1.958** ± 0.001	**1.957** ± 0.002	**1.955** ± 0.001	**1.966** ± 0.001	**1.954** ± 0.001
10	1.958 ± 0.000	1.957 ± 0.002	1.955 ± 0.000	**1.936** ± 0.001	**1.953** ± 0.001
15	**1.958** ± 0.005	**1.959** ± 0.005	**1.959** ± 0.001	**1.953** ± 0.001	**1.953** ± 0.001
20	**1.958** ± 0.003	**1.957** ± 0.003	**1.956** ± 0.001	**1.950** ± 0.001	**1.953** ± 0.001
25	1.958 ± 0.004	1.958 ± 0.004	1.958 ± 0.001	**1.761** ± 0.584	**1.758** ± 0.586
30	1.959 ± 0.003	1.958 ± 0.003	1.957 ± 0.001	**1.868** ± 0.002	**1.874** ± 0.002
40	1.959 ± 0.000	1.957 ± 0.002	1.956 ± 0.001	**1.866** ± 0.000	**1.860** ± 0.000
50	1.959 ± 0.003	1.958 ± 0.003	1.957 ± 0.001	**1.855** ± 0.002	**1.853** ± 0.002
60	1.958 ± 0.003	1.957 ± 0.003	1.956 ± 0.001	**1.857** ± 0.002	**1.872** ± 0.002
70	1.959 ± 0.004	1.958 ± 0.004	1.958 ± 0.001	**1.857** ± 0.002	**1.850** ± 0.002
80	1.959 ± 0.000	1.957 ± 0.002	1.956 ± 0.001	**1.855** ± 0.000	**1.868** ± 0.000
90	1.959 ± 0.004	1.958 ± 0.004	1.958 ± 0.001	**1.841** ± 0.002	**1.851** ± 0.002
100	1.959 ± 0.004	1.958 ± 0.004	1.958 ± 0.001	**1.874** ± 0.002	**1.866** ± 0.002

Table A.18: Davies-Boulding index on the BBC segmented news dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.062 ± 0.003	0.062 ± 0.011	0.062 ± 0.005	0.061 ± 0.001	**0.063** ± 0.003
3	0.068 ± 0.001	**0.068** ± 0.004	**0.069** ± 0.002	0.067 ± 0.001	**0.069** ± 0.001
4	0.073 ± 0.001	0.075 ± 0.004	**0.076** ± 0.001	0.070 ± 0.001	0.074 ± 0.001
5	0.076 ± 0.001	0.078 ± 0.003	**0.079** ± 0.001	0.072 ± 0.000	0.077 ± 0.001
6	0.080 ± 0.001	**0.082** ± 0.004	**0.082** ± 0.001	0.078 ± 0.001	**0.082** ± 0.001
7	0.082 ± 0.001	**0.084** ± 0.004	**0.085** ± 0.001	0.080 ± 0.001	**0.084** ± 0.001
8	0.083 ± 0.001	**0.084** ± 0.003	**0.085** ± 0.001	0.080 ± 0.001	**0.084** ± 0.001
9	0.083 ± 0.001	**0.085** ± 0.003	**0.085** ± 0.001	0.081 ± 0.001	**0.085** ± 0.001
10	0.081 ± 0.001	0.084 ± 0.005	**0.086** ± 0.001	0.076 ± 0.001	0.082 ± 0.001
15	0.078 ± 0.001	0.083 ± 0.006	**0.085** ± 0.001	0.071 ± 0.001	0.079 ± 0.001
20	0.074 ± 0.001	0.077 ± 0.006	**0.079** ± 0.002	0.068 ± 0.001	0.075 ± 0.001
25	0.069 ± 0.001	0.071 ± 0.002	**0.071** ± 0.001	0.067 ± 0.001	0.070 ± 0.001
30	0.067 ± 0.000	0.069 ± 0.002	**0.070** ± 0.000	0.064 ± 0.000	0.068 ± 0.000
40	0.038 ± 0.001	0.021 ± 0.022	0.013 ± 0.003	**0.063** ± 0.000	0.040 ± 0.001
50	0.025 ± 0.002	0.017 ± 0.011	0.013 ± 0.002	**0.036** ± 0.003	0.026 ± 0.003
60	0.017 ± 0.001	0.014 ± 0.011	0.013 ± 0.001	**0.021** ± 0.002	0.017 ± 0.001
70	0.012 ± 0.007	**0.012** ± 0.013	**0.013** ± 0.001	0.012 ± 0.001	**0.013** ± 0.008
80	0.010 ± 0.005	0.011 ± 0.012	**0.012** ± 0.008	0.007 ± 0.001	0.010 ± 0.005
90	0.008 ± 0.004	0.011 ± 0.013	**0.012** ± 0.006	0.004 ± 0.007	0.008 ± 0.004
100	0.007 ± 0.003	0.010 ± 0.010	**0.012** ± 0.004	0.002 ± 0.005	0.007 ± 0.003

Table A.19: One-vs-one SVM classification accuracy on the animal with attributes (AWA) dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.145 ± 0.115	0.145 ± 0.278	0.145 ± 0.110	0.146 ± 0.105	**0.148** ± 0.112
3	0.165 ± 0.135	0.166 ± 0.327	0.168 ± 0.130	**0.172** ± 0.125	**0.172** ± 0.132
4	0.179 ± 0.149	0.180 ± 0.360	0.182 ± 0.143	**0.184** ± 0.138	**0.186** ± 0.146
5	0.184 ± 0.154	0.186 ± 0.374	0.190 ± 0.149	**0.195** ± 0.144	**0.193** ± 0.152
6	0.187 ± 0.157	0.188 ± 0.380	0.190 ± 0.151	**0.194** ± 0.145	**0.195** ± 0.154
7	0.187 ± 0.158	0.188 ± 0.382	0.191 ± 0.152	**0.194** ± 0.146	**0.195** ± 0.155
8	0.188 ± 0.159	0.190 ± 0.384	0.192 ± 0.153	**0.195** ± 0.147	**0.196** ± 0.156
9	0.188 ± 0.159	0.189 ± 0.385	0.192 ± 0.153	**0.195** ± 0.147	**0.195** ± 0.156
10	0.188 ± 0.159	0.191 ± 0.388	0.196 ± 0.156	**0.204** ± 0.153	0.200 ± 0.159
15	0.188 ± 0.159	0.190 ± 0.387	0.196 ± 0.156	**0.204** ± 0.154	0.200 ± 0.159
20	0.188 ± 0.159	0.191 ± 0.388	0.196 ± 0.157	**0.204** ± 0.154	0.200 ± 0.159
25	0.188 ± 0.159	0.191 ± 0.388	0.196 ± 0.157	**0.204** ± 0.154	0.200 ± 0.160
30	0.188 ± 0.159	0.191 ± 0.388	0.196 ± 0.156	**0.204** ± 0.154	0.201 ± 0.160
40	0.038 ± 0.104	0.064 ± 0.283	0.121 ± 0.129	**0.204** ± 0.154	0.126 ± 0.132
50	0.038 ± 0.104	0.051 ± 0.277	0.080 ± 0.130	**0.123** ± 0.156	0.083 ± 0.132
60	0.038 ± 0.104	0.043 ± 0.277	0.056 ± 0.131	**0.074** ± 0.158	0.058 ± 0.135
70	0.038 ± 0.104	0.039 ± 0.277	0.041 ± 0.132	**0.044** ± 0.160	0.042 ± 0.135
80	**0.037** ± 0.104	0.036 ± 0.278	0.032 ± 0.133	0.027 ± 0.162	0.033 ± 0.137
90	**0.037** ± 0.104	0.034 ± 0.279	0.027 ± 0.134	0.016 ± 0.164	0.027 ± 0.137
100	**0.037** ± 0.104	0.033 ± 0.279	0.023 ± 0.134	0.010 ± 0.166	0.024 ± 0.138

Table A.20: Cophenetic correlation on the animal with attributes (AWA) dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.054 ± 0.067	0.056 ± 0.185	**0.057** ± 0.079	0.051 ± 0.055	0.055 ± 0.068
3	0.060 ± 0.076	0.062 ± 0.210	**0.063** ± 0.089	0.057 ± 0.063	0.062 ± 0.078
4	0.065 ± 0.081	0.067 ± 0.224	**0.067** ± 0.095	0.061 ± 0.067	0.066 ± 0.083
5	0.067 ± 0.083	**0.069** ± 0.230	**0.069** ± 0.098	0.065 ± 0.068	0.068 ± 0.085
6	0.067 ± 0.084	0.069 ± 0.233	**0.070** ± 0.099	0.064 ± 0.069	0.068 ± 0.085
7	0.068 ± 0.085	0.070 ± 0.234	**0.071** ± 0.099	0.065 ± 0.070	0.069 ± 0.086
8	0.068 ± 0.085	0.070 ± 0.235	**0.071** ± 0.100	0.065 ± 0.070	0.069 ± 0.087
9	0.068 ± 0.085	0.070 ± 0.236	**0.071** ± 0.100	0.065 ± 0.070	0.069 ± 0.087
10	0.070 ± 0.087	**0.071** ± 0.236	**0.071** ± 0.100	0.068 ± 0.072	**0.071** ± 0.088
15	0.070 ± 0.087	**0.071** ± 0.237	**0.071** ± 0.100	0.068 ± 0.072	**0.071** ± 0.088
20	0.070 ± 0.087	**0.071** ± 0.237	**0.072** ± 0.100	0.068 ± 0.072	**0.071** ± 0.088
25	0.070 ± 0.087	**0.071** ± 0.237	**0.072** ± 0.100	0.068 ± 0.072	**0.071** ± 0.088
30	0.070 ± 0.087	**0.071** ± 0.237	**0.071** ± 0.101	0.068 ± 0.072	**0.071** ± 0.088
40	0.041 ± 0.063	0.023 ± 0.141	0.016 ± 0.054	**0.068** ± 0.072	0.043 ± 0.064
50	0.028 ± 0.059	0.019 ± 0.136	0.015 ± 0.054	**0.041** ± 0.065	0.029 ± 0.061
60	0.020 ± 0.056	0.016 ± 0.134	0.015 ± 0.054	**0.025** ± 0.059	0.020 ± 0.057
70	0.015 ± 0.053	0.014 ± 0.132	0.014 ± 0.054	**0.015** ± 0.053	**0.015** ± 0.055
80	0.012 ± 0.051	0.013 ± 0.129	**0.014** ± 0.054	0.009 ± 0.048	0.012 ± 0.052
90	0.010 ± 0.049	0.013 ± 0.127	**0.014** ± 0.053	0.005 ± 0.043	0.010 ± 0.049
100	0.009 ± 0.046	0.013 ± 0.126	**0.014** ± 0.053	0.003 ± 0.039	0.009 ± 0.047

Table A.21: Area under the curve of the R_{NX} index on the animal with attributes (AWA) dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.106 ± 0.002	0.105 ± 0.005	0.105 ± 0.002	0.104 ± 0.002	**0.107** ± 0.002
3	0.105 ± 0.002	0.105 ± 0.005	0.106 ± 0.002	0.106 ± 0.002	**0.108** ± 0.002
4	0.106 ± 0.001	0.106 ± 0.004	0.106 ± 0.002	0.106 ± 0.003	**0.109** ± 0.002
5	0.106 ± 0.001	0.107 ± 0.002	0.107 ± 0.001	0.108 ± 0.001	**0.109** ± 0.001
6	0.105 ± 0.002	0.105 ± 0.005	0.106 ± 0.002	0.106 ± 0.002	**0.107** ± 0.002
7	0.105 ± 0.001	0.106 ± 0.002	0.106 ± 0.001	0.106 ± 0.001	**0.108** ± 0.001
8	0.106 ± 0.001	0.106 ± 0.003	0.106 ± 0.001	0.106 ± 0.002	**0.108** ± 0.001
9	0.106 ± 0.001	0.106 ± 0.003	0.106 ± 0.001	0.106 ± 0.001	**0.108** ± 0.001
10	0.106 ± 0.001	0.106 ± 0.002	0.107 ± 0.001	0.108 ± 0.001	**0.109** ± 0.001
15	0.106 ± 0.001	0.106 ± 0.003	0.107 ± 0.001	0.107 ± 0.001	**0.109** ± 0.001
20	0.106 ± 0.001	0.106 ± 0.002	0.107 ± 0.001	0.107 ± 0.001	**0.109** ± 0.001
25	0.106 ± 0.001	0.106 ± 0.002	0.107 ± 0.001	0.107 ± 0.001	**0.109** ± 0.001
30	0.106 ± 0.001	0.107 ± 0.001	0.107 ± 0.001	0.107 ± 0.001	**0.109** ± 0.001
40	0.021 ± 0.004	0.035 ± 0.037	0.064 ± 0.002	**0.107** ± 0.001	0.067 ± 0.002
50	0.021 ± 0.003	0.028 ± 0.020	0.043 ± 0.004	**0.065** ± 0.005	0.045 ± 0.004
60	0.021 ± 0.002	0.024 ± 0.015	0.030 ± 0.001	**0.039** ± 0.002	0.031 ± 0.001
70	0.021 ± 0.002	0.022 ± 0.011	0.023 ± 0.006	**0.024** ± 0.001	0.023 ± 0.006
80	**0.022** ± 0.001	0.020 ± 0.009	0.018 ± 0.003	0.014 ± 0.005	0.018 ± 0.003
90	**0.022** ± 0.009	0.020 ± 0.015	0.015 ± 0.001	0.009 ± 0.003	0.015 ± 0.001
100	**0.022** ± 0.006	0.019 ± 0.014	0.013 ± 0.006	0.005 ± 0.001	0.013 ± 0.006

Table A.22: Clustering purity on the animal with attributes (AWA) dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.137 ± 0.001	0.137 ± 0.004	0.138 ± 0.001	0.139 ± 0.001	**0.141** ± 0.001
3	0.137 ± 0.001	0.137 ± 0.004	0.138 ± 0.002	**0.139** ± 0.003	**0.140** ± 0.002
4	0.137 ± 0.001	0.137 ± 0.003	0.138 ± 0.001	0.139 ± 0.001	**0.140** ± 0.001
5	0.137 ± 0.001	0.138 ± 0.004	0.138 ± 0.002	0.140 ± 0.002	**0.142** ± 0.002
6	0.135 ± 0.001	0.136 ± 0.003	0.136 ± 0.001	0.137 ± 0.002	**0.139** ± 0.001
7	0.136 ± 0.001	0.137 ± 0.003	0.137 ± 0.001	0.138 ± 0.002	**0.140** ± 0.001
8	0.135 ± 0.002	0.135 ± 0.005	0.136 ± 0.002	0.137 ± 0.003	**0.139** ± 0.002
9	0.135 ± 0.001	0.136 ± 0.004	0.136 ± 0.002	0.137 ± 0.002	**0.139** ± 0.002
10	0.135 ± 0.002	0.136 ± 0.004	0.137 ± 0.001	**0.140** ± 0.001	**0.140** ± 0.001
15	0.135 ± 0.001	0.136 ± 0.002	0.137 ± 0.001	**0.139** ± 0.001	**0.140** ± 0.001
20	0.135 ± 0.000	0.136 ± 0.002	0.137 ± 0.001	**0.139** ± 0.001	**0.140** ± 0.001
25	0.135 ± 0.001	0.136 ± 0.002	0.137 ± 0.001	**0.139** ± 0.001	**0.140** ± 0.001
30	0.135 ± 0.001	0.136 ± 0.002	0.137 ± 0.001	0.138 ± 0.000	**0.140** ± 0.001
40	0.027 ± 0.005	0.045 ± 0.048	0.083 ± 0.003	**0.139** ± 0.001	0.086 ± 0.003
50	0.027 ± 0.004	0.036 ± 0.027	0.055 ± 0.006	**0.084** ± 0.007	0.057 ± 0.006
60	0.027 ± 0.004	0.031 ± 0.013	0.039 ± 0.005	**0.050** ± 0.009	0.040 ± 0.005
70	0.027 ± 0.003	0.028 ± 0.011	0.029 ± 0.006	**0.030** ± 0.001	0.029 ± 0.006
80	**0.027** ± 0.003	0.026 ± 0.015	0.023 ± 0.008	0.018 ± 0.002	0.023 ± 0.008
90	**0.027** ± 0.002	0.025 ± 0.011	0.019 ± 0.001	0.011 ± 0.002	0.019 ± 0.001
100	**0.028** ± 0.002	0.024 ± 0.015	0.017 ± 0.001	0.007 ± 0.003	0.017 ± 0.001

Table A.23: Clustering normalized mutual information on the animal with attributes (AWA) dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	**1.981** ± 0.003	**1.980** ± 0.009	**1.982** ± 0.003	**1.983** ± 0.002	2.022 ± 0.003
3	**1.981** ± 0.003	**1.980** ± 0.009	**1.981** ± 0.003	**1.981** ± 0.003	2.019 ± 0.003
4	**1.976** ± 0.004	**1.976** ± 0.012	**1.980** ± 0.003	**1.980** ± 0.003	2.015 ± 0.003
5	**1.978** ± 0.004	**1.977** ± 0.011	**1.979** ± 0.003	**1.978** ± 0.003	2.022 ± 0.003
6	**1.979** ± 0.004	**1.978** ± 0.011	**1.979** ± 0.004	**1.979** ± 0.003	2.027 ± 0.004
7	**1.980** ± 0.004	**1.977** ± 0.011	**1.979** ± 0.004	**1.978** ± 0.003	2.022 ± 0.004
8	**1.978** ± 0.004	**1.978** ± 0.010	**1.978** ± 0.004	**1.978** ± 0.003	2.015 ± 0.004
9	**1.978** ± 0.004	**1.977** ± 0.011	**1.979** ± 0.004	**1.981** ± 0.003	2.021 ± 0.004
10	**1.979** ± 0.004	**1.980** ± 0.010	**1.979** ± 0.004	**1.980** ± 0.003	2.019 ± 0.004
15	**1.981** ± 0.004	**1.980** ± 0.011	**1.979** ± 0.004	**1.983** ± 0.003	2.021 ± 0.004
20	**1.985** ± 0.004	**1.980** ± 0.011	**1.979** ± 0.004	**1.982** ± 0.003	2.032 ± 0.004
25	**1.979** ± 0.004	**1.978** ± 0.011	**1.979** ± 0.004	**1.982** ± 0.003	2.021 ± 0.004
30	**1.978** ± 0.004	**1.979** ± 0.013	**1.979** ± 0.004	**1.983** ± 0.003	2.026 ± 0.004
40	1.191 ± 0.008	0.653 ± 0.668	**0.396** ± 0.004	1.983 ± 0.003	1.235 ± 0.004
50	0.792 ± 0.008	0.524 ± 0.334	**0.396** ± 0.009	1.189 ± 0.010	0.819 ± 0.009
60	0.554 ± 0.008	0.447 ± 0.133	**0.396** ± 0.001	0.713 ± 0.003	0.572 ± 0.002
70	0.411 ± 0.008	0.401 ± 0.017	**0.396** ± 0.004	0.427 ± 0.008	0.421 ± 0.004
80	0.326 ± 0.008	0.373 ± 0.061	0.396 ± 0.001	**0.256** ± 0.002	0.330 ± 0.001
90	0.275 ± 0.008	0.356 ± 0.102	0.396 ± 0.004	**0.153** ± 0.007	0.276 ± 0.004
100	0.244 ± 0.008	0.347 ± 0.129	0.396 ± 0.001	**0.092** ± 0.002	0.244 ± 0.001

Table A.24: Davies-Boulding index on the animal with attributes (AWA) dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.730 ± 0.005	0.776 ± 0.101	**0.858** ± 0.002	**0.861** ± 0.002	0.852 ± 0.002
3	0.737 ± 0.005	0.786 ± 0.087	**0.855** ± 0.003	**0.862** ± 0.002	0.844 ± 0.003
4	0.738 ± 0.005	0.783 ± 0.090	**0.855** ± 0.002	**0.864** ± 0.002	0.855 ± 0.003
5	0.733 ± 0.007	0.784 ± 0.093	**0.857** ± 0.002	**0.864** ± 0.002	**0.859** ± 0.002
6	0.733 ± 0.006	0.786 ± 0.091	**0.857** ± 0.002	**0.862** ± 0.002	**0.861** ± 0.003
7	0.736 ± 0.006	0.787 ± 0.089	**0.857** ± 0.002	**0.863** ± 0.002	**0.859** ± 0.002
8	0.733 ± 0.003	0.788 ± 0.090	**0.857** ± 0.001	**0.862** ± 0.002	**0.860** ± 0.003
9	0.736 ± 0.005	0.788 ± 0.088	**0.857** ± 0.002	**0.862** ± 0.003	**0.858** ± 0.002
10	0.735 ± 0.004	0.788 ± 0.087	**0.855** ± 0.002	**0.862** ± 0.001	**0.857** ± 0.002
15	0.739 ± 0.005	0.788 ± 0.086	**0.856** ± 0.001	**0.862** ± 0.002	**0.857** ± 0.003
20	0.736 ± 0.006	0.788 ± 0.087	**0.856** ± 0.002	**0.862** ± 0.001	**0.859** ± 0.004
25	0.741 ± 0.004	0.788 ± 0.087	**0.857** ± 0.001	**0.863** ± 0.002	**0.862** ± 0.002
30	0.742 ± 0.005	0.789 ± 0.086	**0.857** ± 0.001	**0.864** ± 0.001	**0.862** ± 0.001
40	0.745 ± 0.005	0.791 ± 0.084	**0.858** ± 0.001	**0.862** ± 0.002	**0.860** ± 0.001
50	0.747 ± 0.005	0.792 ± 0.082	**0.857** ± 0.001	**0.860** ± 0.002	**0.859** ± 0.002
60	0.742 ± 0.006	0.790 ± 0.085	**0.857** ± 0.001	**0.859** ± 0.001	**0.857** ± 0.002
70	0.743 ± 0.003	0.790 ± 0.084	**0.857** ± 0.001	**0.860** ± 0.001	**0.854** ± 0.001
80	0.734 ± 0.006	0.785 ± 0.090	**0.857** ± 0.001	**0.860** ± 0.001	**0.856** ± 0.001
90	0.729 ± 0.003	0.784 ± 0.094	**0.857** ± 0.001	**0.860** ± 0.001	**0.855** ± 0.002
100	0.731 ± 0.002	0.783 ± 0.092	**0.856** ± 0.001	**0.860** ± 0.001	**0.855** ± 0.001

Table A.25: One-vs-one SVM classification accuracy on the Berkeley protein dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	-0.068 ± 0.040	0.164 ± 0.510	0.312 ± 0.312	**0.325** ± 0.214	0.269 ± 0.227
3	-0.062 ± 0.048	0.174 ± 0.532	0.321 ± 0.324	**0.343** ± 0.239	0.296 ± 0.252
4	-0.062 ± 0.058	0.179 ± 0.545	0.329 ± 0.332	**0.347** ± 0.248	0.310 ± 0.262
5	-0.059 ± 0.067	0.182 ± 0.551	0.333 ± 0.338	**0.349** ± 0.252	0.315 ± 0.268
6	-0.062 ± 0.069	0.181 ± 0.556	0.335 ± 0.341	**0.350** ± 0.255	0.318 ± 0.270
7	-0.062 ± 0.071	0.182 ± 0.558	0.336 ± 0.343	**0.351** ± 0.257	0.319 ± 0.270
8	-0.058 ± 0.079	0.183 ± 0.558	0.337 ± 0.344	**0.352** ± 0.257	0.320 ± 0.270
9	-0.064 ± 0.074	0.182 ± 0.561	0.338 ± 0.345	**0.352** ± 0.257	0.320 ± 0.271
10	-0.059 ± 0.082	0.184 ± 0.561	0.338 ± 0.345	**0.353** ± 0.257	0.320 ± 0.271
15	-0.063 ± 0.082	0.183 ± 0.563	0.338 ± 0.346	**0.353** ± 0.257	0.321 ± 0.271
20	-0.066 ± 0.084	0.182 ± 0.565	0.339 ± 0.346	**0.353** ± 0.257	0.321 ± 0.271
25	-0.069 ± 0.085	0.181 ± 0.566	0.339 ± 0.346	**0.353** ± 0.257	0.321 ± 0.271
30	-0.069 ± 0.085	0.181 ± 0.566	0.339 ± 0.346	**0.353** ± 0.257	0.321 ± 0.271
40	-0.070 ± 0.086	0.180 ± 0.567	0.339 ± 0.346	**0.353** ± 0.257	0.321 ± 0.271
50	-0.072 ± 0.085	0.180 ± 0.568	0.339 ± 0.346	**0.353** ± 0.257	0.321 ± 0.271
60	-0.073 ± 0.084	0.179 ± 0.568	0.339 ± 0.346	**0.353** ± 0.257	0.321 ± 0.271
70	-0.074 ± 0.083	0.179 ± 0.569	0.339 ± 0.346	**0.353** ± 0.257	0.321 ± 0.271
80	-0.073 ± 0.084	0.179 ± 0.568	0.339 ± 0.346	**0.353** ± 0.257	0.321 ± 0.271
90	-0.074 ± 0.084	0.179 ± 0.569	0.339 ± 0.346	**0.353** ± 0.257	0.321 ± 0.271
100	-0.074 ± 0.083	0.179 ± 0.569	0.339 ± 0.346	**0.353** ± 0.257	0.321 ± 0.271

Table A.26: Cophenetic correlation on the Berkeley protein dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.045 ± 0.028	0.102 ± 0.167	0.148 ± 0.101	**0.158** ± 0.025	0.144 ± 0.026
3	0.045 ± 0.027	0.105 ± 0.173	0.151 ± 0.103	**0.166** ± 0.029	0.150 ± 0.031
4	0.044 ± 0.026	0.107 ± 0.177	0.155 ± 0.106	**0.169** ± 0.030	0.156 ± 0.033
5	0.045 ± 0.025	0.108 ± 0.179	0.156 ± 0.107	**0.170** ± 0.031	0.158 ± 0.035
6	0.044 ± 0.024	0.108 ± 0.182	0.158 ± 0.108	**0.170** ± 0.031	0.160 ± 0.035
7	0.043 ± 0.024	0.108 ± 0.182	0.159 ± 0.109	**0.171** ± 0.031	0.160 ± 0.035
8	0.045 ± 0.024	0.109 ± 0.182	0.159 ± 0.109	**0.171** ± 0.031	0.160 ± 0.035
9	0.043 ± 0.023	0.108 ± 0.183	0.159 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
10	0.044 ± 0.024	0.109 ± 0.183	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
15	0.042 ± 0.022	0.108 ± 0.183	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
20	0.040 ± 0.021	0.108 ± 0.184	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
25	0.038 ± 0.020	0.107 ± 0.185	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
30	0.038 ± 0.020	0.107 ± 0.185	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
40	0.037 ± 0.019	0.107 ± 0.185	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
50	0.036 ± 0.019	0.106 ± 0.185	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
60	0.035 ± 0.019	0.106 ± 0.186	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
70	0.035 ± 0.019	0.106 ± 0.186	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
80	0.035 ± 0.019	0.106 ± 0.186	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
90	0.034 ± 0.019	0.106 ± 0.186	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035
100	0.035 ± 0.019	0.106 ± 0.186	0.160 ± 0.110	**0.171** ± 0.031	0.161 ± 0.035

Table A.27: Area under the curve of the R_{NX} index on the Berkeley protein dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.700 ± 0.000	0.712 ± 0.030	**0.736** ± 0.005	0.700 ± 0.000	0.700 ± 0.000
3	0.700 ± 0.000	0.718 ± 0.044	**0.753** ± 0.005	0.744 ± 0.006	0.732 ± 0.002
4	0.700 ± 0.000	0.719 ± 0.046	**0.756** ± 0.004	**0.755** ± 0.002	0.747 ± 0.003
5	0.700 ± 0.000	0.719 ± 0.046	**0.756** ± 0.005	**0.756** ± 0.001	0.748 ± 0.002
6	0.700 ± 0.000	0.719 ± 0.048	**0.758** ± 0.003	**0.756** ± 0.001	0.750 ± 0.001
7	0.700 ± 0.000	0.720 ± 0.049	**0.760** ± 0.002	**0.756** ± 0.001	0.749 ± 0.001
8	0.700 ± 0.000	0.720 ± 0.049	**0.760** ± 0.001	**0.756** ± 0.000	0.749 ± 0.000
9	0.700 ± 0.000	0.720 ± 0.049	**0.760** ± 0.001	**0.756** ± 0.000	0.749 ± 0.001
10	0.700 ± 0.000	0.720 ± 0.049	**0.760** ± 0.001	**0.756** ± 0.000	0.749 ± 0.000
15	0.700 ± 0.000	0.720 ± 0.049	**0.761** ± 0.000	**0.756** ± 0.000	0.749 ± 0.000
20	0.700 ± 0.000	0.720 ± 0.049	**0.761** ± 0.000	**0.756** ± 0.000	0.749 ± 0.001
25	0.700 ± 0.000	0.720 ± 0.049	**0.760** ± 0.000	**0.756** ± 0.000	0.749 ± 0.000
30	0.700 ± 0.000	0.720 ± 0.049	**0.761** ± 0.000	**0.756** ± 0.000	0.749 ± 0.000
40	0.700 ± 0.000	0.720 ± 0.049	**0.761** ± 0.000	**0.756** ± 0.000	0.749 ± 0.000
50	0.700 ± 0.000	0.720 ± 0.049	**0.761** ± 0.000	**0.756** ± 0.000	0.749 ± 0.000
60	0.700 ± 0.000	0.720 ± 0.049	**0.761** ± 0.000	**0.756** ± 0.000	0.749 ± 0.000
70	0.700 ± 0.000	0.720 ± 0.049	**0.761** ± 0.000	**0.756** ± 0.000	0.749 ± 0.000
80	0.700 ± 0.000	0.720 ± 0.049	**0.761** ± 0.000	**0.756** ± 0.000	0.749 ± 0.000
90	0.700 ± 0.000	0.720 ± 0.049	**0.761** ± 0.000	**0.756** ± 0.000	0.749 ± 0.000
100	0.700 ± 0.000	0.720 ± 0.049	**0.761** ± 0.000	**0.756** ± 0.000	0.749 ± 0.000

Table A.28: Clustering purity on the Berkeley protein dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.015 ± 0.012	0.120 ± 0.164	**0.243** ± 0.007	0.199 ± 0.004	0.187 ± 0.004
3	0.016 ± 0.007	0.127 ± 0.180	0.265 ± 0.006	**0.277** ± 0.007	0.247 ± 0.007
4	0.013 ± 0.013	0.134 ± 0.184	0.270 ± 0.003	**0.290** ± 0.004	0.269 ± 0.003
5	0.024 ± 0.013	0.145 ± 0.175	0.270 ± 0.005	**0.291** ± 0.002	0.270 ± 0.003
6	0.036 ± 0.015	0.151 ± 0.168	0.272 ± 0.004	**0.292** ± 0.001	0.272 ± 0.001
7	0.035 ± 0.012	0.151 ± 0.171	0.275 ± 0.002	**0.292** ± 0.001	0.272 ± 0.001
8	0.047 ± 0.011	0.154 ± 0.162	0.274 ± 0.002	**0.292** ± 0.001	0.272 ± 0.000
9	0.049 ± 0.009	0.154 ± 0.160	0.273 ± 0.001	**0.292** ± 0.000	0.271 ± 0.001
10	0.053 ± 0.009	0.157 ± 0.158	0.275 ± 0.002	**0.292** ± 0.000	0.271 ± 0.000
15	0.065 ± 0.009	0.160 ± 0.150	0.274 ± 0.000	**0.292** ± 0.000	0.271 ± 0.000
20	0.058 ± 0.008	0.158 ± 0.154	0.274 ± 0.000	**0.291** ± 0.000	0.272 ± 0.001
25	0.066 ± 0.013	0.160 ± 0.150	0.274 ± 0.000	**0.291** ± 0.000	0.271 ± 0.000
30	0.064 ± 0.006	0.161 ± 0.150	0.274 ± 0.000	**0.291** ± 0.000	0.271 ± 0.000
40	0.061 ± 0.006	0.160 ± 0.152	0.274 ± 0.000	**0.292** ± 0.000	0.271 ± 0.000
50	0.059 ± 0.006	0.158 ± 0.154	0.274 ± 0.000	**0.291** ± 0.000	0.271 ± 0.000
60	0.056 ± 0.006	0.156 ± 0.156	0.274 ± 0.000	**0.291** ± 0.000	0.271 ± 0.000
70	0.059 ± 0.006	0.158 ± 0.153	0.274 ± 0.000	**0.291** ± 0.000	0.271 ± 0.000
80	0.052 ± 0.011	0.156 ± 0.159	0.274 ± 0.000	**0.292** ± 0.000	0.271 ± 0.000
90	0.057 ± 0.009	0.158 ± 0.155	0.274 ± 0.000	**0.291** ± 0.000	0.271 ± 0.000
100	0.059 ± 0.009	0.159 ± 0.153	0.274 ± 0.000	**0.291** ± 0.000	0.271 ± 0.000

Table A.29: Clustering normalized mutual information on the Berkeley protein dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	1.976 ± 0.173	1.826 ± 0.315	1.750 ± 0.026	**1.721** ± 0.068	**1.735** ± 0.059
3	1.966 ± 0.235	1.801 ± 0.369	1.695 ± 0.033	**1.677** ± 0.067	**1.689** ± 0.064
4	1.965 ± 0.309	1.763 ± 0.459	**1.584** ± 0.039	1.664 ± 0.070	1.677 ± 0.070
5	1.943 ± 0.293	1.748 ± 0.450	**1.563** ± 0.048	1.664 ± 0.070	1.676 ± 0.071
6	1.920 ± 0.314	1.732 ± 0.466	**1.537** ± 0.057	1.663 ± 0.070	1.674 ± 0.072
7	1.904 ± 0.304	1.731 ± 0.451	**1.548** ± 0.068	1.664 ± 0.070	1.674 ± 0.072
8	1.875 ± 0.303	1.722 ± 0.441	**1.552** ± 0.081	1.663 ± 0.070	1.674 ± 0.072
9	1.880 ± 0.309	1.721 ± 0.449	**1.544** ± 0.076	1.663 ± 0.070	1.674 ± 0.072
10	1.855 ± 0.298	1.717 ± 0.432	**1.556** ± 0.091	1.663 ± 0.070	1.674 ± 0.072
15	1.821 ± 0.307	1.702 ± 0.436	**1.545** ± 0.107	1.664 ± 0.070	1.674 ± 0.072
20	1.803 ± 0.299	1.699 ± 0.430	**1.555** ± 0.135	1.664 ± 0.070	1.674 ± 0.072
25	1.789 ± 0.304	1.693 ± 0.435	**1.550** ± 0.145	1.664 ± 0.070	1.674 ± 0.072
30	1.764 ± 0.297	1.687 ± 0.425	**1.557** ± 0.152	1.664 ± 0.070	1.674 ± 0.072
40	1.745 ± 0.302	1.679 ± 0.435	**1.551** ± 0.171	1.664 ± 0.070	1.674 ± 0.072
50	1.764 ± 0.305	1.684 ± 0.434	**1.550** ± 0.154	1.664 ± 0.070	1.674 ± 0.072
60	1.765 ± 0.304	1.685 ± 0.432	**1.550** ± 0.151	1.664 ± 0.070	1.674 ± 0.072
70	1.758 ± 0.300	1.684 ± 0.428	**1.554** ± 0.155	1.664 ± 0.070	1.674 ± 0.072
80	1.762 ± 0.298	1.686 ± 0.427	**1.556** ± 0.155	1.664 ± 0.070	1.674 ± 0.072
90	1.772 ± 0.303	1.687 ± 0.432	**1.551** ± 0.147	1.664 ± 0.070	1.674 ± 0.072
100	1.769 ± 0.300	1.688 ± 0.426	**1.555** ± 0.145	1.664 ± 0.070	1.674 ± 0.072

Table A.30: Davies-Boulding index on the Berkeley protein dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.432 ± 0.007	0.482 ± 0.073	**0.533** ± 0.009	0.484 ± 0.008	**0.538** ± 0.009
3	0.438 ± 0.007	0.491 ± 0.076	**0.544** ± 0.009	0.494 ± 0.008	**0.550** ± 0.009
4	0.457 ± 0.005	0.508 ± 0.074	**0.560** ± 0.006	0.509 ± 0.006	**0.565** ± 0.006
5	0.464 ± 0.004	0.518 ± 0.077	**0.572** ± 0.005	0.519 ± 0.004	**0.577** ± 0.005
6	0.468 ± 0.002	0.525 ± 0.080	**0.582** ± 0.003	0.530 ± 0.003	**0.587** ± 0.003
7	0.476 ± 0.005	0.531 ± 0.078	0.586 ± 0.006	0.534 ± 0.005	**0.593** ± 0.006
8	0.474 ± 0.004	0.532 ± 0.083	0.591 ± 0.005	0.540 ± 0.004	**0.597** ± 0.005
9	0.485 ± 0.002	0.545 ± 0.085	**0.606** ± 0.002	0.543 ± 0.002	**0.605** ± 0.002
10	0.546 ± 0.006	0.615 ± 0.098	**0.684** ± 0.008	0.616 ± 0.007	**0.684** ± 0.008
15	0.572 ± 0.006	0.645 ± 0.104	**0.718** ± 0.008	0.645 ± 0.007	**0.718** ± 0.008
20	0.582 ± 0.003	0.652 ± 0.099	**0.722** ± 0.003	0.648 ± 0.003	**0.722** ± 0.003
25	0.566 ± 0.003	0.638 ± 0.102	**0.710** ± 0.003	0.637 ± 0.003	**0.710** ± 0.003
30	0.558 ± 0.003	0.627 ± 0.097	**0.695** ± 0.003	0.626 ± 0.003	**0.696** ± 0.003
40	0.328 ± 0.003	0.367 ± 0.055	**0.406** ± 0.003	0.365 ± 0.003	**0.406** ± 0.003
50	0.107 ± 0.002	0.120 ± 0.019	**0.133** ± 0.003	0.120 ± 0.002	**0.133** ± 0.003
60	0.063 ± 0.002	0.070 ± 0.011	**0.078** ± 0.003	0.070 ± 0.002	**0.078** ± 0.003
70	0.036 ± 0.002	0.041 ± 0.007	**0.045** ± 0.003	0.041 ± 0.003	**0.045** ± 0.003
80	0.021 ± 0.002	0.024 ± 0.005	**0.026** ± 0.003	0.024 ± 0.003	**0.026** ± 0.003
90	0.012 ± 0.002	0.014 ± 0.005	**0.015** ± 0.003	0.014 ± 0.003	**0.015** ± 0.003
100	0.007 ± 0.003	0.008 ± 0.004	**0.009** ± 0.003	0.008 ± 0.003	**0.009** ± 0.003

Table A.31: One-vs-one SVM classification accuracy on the Cora dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.330 ± 0.001	0.370 ± 0.058	**0.411** ± 0.001	0.371 ± 0.001	**0.411** ± 0.001
3	0.349 ± 0.001	0.391 ± 0.060	**0.434** ± 0.002	0.390 ± 0.002	**0.434** ± 0.002
4	0.341 ± 0.006	0.382 ± 0.060	**0.424** ± 0.007	0.382 ± 0.007	**0.424** ± 0.007
5	0.334 ± 0.007	0.377 ± 0.062	**0.420** ± 0.008	0.378 ± 0.007	**0.421** ± 0.008
6	0.332 ± 0.004	0.375 ± 0.061	**0.418** ± 0.005	0.375 ± 0.005	**0.419** ± 0.005
7	0.337 ± 0.006	0.381 ± 0.062	**0.424** ± 0.007	0.380 ± 0.007	**0.423** ± 0.007
8	0.339 ± 0.004	0.380 ± 0.059	**0.421** ± 0.005	0.379 ± 0.005	**0.421** ± 0.005
9	0.329 ± 0.001	0.372 ± 0.061	**0.415** ± 0.001	0.374 ± 0.001	**0.416** ± 0.001
10	0.374 ± 0.001	0.422 ± 0.068	**0.469** ± 0.002	0.423 ± 0.002	**0.470** ± 0.002
15	0.402 ± 0.002	0.451 ± 0.070	**0.501** ± 0.002	0.450 ± 0.002	**0.501** ± 0.002
20	0.397 ± 0.010	0.444 ± 0.068	**0.492** ± 0.001	0.442 ± 0.001	**0.492** ± 0.001
25	0.383 ± 0.002	0.432 ± 0.069	**0.480** ± 0.002	0.432 ± 0.002	**0.480** ± 0.002
30	0.382 ± 0.002	0.430 ± 0.068	**0.478** ± 0.002	0.429 ± 0.002	**0.477** ± 0.002
40	0.231 ± 0.002	0.261 ± 0.042	**0.290** ± 0.002	0.261 ± 0.002	**0.290** ± 0.002
50	0.074 ± 0.001	0.083 ± 0.013	**0.092** ± 0.002	0.082 ± 0.002	**0.092** ± 0.002
60	0.045 ± 0.002	0.050 ± 0.008	**0.056** ± 0.002	0.050 ± 0.002	**0.056** ± 0.002
70	0.027 ± 0.002	0.031 ± 0.006	**0.034** ± 0.003	0.030 ± 0.002	**0.034** ± 0.003
80	0.016 ± 0.002	0.018 ± 0.005	**0.021** ± 0.003	0.018 ± 0.003	**0.021** ± 0.003
90	0.010 ± 0.003	0.011 ± 0.005	**0.012** ± 0.003	0.011 ± 0.003	**0.012** ± 0.003
100	0.006 ± 0.003	0.007 ± 0.005	**0.008** ± 0.004	0.007 ± 0.004	**0.008** ± 0.004

Table A.32: Clustering purity on the Cora dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-tSNE
	Worst	Average	Best		
2	0.110 ± 0.007	0.125 ± 0.023	**0.139** ± 0.009	0.125 ± 0.008	**0.139** ± 0.009
3	0.129 ± 0.008	0.145 ± 0.026	**0.162** ± 0.010	0.145 ± 0.009	**0.161** ± 0.010
4	0.128 ± 0.004	0.144 ± 0.024	**0.160** ± 0.005	0.144 ± 0.005	**0.160** ± 0.005
5	0.124 ± 0.003	0.140 ± 0.023	**0.157** ± 0.004	0.141 ± 0.004	**0.157** ± 0.004
6	0.126 ± 0.006	0.141 ± 0.024	**0.157** ± 0.007	0.142 ± 0.006	**0.157** ± 0.007
7	0.129 ± 0.006	0.146 ± 0.025	**0.162** ± 0.007	0.146 ± 0.007	**0.162** ± 0.007
8	0.128 ± 0.007	0.144 ± 0.025	**0.159** ± 0.008	0.144 ± 0.007	**0.159** ± 0.008
9	0.148 ± 0.005	0.168 ± 0.029	**0.187** ± 0.006	0.168 ± 0.006	**0.187** ± 0.006
10	0.176 ± 0.001	0.197 ± 0.031	**0.219** ± 0.002	0.197 ± 0.002	**0.219** ± 0.002
15	0.201 ± 0.002	0.226 ± 0.035	**0.251** ± 0.002	0.226 ± 0.002	**0.251** ± 0.002
20	0.206 ± 0.008	0.231 ± 0.038	**0.256** ± 0.010	0.230 ± 0.009	**0.256** ± 0.010
25	0.201 ± 0.001	0.225 ± 0.034	**0.250** ± 0.002	0.225 ± 0.001	**0.249** ± 0.002
30	0.202 ± 0.001	0.227 ± 0.036	**0.252** ± 0.001	0.226 ± 0.001	**0.252** ± 0.002
40	0.121 ± 0.010	0.137 ± 0.024	**0.152** ± 0.001	0.138 ± 0.001	**0.153** ± 0.001
50	0.037 ± 0.008	0.042 ± 0.014	**0.047** ± 0.009	0.042 ± 0.008	**0.047** ± 0.009
60	0.023 ± 0.006	0.026 ± 0.011	**0.029** ± 0.008	0.026 ± 0.007	**0.029** ± 0.008
70	0.014 ± 0.005	0.016 ± 0.009	**0.017** ± 0.007	0.015 ± 0.006	**0.017** ± 0.007
80	0.008 ± 0.004	0.009 ± 0.007	**0.010** ± 0.005	0.009 ± 0.005	**0.010** ± 0.005
90	0.005 ± 0.004	0.006 ± 0.006	**0.006** ± 0.005	0.006 ± 0.004	**0.006** ± 0.005
100	0.003 ± 0.003	0.003 ± 0.005	**0.004** ± 0.004	0.003 ± 0.003	**0.004** ± 0.004

Table A.33: Clustering normalized mutual information on the Cora dataset of MV-tSNE compared with single view and stacked views tSNE. K is the dimensionality of the projection.

Appendix B

Results of MV-MDS experiments

In this appendix, the detailed results of the experiments with the MV-MDS method are presented. There is a results table for each combination of dataset and evaluation method, yielding a total of 36 tables. The methods are compared with the counterpart single view method, either applied to single views individually or to all views stacked on a single matrix. For single views, the worst, average and best views (on average) are given.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.516 ± 0.016	0.571 ± 0.144	0.488 ± 0.014	0.650 ± 0.016	**0.741** ± 0.009
3	0.609 ± 0.011	0.719 ± 0.143	0.692 ± 0.015	0.811 ± 0.015	**0.880** ± 0.004
4	0.680 ± 0.009	0.777 ± 0.150	0.787 ± 0.013	0.880 ± 0.010	**0.941** ± 0.008
5	0.706 ± 0.012	0.816 ± 0.179	0.864 ± 0.011	0.921 ± 0.009	**0.951** ± 0.005
6	0.749 ± 0.011	0.834 ± 0.184	0.875 ± 0.007	0.939 ± 0.006	**0.968** ± 0.004
7	0.769 ± 0.015	0.861 ± 0.190	0.898 ± 0.008	0.957 ± 0.004	**0.971** ± 0.006
8	0.798 ± 0.020	0.873 ± 0.188	0.915 ± 0.009	**0.968** ± 0.005	**0.977** ± 0.005
9	0.810 ± 0.012	0.880 ± 0.185	0.928 ± 0.008	**0.970** ± 0.004	**0.979** ± 0.007
10	0.824 ± 0.015	0.891 ± 0.175	0.942 ± 0.006	**0.973** ± 0.005	**0.975** ± 0.007
15	0.830 ± 0.017	0.902 ± 0.177	0.966 ± 0.010	**0.977** ± 0.007	**0.980** ± 0.004
20	0.831 ± 0.017	0.910 ± 0.163	0.970 ± 0.007	**0.980** ± 0.007	**0.975** ± 0.006
25	0.822 ± 0.016	0.911 ± 0.168	**0.968** ± 0.009	**0.977** ± 0.005	**0.972** ± 0.005
30	0.800 ± 0.019	0.900 ± 0.178	0.961 ± 0.009	**0.974** ± 0.004	0.963 ± 0.005
40	0.722 ± 0.024	0.839 ± 0.189	0.922 ± 0.016	**0.928** ± 0.010	**0.932** ± 0.007
50	0.701 ± 0.018	0.748 ± 0.161	0.832 ± 0.025	0.835 ± 0.016	**0.844** ± 0.025
60	0.639 ± 0.030	0.589 ± 0.231	0.688 ± 0.054	0.663 ± 0.045	**0.700** ± 0.056
70	**0.520** ± 0.064	0.426 ± 0.338	0.496 ± 0.081	0.421 ± 0.075	0.510 ± 0.081
80	**0.343** ± 0.105	0.300 ± 0.319	0.339 ± 0.079	0.247 ± 0.082	**0.342** ± 0.085
90	0.200 ± 0.091	0.215 ± 0.291	0.233 ± 0.074	0.156 ± 0.055	**0.238** ± 0.081
100	0.116 ± 0.039	0.158 ± 0.232	**0.166** ± 0.058	0.112 ± 0.022	**0.164** ± 0.057

Table B.1: One-vs-one SVM classification accuracy on the digits dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.214 ± 0.269	0.293 ± 0.531	0.359 ± 0.238	**0.392** ± 0.212	**0.394** ± 0.082
3	0.214 ± 0.269	0.328 ± 0.611	0.389 ± 0.278	**0.441** ± 0.269	**0.437** ± 0.112
4	0.214 ± 0.269	0.347 ± 0.632	0.430 ± 0.291	**0.483** ± 0.278	0.461 ± 0.121
5	0.214 ± 0.269	0.363 ± 0.636	0.453 ± 0.288	**0.516** ± 0.259	0.462 ± 0.109
6	0.214 ± 0.269	0.374 ± 0.649	0.475 ± 0.290	**0.525** ± 0.268	0.476 ± 0.107
7	0.214 ± 0.269	0.382 ± 0.654	0.487 ± 0.294	**0.542** ± 0.272	0.482 ± 0.119
8	0.214 ± 0.269	0.387 ± 0.659	0.497 ± 0.293	**0.552** ± 0.268	0.486 ± 0.121
9	0.214 ± 0.269	0.391 ± 0.666	0.504 ± 0.297	**0.559** ± 0.273	0.495 ± 0.118
10	0.214 ± 0.269	0.394 ± 0.672	0.510 ± 0.301	**0.562** ± 0.275	0.489 ± 0.117
15	0.214 ± 0.269	0.402 ± 0.686	0.524 ± 0.309	**0.577** ± 0.283	0.464 ± 0.117
20	0.214 ± 0.269	0.407 ± 0.694	0.533 ± 0.314	**0.585** ± 0.284	0.441 ± 0.120
25	0.214 ± 0.269	0.409 ± 0.699	0.537 ± 0.316	**0.590** ± 0.285	0.408 ± 0.122
30	0.214 ± 0.269	0.411 ± 0.701	0.540 ± 0.316	**0.593** ± 0.285	0.367 ± 0.115
40	0.214 ± 0.269	0.412 ± 0.704	0.542 ± 0.317	**0.598** ± 0.285	0.345 ± 0.116
50	0.214 ± 0.269	0.413 ± 0.706	0.543 ± 0.318	**0.599** ± 0.285	0.324 ± 0.114
60	0.214 ± 0.269	0.413 ± 0.707	0.543 ± 0.318	**0.601** ± 0.285	0.311 ± 0.111
70	0.214 ± 0.269	0.413 ± 0.708	0.543 ± 0.318	**0.602** ± 0.285	0.297 ± 0.105
80	0.214 ± 0.269	0.413 ± 0.708	0.543 ± 0.318	**0.603** ± 0.285	0.290 ± 0.104
90	0.214 ± 0.269	0.413 ± 0.708	0.543 ± 0.318	**0.603** ± 0.285	0.282 ± 0.105
100	0.214 ± 0.269	0.413 ± 0.708	0.543 ± 0.318	**0.604** ± 0.284	0.275 ± 0.105

Table B.2: Cophenetic correlation on the digits dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.116 ± 0.071	0.130 ± 0.144	0.148 ± 0.068	0.159 ± 0.065	**0.161** ± 0.031
3	0.116 ± 0.071	0.171 ± 0.212	0.207 ± 0.103	**0.226** ± 0.106	0.209 ± 0.049
4	0.116 ± 0.071	0.197 ± 0.262	0.247 ± 0.127	**0.268** ± 0.134	0.244 ± 0.064
5	0.116 ± 0.071	0.215 ± 0.300	0.274 ± 0.146	**0.298** ± 0.148	0.258 ± 0.069
6	0.116 ± 0.071	0.229 ± 0.331	0.294 ± 0.160	**0.321** ± 0.162	0.273 ± 0.075
7	0.116 ± 0.071	0.241 ± 0.354	0.316 ± 0.172	**0.340** ± 0.176	0.285 ± 0.083
8	0.116 ± 0.071	0.251 ± 0.372	0.334 ± 0.181	**0.352** ± 0.181	0.290 ± 0.090
9	0.116 ± 0.071	0.256 ± 0.386	0.341 ± 0.188	**0.362** ± 0.185	0.298 ± 0.093
10	0.116 ± 0.071	0.262 ± 0.398	0.350 ± 0.195	**0.371** ± 0.190	0.302 ± 0.095
15	0.116 ± 0.071	0.277 ± 0.437	0.375 ± 0.216	**0.397** ± 0.211	0.306 ± 0.097
20	0.116 ± 0.071	0.285 ± 0.460	0.392 ± 0.232	**0.410** ± 0.220	0.310 ± 0.102
25	0.116 ± 0.071	0.289 ± 0.474	0.401 ± 0.240	**0.416** ± 0.223	0.312 ± 0.108
30	0.116 ± 0.071	0.292 ± 0.482	0.405 ± 0.245	**0.420** ± 0.225	0.308 ± 0.107
40	0.116 ± 0.071	0.295 ± 0.493	0.411 ± 0.251	**0.424** ± 0.226	0.297 ± 0.105
50	0.116 ± 0.071	0.296 ± 0.500	0.414 ± 0.255	**0.425** ± 0.224	0.286 ± 0.101
60	0.116 ± 0.071	0.297 ± 0.504	0.415 ± 0.257	**0.426** ± 0.223	0.276 ± 0.097
70	0.116 ± 0.071	0.298 ± 0.507	0.416 ± 0.258	**0.427** ± 0.223	0.268 ± 0.096
80	0.116 ± 0.071	0.298 ± 0.508	0.417 ± 0.259	**0.427** ± 0.222	0.260 ± 0.094
90	0.116 ± 0.071	0.298 ± 0.508	0.417 ± 0.260	**0.428** ± 0.221	0.253 ± 0.092
100	0.116 ± 0.071	0.298 ± 0.508	0.417 ± 0.261	**0.428** ± 0.221	0.247 ± 0.091

Table B.3: Area under the curve of the R_{NX} index on the digits dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.445	0.510 ± 0.048	0.538	0.556	**0.616**
3	0.445	0.550 ± 0.072	0.583	0.697	**0.743**
4	0.445	0.602 ± 0.088	0.684	**0.767**	0.758
5	0.445	0.597 ± 0.087	0.659	**0.798**	**0.792**
6	0.445	0.600 ± 0.091	0.694	0.832	**0.844**
7	0.445	0.634 ± 0.116	0.775	**0.854**	0.801
8	0.445	0.653 ± 0.126	0.787	0.798	**0.886**
9	0.445	0.671 ± 0.144	0.837	**0.880**	**0.873**
10	0.445	0.671 ± 0.144	0.812	**0.885**	**0.887**
15	0.445	0.696 ± 0.163	0.871	**0.914**	0.849
20	0.445	0.695 ± 0.161	0.861	**0.918**	0.824
25	0.445	0.697 ± 0.165	0.876	**0.921**	0.850
30	0.445	0.697 ± 0.165	**0.875**	0.837	0.855
40	0.445	0.697 ± 0.165	**0.877**	0.838	**0.873**
50	0.445	0.695 ± 0.163	**0.876**	0.837	0.848
60	0.445	0.695 ± 0.163	**0.877**	0.838	0.849
70	0.445	0.695 ± 0.163	**0.877**	0.839	0.831
80	0.445	0.695 ± 0.163	**0.877**	0.839	**0.873**
90	0.445	0.695 ± 0.163	**0.877**	0.839	**0.871**
100	0.445	0.695 ± 0.163	**0.877**	0.839	**0.872**

Table B.4: Clustering purity on the digits dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.412	0.471 ± 0.052	0.489	0.530	**0.628**
3	0.416	0.524 ± 0.064	0.551	0.631	**0.742**
4	0.461	0.553 ± 0.065	0.606	0.661	**0.737**
5	0.441	0.564 ± 0.076	0.616	0.717	**0.756**
6	0.453	0.580 ± 0.085	0.665	0.741	**0.802**
7	0.457	0.599 ± 0.096	0.706	**0.767**	**0.769**
8	0.476	0.614 ± 0.103	0.719	0.767	**0.820**
9	0.463	0.621 ± 0.114	0.745	**0.804**	**0.810**
10	0.465	0.623 ± 0.114	0.734	0.810	**0.821**
15	0.474	0.642 ± 0.130	0.788	**0.842**	**0.836**
20	0.472	0.641 ± 0.128	0.771	**0.847**	0.824
25	0.474	0.646 ± 0.133	0.799	**0.851**	**0.845**
30	0.474	0.646 ± 0.134	0.797	0.806	**0.864**
40	0.474	0.646 ± 0.134	0.800	0.807	**0.888**
50	0.474	0.643 ± 0.131	0.798	0.806	**0.863**
60	0.474	0.644 ± 0.131	0.798	0.807	**0.872**
70	0.474	0.644 ± 0.131	0.798	0.809	**0.853**
80	0.474	0.644 ± 0.131	0.798	0.808	**0.891**
90	0.474	0.644 ± 0.131	0.798	0.809	**0.891**
100	0.474	0.644 ± 0.131	0.798	0.809	**0.892**

Table B.5: Clustering normalized mutual information on the digits dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	1.904 ± 0.053	**1.878** ± 0.179	**1.870** ± 0.094	**1.870** ± 0.054	**1.874** ± 0.045
3	1.904 ± 0.076	1.850 ± 0.205	1.829 ± 0.094	**1.803** ± 0.085	**1.788** ± 0.092
4	1.904 ± 0.084	1.827 ± 0.238	1.789 ± 0.094	**1.762** ± 0.095	**1.778** ± 0.106
5	1.904 ± 0.092	1.825 ± 0.239	1.782 ± 0.094	**1.742** ± 0.108	**1.758** ± 0.128
6	1.904 ± 0.089	1.822 ± 0.246	1.784 ± 0.094	**1.724** ± 0.116	**1.738** ± 0.144
7	1.904 ± 0.109	1.813 ± 0.266	1.738 ± 0.094	**1.717** ± 0.121	1.737 ± 0.114
8	1.904 ± 0.112	1.796 ± 0.289	1.732 ± 0.094	**1.720** ± 0.126	**1.709** ± 0.134
9	1.904 ± 0.112	1.793 ± 0.290	1.731 ± 0.094	**1.704** ± 0.128	**1.708** ± 0.129
10	1.904 ± 0.116	1.790 ± 0.297	1.727 ± 0.094	**1.702** ± 0.130	**1.703** ± 0.132
15	1.904 ± 0.114	1.791 ± 0.298	**1.719** ± 0.094	**1.703** ± 0.131	**1.718** ± 0.130
20	1.904 ± 0.118	1.791 ± 0.300	1.721 ± 0.094	**1.699** ± 0.133	1.720 ± 0.107
25	1.904 ± 0.115	1.790 ± 0.300	**1.716** ± 0.094	**1.700** ± 0.132	**1.715** ± 0.143
30	1.904 ± 0.114	1.790 ± 0.300	**1.717** ± 0.094	**1.709** ± 0.122	**1.714** ± 0.106
40	1.904 ± 0.115	1.790 ± 0.302	**1.716** ± 0.094	**1.709** ± 0.122	**1.724** ± 0.114
50	1.904 ± 0.116	1.791 ± 0.300	**1.716** ± 0.094	**1.709** ± 0.122	**1.726** ± 0.121
60	1.904 ± 0.115	1.791 ± 0.300	**1.716** ± 0.094	**1.709** ± 0.122	**1.708** ± 0.136
70	1.904 ± 0.115	1.791 ± 0.299	**1.716** ± 0.094	**1.709** ± 0.122	**1.720** ± 0.136
80	1.904 ± 0.115	1.791 ± 0.299	1.716 ± 0.094	**1.709** ± 0.122	**1.698** ± 0.133
90	1.904 ± 0.115	1.791 ± 0.299	**1.716** ± 0.094	**1.708** ± 0.122	**1.701** ± 0.130
100	1.904 ± 0.115	1.791 ± 0.299	**1.716** ± 0.094	**1.708** ± 0.122	**1.712** ± 0.135

Table B.6: Davies-Boulding index on the digits dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.532 ± 0.015	0.555 ± 0.058	0.591 ± 0.009	**0.607** ± 0.021	0.586 ± 0.019
3	0.570 ± 0.028	0.592 ± 0.089	**0.660** ± 0.009	0.610 ± 0.011	0.611 ± 0.025
4	0.614 ± 0.022	0.636 ± 0.077	**0.693** ± 0.009	0.612 ± 0.025	0.637 ± 0.019
5	0.640 ± 0.018	0.661 ± 0.097	**0.742** ± 0.006	0.657 ± 0.023	0.672 ± 0.020
6	0.659 ± 0.016	0.683 ± 0.094	**0.761** ± 0.005	0.670 ± 0.019	0.683 ± 0.019
7	0.685 ± 0.017	0.696 ± 0.091	**0.770** ± 0.006	0.705 ± 0.013	0.713 ± 0.015
8	0.691 ± 0.021	0.719 ± 0.085	**0.787** ± 0.007	0.713 ± 0.009	0.724 ± 0.016
9	0.702 ± 0.017	0.726 ± 0.082	**0.790** ± 0.005	0.732 ± 0.010	0.745 ± 0.019
10	0.695 ± 0.017	0.728 ± 0.085	**0.793** ± 0.005	0.731 ± 0.018	0.745 ± 0.017
15	0.704 ± 0.015	0.743 ± 0.083	**0.806** ± 0.006	0.774 ± 0.012	0.776 ± 0.011
20	0.719 ± 0.014	0.753 ± 0.087	**0.818** ± 0.008	0.779 ± 0.014	0.785 ± 0.017
25	0.714 ± 0.023	0.749 ± 0.098	**0.827** ± 0.006	0.783 ± 0.016	0.787 ± 0.012
30	0.722 ± 0.010	0.742 ± 0.102	**0.828** ± 0.008	0.773 ± 0.013	0.777 ± 0.014
40	0.672 ± 0.009	0.703 ± 0.133	**0.814** ± 0.008	0.711 ± 0.016	0.710 ± 0.020
50	0.619 ± 0.018	0.643 ± 0.134	**0.756** ± 0.012	0.620 ± 0.014	0.626 ± 0.015
60	0.534 ± 0.008	0.571 ± 0.155	**0.687** ± 0.016	0.535 ± 0.025	0.543 ± 0.029
70	0.464 ± 0.066	0.464 ± 0.155	**0.506** ± 0.037	0.455 ± 0.030	0.456 ± 0.031
80	0.351 ± 0.036	0.368 ± 0.105	**0.420** ± 0.014	0.382 ± 0.038	0.380 ± 0.039
90	0.334 ± 0.063	0.342 ± 0.108	**0.401** ± 0.013	0.328 ± 0.029	0.331 ± 0.030
100	0.306 ± 0.017	0.327 ± 0.082	**0.388** ± 0.013	0.306 ± 0.019	0.307 ± 0.019

Table B.7: One-vs-one SVM classification accuracy on the Reuters multilingual corpus dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.766 ± 0.067	0.759 ± 0.250	0.790 ± 0.082	**0.846** ± 0.077	0.784 ± 0.024
3	0.738 ± 0.079	0.780 ± 0.206	0.821 ± 0.060	**0.882** ± 0.053	0.828 ± 0.028
4	0.735 ± 0.076	0.787 ± 0.197	0.833 ± 0.054	**0.899** ± 0.041	0.837 ± 0.049
5	0.751 ± 0.084	0.795 ± 0.185	0.831 ± 0.048	**0.905** ± 0.039	0.813 ± 0.040
6	0.740 ± 0.087	0.796 ± 0.184	0.833 ± 0.046	**0.913** ± 0.035	0.810 ± 0.030
7	0.742 ± 0.082	0.798 ± 0.177	0.840 ± 0.040	**0.915** ± 0.033	0.802 ± 0.021
8	0.743 ± 0.081	0.794 ± 0.175	0.840 ± 0.039	**0.917** ± 0.032	0.793 ± 0.022
9	0.740 ± 0.080	0.789 ± 0.171	0.834 ± 0.039	**0.918** ± 0.031	0.770 ± 0.023
10	0.740 ± 0.078	0.787 ± 0.165	0.829 ± 0.038	**0.920** ± 0.030	0.713 ± 0.026
15	0.713 ± 0.072	0.770 ± 0.155	0.812 ± 0.037	**0.923** ± 0.028	0.630 ± 0.032
20	0.696 ± 0.069	0.754 ± 0.150	0.797 ± 0.036	**0.924** ± 0.027	0.604 ± 0.032
25	0.685 ± 0.067	0.741 ± 0.145	0.786 ± 0.035	**0.925** ± 0.027	0.544 ± 0.029
30	0.669 ± 0.066	0.728 ± 0.144	0.775 ± 0.035	**0.926** ± 0.027	0.505 ± 0.030
40	0.645 ± 0.063	0.708 ± 0.138	0.750 ± 0.034	**0.926** ± 0.027	0.454 ± 0.027
50	0.624 ± 0.060	0.690 ± 0.135	0.732 ± 0.034	**0.927** ± 0.027	0.416 ± 0.022
60	0.604 ± 0.058	0.674 ± 0.136	0.718 ± 0.034	**0.927** ± 0.026	0.380 ± 0.019
70	0.589 ± 0.057	0.660 ± 0.135	0.705 ± 0.034	**0.927** ± 0.026	0.341 ± 0.015
80	0.573 ± 0.056	0.647 ± 0.135	0.693 ± 0.034	**0.927** ± 0.026	0.299 ± 0.015
90	0.561 ± 0.054	0.633 ± 0.134	0.680 ± 0.034	**0.928** ± 0.026	0.273 ± 0.016
100	0.549 ± 0.054	0.622 ± 0.133	0.668 ± 0.034	**0.928** ± 0.026	0.242 ± 0.018

Table B.8: Cophenetic correlation on the Reuters multilingual corpus dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.144 ± 0.028	0.179 ± 0.125	0.202 ± 0.050	**0.247** ± 0.060	0.206 ± 0.026
3	0.167 ± 0.036	0.214 ± 0.152	0.232 ± 0.057	**0.275** ± 0.055	0.257 ± 0.035
4	0.195 ± 0.047	0.242 ± 0.170	0.268 ± 0.067	**0.306** ± 0.056	0.296 ± 0.042
5	0.218 ± 0.056	0.259 ± 0.182	0.276 ± 0.068	**0.331** ± 0.059	0.314 ± 0.041
6	0.228 ± 0.061	0.270 ± 0.197	0.292 ± 0.077	**0.345** ± 0.058	0.321 ± 0.035
7	0.236 ± 0.065	0.282 ± 0.208	0.315 ± 0.080	**0.354** ± 0.057	0.322 ± 0.031
8	0.242 ± 0.069	0.288 ± 0.215	0.324 ± 0.083	**0.360** ± 0.056	0.321 ± 0.028
9	0.251 ± 0.074	0.293 ± 0.220	0.327 ± 0.084	**0.370** ± 0.058	0.322 ± 0.026
10	0.253 ± 0.076	0.295 ± 0.225	0.329 ± 0.086	**0.375** ± 0.057	0.323 ± 0.024
15	0.271 ± 0.088	0.304 ± 0.238	0.335 ± 0.090	**0.394** ± 0.057	0.310 ± 0.018
20	0.279 ± 0.095	0.307 ± 0.245	0.335 ± 0.093	**0.407** ± 0.058	0.299 ± 0.012
25	0.284 ± 0.099	0.310 ± 0.249	0.337 ± 0.095	**0.417** ± 0.060	0.289 ± 0.014
30	0.287 ± 0.103	0.311 ± 0.249	0.335 ± 0.095	**0.426** ± 0.062	0.284 ± 0.015
40	0.292 ± 0.106	0.310 ± 0.246	0.328 ± 0.092	**0.435** ± 0.063	0.267 ± 0.013
50	0.293 ± 0.109	0.307 ± 0.242	0.323 ± 0.091	**0.442** ± 0.064	0.257 ± 0.013
60	0.293 ± 0.110	0.305 ± 0.237	0.320 ± 0.090	**0.447** ± 0.065	0.245 ± 0.012
70	0.292 ± 0.111	0.302 ± 0.233	0.317 ± 0.088	**0.450** ± 0.066	0.237 ± 0.011
80	0.291 ± 0.111	0.299 ± 0.229	0.313 ± 0.087	**0.454** ± 0.066	0.229 ± 0.012
90	0.289 ± 0.110	0.297 ± 0.224	0.311 ± 0.086	**0.456** ± 0.066	0.224 ± 0.013
100	0.287 ± 0.109	0.295 ± 0.221	0.309 ± 0.085	**0.458** ± 0.066	0.216 ± 0.014

Table B.9: Area under the curve of the R_{NX} index on the Reuters multilingual corpus dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.505	0.534 ± 0.015	**0.545**	0.535	0.523
3	0.519	0.530 ± 0.020	**0.558**	0.492	0.534
4	0.499	0.525 ± 0.037	**0.599**	0.454	0.483
5	0.503	0.522 ± 0.034	**0.583**	0.450	0.485
6	0.499	0.530 ± 0.038	**0.600**	0.449	0.507
7	0.496	0.533 ± 0.041	**0.611**	0.449	0.517
8	0.509	0.538 ± 0.041	**0.618**	0.449	0.485
9	0.496	0.520 ± 0.051	**0.613**	0.449	0.483
10	0.499	0.532 ± 0.043	**0.616**	0.449	0.523
15	0.511	0.529 ± 0.024	**0.576**	0.449	0.487
20	0.492	0.532 ± 0.043	**0.617**	0.448	0.532
25	0.496	0.533 ± 0.042	**0.616**	0.448	0.493
30	0.508	0.533 ± 0.036	**0.604**	0.448	0.451
40	0.502	0.532 ± 0.038	**0.607**	0.448	0.392
50	0.536	0.539 ± 0.034	**0.605**	0.448	0.404
60	0.538	0.541 ± 0.033	**0.605**	0.448	0.418
70	0.533	0.544 ± 0.044	**0.631**	0.448	0.424
80	0.533	0.544 ± 0.044	**0.630**	0.448	0.443
90	0.533	0.544 ± 0.044	**0.630**	0.448	0.352
100	0.533	0.545 ± 0.043	**0.631**	0.448	0.359

Table B.10: Clustering purity on the Reuters multilingual corpus dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.191	0.234 ± 0.031	**0.284**	0.230	0.214
3	0.219	0.253 ± 0.045	**0.339**	0.200	0.248
4	0.212	0.255 ± 0.045	**0.340**	0.149	0.215
5	0.216	0.259 ± 0.053	**0.362**	0.150	0.223
6	0.231	0.271 ± 0.057	**0.377**	0.153	0.250
7	0.199	0.272 ± 0.063	**0.387**	0.153	0.280
8	0.210	0.281 ± 0.065	**0.404**	0.153	0.217
9	0.218	0.262 ± 0.076	**0.395**	0.153	0.238
10	0.218	0.277 ± 0.065	**0.401**	0.153	0.275
15	0.215	0.266 ± 0.044	**0.346**	0.153	0.219
20	0.196	0.265 ± 0.055	**0.365**	0.154	0.265
25	0.224	0.271 ± 0.049	**0.364**	0.153	0.254
30	0.212	0.263 ± 0.040	**0.333**	0.153	0.222
40	0.216	0.264 ± 0.040	**0.337**	0.152	0.138
50	0.251	0.270 ± 0.033	**0.335**	0.152	0.144
60	0.254	0.266 ± 0.037	**0.335**	0.153	0.195
70	0.250	0.275 ± 0.045	**0.364**	0.153	0.185
80	0.249	0.276 ± 0.043	**0.361**	0.153	0.174
90	0.249	0.277 ± 0.044	**0.364**	0.153	0.094
100	0.248	0.274 ± 0.046	**0.365**	0.153	0.101

Table B.11: Clustering normalized mutual information on the Reuters multilingual corpus dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	1.782 ± 0.047	1.748 ± 0.155	1.832 ± 0.074	**1.683** ± 0.049	**1.699** ± 0.028
3	1.687 ± 0.054	1.691 ± 0.114	**1.662** ± 0.052	**1.657** ± 0.065	**1.662** ± 0.041
4	1.661 ± 0.087	1.654 ± 0.135	1.638 ± 0.057	1.647 ± 0.073	**1.592** ± 0.063
5	1.674 ± 0.092	1.646 ± 0.147	1.594 ± 0.052	1.638 ± 0.073	**1.577** ± 0.070
6	1.672 ± 0.065	1.638 ± 0.152	1.617 ± 0.048	1.636 ± 0.073	**1.581** ± 0.065
7	1.670 ± 0.066	1.655 ± 0.133	1.613 ± 0.048	1.637 ± 0.073	**1.587** ± 0.054
8	1.666 ± 0.068	1.649 ± 0.123	**1.610** ± 0.049	1.637 ± 0.073	1.667 ± 0.042
9	1.671 ± 0.068	1.648 ± 0.127	**1.610** ± 0.055	1.636 ± 0.072	1.647 ± 0.053
10	1.652 ± 0.068	1.644 ± 0.131	**1.609** ± 0.067	1.636 ± 0.072	1.682 ± 0.041
15	1.677 ± 0.069	1.654 ± 0.141	**1.604** ± 0.071	1.636 ± 0.072	1.634 ± 0.048
20	1.674 ± 0.069	1.644 ± 0.141	**1.604** ± 0.071	1.635 ± 0.073	1.652 ± 0.046
25	1.674 ± 0.069	1.643 ± 0.138	**1.604** ± 0.071	1.635 ± 0.073	1.622 ± 0.058
30	1.673 ± 0.069	1.650 ± 0.139	**1.604** ± 0.071	1.635 ± 0.073	1.722 ± 0.009
40	1.672 ± 0.069	1.648 ± 0.136	**1.603** ± 0.071	1.635 ± 0.072	1.744 ± 0.006
50	1.673 ± 0.069	1.644 ± 0.137	**1.603** ± 0.070	1.635 ± 0.072	1.854 ± 0.006
60	1.672 ± 0.069	1.646 ± 0.143	**1.603** ± 0.070	1.635 ± 0.072	1.738 ± 0.007
70	1.673 ± 0.069	1.649 ± 0.140	**1.603** ± 0.070	1.635 ± 0.072	1.870 ± 0.004
80	1.672 ± 0.069	1.648 ± 0.137	**1.603** ± 0.070	1.635 ± 0.072	1.881 ± 0.012
90	1.672 ± 0.069	1.648 ± 0.136	**1.603** ± 0.070	1.635 ± 0.072	1.937 ± 0.010
100	1.672 ± 0.069	1.649 ± 0.140	**1.603** ± 0.070	1.635 ± 0.072	1.943 ± 0.008

Table B.12: Davies-Boulding index on the Reuters multilingual corpus dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.827 ± 0.011	0.828 ± 0.015	0.830 ± 0.010	**0.871** ± 0.006	**0.871** ± 0.006
3	0.926 ± 0.010	0.920 ± 0.017	0.914 ± 0.011	**0.949** ± 0.007	**0.949** ± 0.007
4	0.945 ± 0.009	0.941 ± 0.012	0.937 ± 0.006	**0.955** ± 0.007	**0.955** ± 0.007
5	0.942 ± 0.011	0.942 ± 0.015	0.941 ± 0.010	**0.960** ± 0.011	**0.960** ± 0.011
6	0.950 ± 0.009	0.946 ± 0.013	0.942 ± 0.007	**0.962** ± 0.008	**0.962** ± 0.007
7	0.951 ± 0.009	0.951 ± 0.011	0.951 ± 0.007	**0.966** ± 0.006	**0.967** ± 0.006
8	0.953 ± 0.008	0.954 ± 0.011	0.954 ± 0.007	**0.969** ± 0.006	**0.968** ± 0.006
9	0.953 ± 0.009	0.955 ± 0.011	0.957 ± 0.006	**0.968** ± 0.006	**0.969** ± 0.006
10	0.953 ± 0.009	0.955 ± 0.011	**0.956** ± 0.006	**0.965** ± 0.007	**0.965** ± 0.007
15	0.955 ± 0.006	0.954 ± 0.010	0.952 ± 0.008	**0.968** ± 0.006	**0.969** ± 0.006
20	0.955 ± 0.008	0.953 ± 0.012	0.950 ± 0.008	**0.967** ± 0.005	**0.967** ± 0.005
25	0.951 ± 0.007	0.949 ± 0.013	0.947 ± 0.010	**0.962** ± 0.004	**0.962** ± 0.004
30	0.943 ± 0.007	0.946 ± 0.012	0.948 ± 0.009	**0.959** ± 0.004	**0.959** ± 0.004
40	0.931 ± 0.011	0.935 ± 0.014	0.939 ± 0.006	**0.951** ± 0.007	**0.951** ± 0.007
50	0.914 ± 0.012	0.921 ± 0.018	0.928 ± 0.008	**0.942** ± 0.008	**0.942** ± 0.007
60	0.859 ± 0.014	0.871 ± 0.030	0.882 ± 0.020	**0.923** ± 0.013	**0.923** ± 0.013
70	0.737 ± 0.031	0.754 ± 0.052	0.771 ± 0.035	**0.864** ± 0.025	**0.863** ± 0.025
80	0.565 ± 0.044	0.587 ± 0.071	0.609 ± 0.046	**0.758** ± 0.037	**0.758** ± 0.036
90	0.421 ± 0.040	0.439 ± 0.064	0.456 ± 0.043	**0.615** ± 0.044	**0.615** ± 0.045
100	0.340 ± 0.025	0.352 ± 0.043	0.364 ± 0.031	**0.484** ± 0.038	**0.485** ± 0.038

Table B.13: One-vs-one SVM classification accuracy on the BBC segmented news dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.194 ± 0.006	0.193 ± 0.007	0.191 ± 0.003	**0.212** ± 0.001	**0.213** ± 0.001
3	0.211 ± 0.005	0.213 ± 0.006	0.214 ± 0.004	**0.243** ± 0.000	**0.246** ± 0.000
4	0.225 ± 0.006	0.228 ± 0.009	0.231 ± 0.005	**0.253** ± 0.001	**0.253** ± 0.000
5	0.226 ± 0.008	0.227 ± 0.010	0.228 ± 0.005	**0.253** ± 0.001	0.248 ± 0.001
6	0.225 ± 0.008	0.226 ± 0.010	0.227 ± 0.006	**0.249** ± 0.001	0.235 ± 0.001
7	0.209 ± 0.007	0.216 ± 0.014	0.222 ± 0.008	**0.246** ± 0.000	0.225 ± 0.000
8	0.207 ± 0.007	0.211 ± 0.013	0.216 ± 0.008	**0.235** ± 0.000	0.208 ± 0.001
9	0.209 ± 0.008	0.209 ± 0.012	0.208 ± 0.009	**0.227** ± 0.000	0.196 ± 0.001
10	0.200 ± 0.008	0.202 ± 0.013	0.203 ± 0.010	**0.228** ± 0.001	0.196 ± 0.002
15	0.192 ± 0.013	0.192 ± 0.019	0.192 ± 0.014	**0.220** ± 0.001	0.185 ± 0.001
20	0.185 ± 0.017	0.184 ± 0.024	0.184 ± 0.017	**0.209** ± 0.001	0.170 ± 0.001
25	0.182 ± 0.019	0.183 ± 0.027	0.184 ± 0.019	**0.211** ± 0.000	0.172 ± 0.000
30	0.182 ± 0.022	0.182 ± 0.030	0.181 ± 0.020	**0.208** ± 0.000	0.167 ± 0.000
40	0.182 ± 0.025	0.181 ± 0.034	0.179 ± 0.023	**0.205** ± 0.000	0.160 ± 0.001
50	0.180 ± 0.027	0.180 ± 0.038	0.179 ± 0.027	**0.202** ± 0.001	0.156 ± 0.001
60	0.181 ± 0.029	0.180 ± 0.042	0.180 ± 0.031	**0.203** ± 0.001	0.158 ± 0.001
70	0.183 ± 0.032	0.183 ± 0.047	0.183 ± 0.034	**0.207** ± 0.001	0.162 ± 0.002
80	0.188 ± 0.034	0.187 ± 0.051	0.187 ± 0.037	**0.211** ± 0.002	0.167 ± 0.002
90	0.192 ± 0.037	0.191 ± 0.055	0.190 ± 0.040	**0.216** ± 0.002	0.172 ± 0.003
100	0.197 ± 0.039	0.196 ± 0.059	0.196 ± 0.044	**0.220** ± 0.003	0.176 ± 0.004

Table B.14: Cophenetic correlation on the BBC segmented news dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.116 ± 0.008	0.115 ± 0.011	0.114 ± 0.007	**0.134** ± 0.000	**0.133** ± 0.000
3	0.136 ± 0.010	0.135 ± 0.015	0.135 ± 0.010	**0.161** ± 0.001	**0.160** ± 0.001
4	0.146 ± 0.012	0.146 ± 0.016	0.146 ± 0.012	**0.169** ± 0.001	**0.169** ± 0.001
5	0.166 ± 0.015	0.165 ± 0.021	0.164 ± 0.015	**0.190** ± 0.001	**0.189** ± 0.000
6	0.173 ± 0.017	0.174 ± 0.024	0.176 ± 0.017	**0.199** ± 0.000	**0.197** ± 0.000
7	0.176 ± 0.018	0.180 ± 0.027	0.184 ± 0.019	**0.207** ± 0.000	0.204 ± 0.000
8	0.182 ± 0.019	0.185 ± 0.029	0.188 ± 0.021	**0.211** ± 0.000	0.208 ± 0.000
9	0.187 ± 0.020	0.188 ± 0.030	0.190 ± 0.022	**0.213** ± 0.000	0.209 ± 0.001
10	0.189 ± 0.022	0.193 ± 0.032	0.196 ± 0.023	**0.219** ± 0.000	0.214 ± 0.000
15	0.207 ± 0.029	0.209 ± 0.042	0.212 ± 0.030	**0.240** ± 0.000	0.230 ± 0.000
20	0.217 ± 0.035	0.218 ± 0.049	0.219 ± 0.034	**0.247** ± 0.001	0.234 ± 0.001
25	0.224 ± 0.040	0.225 ± 0.055	0.226 ± 0.038	**0.254** ± 0.000	0.241 ± 0.001
30	0.226 ± 0.042	0.228 ± 0.059	0.230 ± 0.041	**0.259** ± 0.000	0.244 ± 0.000
40	0.235 ± 0.048	0.235 ± 0.067	0.235 ± 0.046	**0.264** ± 0.000	0.247 ± 0.001
50	0.239 ± 0.052	0.240 ± 0.074	0.241 ± 0.052	**0.268** ± 0.001	0.250 ± 0.001
60	0.243 ± 0.056	0.245 ± 0.080	0.247 ± 0.058	**0.273** ± 0.000	0.253 ± 0.000
70	0.247 ± 0.060	0.250 ± 0.087	0.252 ± 0.062	**0.277** ± 0.001	0.256 ± 0.001
80	0.253 ± 0.064	0.255 ± 0.092	0.256 ± 0.067	**0.282** ± 0.000	0.259 ± 0.000
90	0.257 ± 0.068	0.258 ± 0.098	0.259 ± 0.071	**0.286** ± 0.000	0.264 ± 0.000
100	0.260 ± 0.071	0.262 ± 0.104	0.264 ± 0.075	**0.290** ± 0.000	0.268 ± 0.000

Table B.15: Area under the curve of the R_{NX} index on the BBC segmented news dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.529	0.553 ± 0.024	0.577	**0.697**	**0.691**
3	0.617	0.652 ± 0.034	0.686	**0.857**	**0.862**
4	0.856	0.869 ± 0.013	0.882	**0.914**	**0.921**
5	0.817	0.820 ± 0.002	0.822	**0.875**	0.689
6	0.699	0.708 ± 0.009	0.716	**0.736**	0.715
7	0.706	0.695 ± 0.011	0.683	**0.751**	0.567
8	0.581	0.653 ± 0.073	0.726	**0.752**	0.552
9	0.577	0.652 ± 0.075	0.727	**0.746**	0.565
10	0.581	0.654 ± 0.073	0.727	**0.744**	0.553
15	0.584	0.659 ± 0.074	0.733	**0.746**	0.443
20	0.735	0.728 ± 0.007	0.720	**0.747**	0.526
25	0.732	0.726 ± 0.006	0.720	**0.746**	0.432
30	0.737	0.729 ± 0.007	0.722	**0.746**	0.427
40	0.736	0.737 ± 0.001	**0.738**	**0.745**	0.384
50	0.733	0.735 ± 0.001	0.736	**0.749**	0.379
60	0.734	0.735 ± 0.001	0.737	**0.747**	0.414
70	**0.739**	0.738 ± 0.000	0.738	**0.746**	0.338
80	0.736	0.729 ± 0.007	0.722	**0.746**	0.384
90	0.736	0.729 ± 0.006	0.723	**0.745**	0.323
100	0.734	0.728 ± 0.006	0.722	**0.745**	0.415

Table B.16: Clustering purity on the BBC segmented news dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.484	0.476 ± 0.008	0.468	**0.572**	0.565
3	0.540	0.555 ± 0.015	0.570	**0.716**	**0.722**
4	0.696	0.705 ± 0.009	0.715	0.778	**0.788**
5	0.651	0.652 ± 0.001	0.653	**0.732**	0.562
6	0.606	0.614 ± 0.008	0.622	**0.672**	0.637
7	0.637	0.619 ± 0.018	0.601	**0.704**	0.483
8	0.487	0.572 ± 0.084	0.656	**0.715**	0.472
9	0.488	0.574 ± 0.087	0.661	**0.703**	0.490
10	0.489	0.577 ± 0.088	0.665	**0.699**	0.463
15	0.496	0.590 ± 0.094	0.683	**0.703**	0.314
20	0.686	0.667 ± 0.019	0.647	**0.707**	0.422
25	0.678	0.664 ± 0.014	0.650	**0.706**	0.290
30	0.689	0.672 ± 0.018	0.654	**0.706**	0.251
40	0.687	0.690 ± 0.002	0.692	**0.705**	0.213
50	0.681	0.685 ± 0.004	0.690	**0.714**	0.201
60	0.683	0.687 ± 0.004	0.692	**0.710**	0.232
70	0.693	0.694 ± 0.001	0.695	**0.708**	0.109
80	0.686	0.671 ± 0.015	0.656	**0.707**	0.169
90	0.686	0.672 ± 0.014	0.658	**0.706**	0.147
100	0.685	0.671 ± 0.014	0.657	**0.706**	0.260

Table B.17: Clustering normalized mutual information on the BBC segmented news dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	**1.983** ± 0.002	**1.983** ± 0.003	**1.983** ± 0.002	**1.979** ± 0.000	**1.979** ± 0.000
3	**1.979** ± 0.002	**1.979** ± 0.003	**1.979** ± 0.002	**1.972** ± 0.000	**1.972** ± 0.000
4	**1.971** ± 0.001	**1.972** ± 0.002	**1.972** ± 0.001	**1.970** ± 0.000	**1.970** ± 0.000
5	**1.972** ± 0.002	**1.972** ± 0.002	**1.972** ± 0.001	**1.971** ± 0.000	**1.976** ± 0.000
6	**1.971** ± 0.002	**1.971** ± 0.002	**1.970** ± 0.001	**1.970** ± 0.000	**1.969** ± 0.000
7	**1.971** ± 0.004	**1.969** ± 0.005	**1.967** ± 0.001	**1.970** ± 0.000	**1.967** ± 0.001
8	**1.969** ± 0.004	**1.965** ± 0.007	**1.962** ± 0.001	**1.969** ± 0.001	**1.960** ± 0.001
9	**1.967** ± 0.004	**1.964** ± 0.006	**1.961** ± 0.002	**1.967** ± 0.000	**1.957** ± 0.002
10	**1.967** ± 0.004	**1.965** ± 0.005	**1.963** ± 0.002	**1.967** ± 0.000	**1.958** ± 0.002
15	**1.964** ± 0.005	**1.962** ± 0.006	**1.960** ± 0.003	**1.965** ± 0.001	**1.946** ± 0.003
20	**1.965** ± 0.003	**1.965** ± 0.005	**1.966** ± 0.003	**1.963** ± 0.000	**1.957** ± 0.001
25	**1.964** ± 0.003	**1.965** ± 0.005	**1.965** ± 0.003	**1.962** ± 0.000	**1.950** ± 0.001
30	**1.964** ± 0.003	**1.965** ± 0.005	**1.965** ± 0.003	**1.962** ± 0.000	**1.956** ± 0.001
40	**1.964** ± 0.003	**1.965** ± 0.005	**1.965** ± 0.003	**1.962** ± 0.000	**1.961** ± 0.001
50	**1.964** ± 0.003	**1.965** ± 0.005	**1.965** ± 0.003	**1.962** ± 0.000	**1.959** ± 0.001
60	**1.964** ± 0.003	**1.965** ± 0.005	**1.965** ± 0.003	**1.962** ± 0.000	**1.959** ± 0.001
70	**1.963** ± 0.003	**1.964** ± 0.005	**1.965** ± 0.003	**1.962** ± 0.000	**1.964** ± 0.001
80	**1.962** ± 0.003	**1.964** ± 0.005	**1.965** ± 0.003	**1.962** ± 0.000	**1.961** ± 0.002
90	**1.962** ± 0.003	**1.964** ± 0.005	**1.965** ± 0.003	**1.962** ± 0.000	**1.957** ± 0.002
100	**1.962** ± 0.003	**1.964** ± 0.005	**1.965** ± 0.003	**1.962** ± 0.000	**1.955** ± 0.002

Table B.18: Davies-Boulding index on the BBC segmented news dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.063 ± 0.007	0.056 ± 0.025	**0.065** ± 0.007	0.063 ± 0.007	0.062 ± 0.007
3	0.069 ± 0.006	0.064 ± 0.023	**0.077** ± 0.006	0.069 ± 0.007	0.066 ± 0.006
4	0.077 ± 0.014	0.072 ± 0.031	**0.089** ± 0.006	0.077 ± 0.007	0.080 ± 0.007
5	0.047 ± 0.012	0.070 ± 0.041	**0.094** ± 0.006	0.060 ± 0.025	0.088 ± 0.008
6	0.038 ± 0.005	0.067 ± 0.052	**0.097** ± 0.010	0.044 ± 0.017	0.065 ± 0.027
7	0.037 ± 0.004	0.065 ± 0.060	**0.096** ± 0.022	0.038 ± 0.005	0.055 ± 0.031
8	0.037 ± 0.004	0.051 ± 0.049	**0.062** ± 0.029	0.038 ± 0.004	0.043 ± 0.023
9	0.037 ± 0.004	**0.044** ± 0.034	0.041 ± 0.004	0.037 ± 0.004	0.043 ± 0.025
10	0.036 ± 0.004	0.045 ± 0.047	**0.056** ± 0.037	0.037 ± 0.004	0.043 ± 0.023
15	0.038 ± 0.008	0.052 ± 0.086	**0.121** ± 0.039	0.040 ± 0.007	0.045 ± 0.013
20	0.038 ± 0.008	0.051 ± 0.080	**0.096** ± 0.057	0.050 ± 0.011	0.060 ± 0.018
25	0.037 ± 0.005	0.054 ± 0.082	**0.098** ± 0.055	0.048 ± 0.007	0.070 ± 0.030
30	0.042 ± 0.008	0.053 ± 0.080	0.094 ± 0.055	0.059 ± 0.036	**0.104** ± 0.055
40	0.044 ± 0.016	0.059 ± 0.090	0.104 ± 0.047	0.083 ± 0.027	**0.145** ± 0.038
50	0.040 ± 0.009	0.061 ± 0.082	0.102 ± 0.027	0.074 ± 0.017	**0.125** ± 0.014
60	0.047 ± 0.012	0.071 ± 0.069	0.095 ± 0.009	0.063 ± 0.013	**0.101** ± 0.014
70	0.044 ± 0.010	0.065 ± 0.049	0.076 ± 0.012	0.056 ± 0.010	**0.088** ± 0.010
80	0.040 ± 0.007	0.059 ± 0.039	0.064 ± 0.009	0.053 ± 0.006	**0.072** ± 0.010
90	0.039 ± 0.006	0.053 ± 0.032	0.054 ± 0.010	0.049 ± 0.007	**0.060** ± 0.008
100	0.042 ± 0.008	0.049 ± 0.025	0.045 ± 0.005	0.047 ± 0.006	**0.053** ± 0.008

Table B.19: One-vs-one SVM classification accuracy on the animal with attributes (AWA) dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.161 ± 0.292	0.207 ± 0.596	0.282 ± 0.277	0.283 ± 0.143	**0.391** ± 0.169
3	0.177 ± 0.308	0.242 ± 0.625	0.319 ± 0.272	0.289 ± 0.159	**0.404** ± 0.160
4	0.175 ± 0.320	0.261 ± 0.653	0.340 ± 0.285	0.306 ± 0.176	**0.410** ± 0.175
5	0.178 ± 0.325	0.268 ± 0.664	0.350 ± 0.286	0.317 ± 0.186	**0.405** ± 0.188
6	0.182 ± 0.326	0.272 ± 0.674	0.353 ± 0.293	0.318 ± 0.192	**0.410** ± 0.184
7	0.184 ± 0.329	0.278 ± 0.683	0.357 ± 0.296	0.333 ± 0.189	**0.422** ± 0.184
8	0.190 ± 0.332	0.285 ± 0.691	0.361 ± 0.296	0.346 ± 0.192	**0.430** ± 0.183
9	0.192 ± 0.335	0.291 ± 0.698	0.366 ± 0.296	0.348 ± 0.199	**0.439** ± 0.183
10	0.195 ± 0.336	0.294 ± 0.704	0.368 ± 0.298	0.353 ± 0.197	**0.441** ± 0.180
15	0.200 ± 0.341	0.310 ± 0.722	0.375 ± 0.297	0.375 ± 0.215	**0.461** ± 0.172
20	0.204 ± 0.343	0.318 ± 0.733	0.379 ± 0.296	0.391 ± 0.229	**0.471** ± 0.169
25	0.207 ± 0.345	0.324 ± 0.742	0.381 ± 0.296	0.400 ± 0.239	**0.474** ± 0.164
30	0.209 ± 0.346	0.327 ± 0.748	0.382 ± 0.296	0.408 ± 0.245	**0.483** ± 0.161
40	0.214 ± 0.347	0.333 ± 0.757	0.383 ± 0.295	0.420 ± 0.257	**0.494** ± 0.154
50	0.217 ± 0.348	0.338 ± 0.763	0.385 ± 0.294	0.427 ± 0.264	**0.501** ± 0.143
60	0.218 ± 0.349	0.340 ± 0.766	0.386 ± 0.294	0.434 ± 0.272	**0.509** ± 0.132
70	0.220 ± 0.349	0.342 ± 0.769	0.386 ± 0.294	0.439 ± 0.278	**0.516** ± 0.120
80	0.221 ± 0.350	0.344 ± 0.771	0.387 ± 0.293	0.443 ± 0.282	**0.519** ± 0.113
90	0.222 ± 0.350	0.345 ± 0.774	0.387 ± 0.293	0.446 ± 0.285	**0.520** ± 0.107
100	0.222 ± 0.351	0.346 ± 0.775	0.387 ± 0.293	0.448 ± 0.288	**0.520** ± 0.101

Table B.20: Cophenetic correlation on the animal with attributes (AWA) dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.029 ± 0.028	0.042 ± 0.083	0.054 ± 0.038	0.044 ± 0.017	**0.060** ± 0.030
3	0.036 ± 0.039	0.053 ± 0.110	**0.064** ± 0.053	0.051 ± 0.024	**0.065** ± 0.031
4	0.047 ± 0.049	0.062 ± 0.137	**0.084** ± 0.076	0.058 ± 0.031	0.072 ± 0.030
5	0.050 ± 0.054	0.067 ± 0.153	**0.090** ± 0.083	0.064 ± 0.037	0.079 ± 0.030
6	0.054 ± 0.059	0.071 ± 0.165	**0.092** ± 0.085	0.068 ± 0.042	0.084 ± 0.031
7	0.058 ± 0.067	0.077 ± 0.184	**0.102** ± 0.097	0.074 ± 0.043	0.090 ± 0.030
8	0.062 ± 0.073	0.084 ± 0.204	**0.114** ± 0.110	0.080 ± 0.048	0.095 ± 0.029
9	0.064 ± 0.076	0.090 ± 0.221	**0.124** ± 0.122	0.083 ± 0.052	0.099 ± 0.029
10	0.067 ± 0.081	0.092 ± 0.230	**0.125** ± 0.124	0.086 ± 0.054	0.102 ± 0.030
15	0.080 ± 0.100	0.110 ± 0.294	**0.160** ± 0.172	0.099 ± 0.065	0.117 ± 0.028
20	0.088 ± 0.113	0.121 ± 0.339	**0.182** ± 0.203	0.109 ± 0.076	0.124 ± 0.026
25	0.094 ± 0.123	0.128 ± 0.368	**0.191** ± 0.216	0.116 ± 0.085	0.130 ± 0.024
30	0.100 ± 0.134	0.134 ± 0.392	**0.197** ± 0.225	0.123 ± 0.092	0.135 ± 0.025
40	0.109 ± 0.150	0.145 ± 0.432	**0.207** ± 0.240	0.132 ± 0.104	0.142 ± 0.025
50	0.116 ± 0.162	0.152 ± 0.463	**0.215** ± 0.251	0.139 ± 0.114	0.147 ± 0.026
60	0.123 ± 0.174	0.158 ± 0.488	**0.220** ± 0.258	0.145 ± 0.123	0.152 ± 0.026
70	0.128 ± 0.183	0.163 ± 0.508	**0.224** ± 0.264	0.149 ± 0.129	0.155 ± 0.027
80	0.133 ± 0.192	0.168 ± 0.527	**0.227** ± 0.268	0.153 ± 0.134	0.157 ± 0.027
90	0.137 ± 0.199	0.171 ± 0.542	**0.229** ± 0.272	0.156 ± 0.138	0.158 ± 0.027
100	0.141 ± 0.206	0.174 ± 0.556	**0.231** ± 0.274	0.158 ± 0.142	0.159 ± 0.027

Table B.21: Area under the curve of the R_{NX} index on the animal with attributes (AWA) dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.086	0.090 ± 0.004	**0.099**	0.095	**0.098**
3	0.090	0.096 ± 0.005	**0.103**	0.098	0.098
4	0.095	0.097 ± 0.004	0.103	0.100	**0.106**
5	0.094	0.096 ± 0.005	**0.106**	0.104	**0.105**
6	0.090	0.098 ± 0.006	0.110	0.101	**0.113**
7	0.097	0.102 ± 0.006	**0.115**	0.106	0.111
8	0.096	0.102 ± 0.006	**0.115**	0.107	0.113
9	0.096	0.102 ± 0.006	**0.116**	0.109	**0.115**
10	0.095	0.103 ± 0.006	**0.116**	0.111	**0.116**
15	0.095	0.103 ± 0.008	**0.120**	0.109	0.115
20	0.097	0.105 ± 0.007	**0.120**	0.110	0.112
25	0.098	0.106 ± 0.006	**0.118**	0.110	0.108
30	0.095	0.106 ± 0.008	**0.120**	0.112	0.113
40	0.097	0.106 ± 0.007	**0.119**	0.112	**0.120**
50	0.100	0.106 ± 0.007	**0.120**	0.110	0.118
60	0.096	0.105 ± 0.006	**0.114**	0.110	**0.114**
70	0.098	0.106 ± 0.007	**0.119**	0.111	0.114
80	0.096	0.105 ± 0.008	**0.122**	0.108	0.114
90	0.097	0.106 ± 0.008	**0.122**	0.111	0.117
100	0.096	0.104 ± 0.007	**0.117**	0.111	0.112

Table B.22: Clustering purity on the animal with attributes (AWA) dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.113	0.123 ± 0.009	**0.142**	0.129	0.136
3	0.121	0.129 ± 0.011	**0.151**	0.134	0.138
4	0.131	0.134 ± 0.008	**0.151**	0.140	0.149
5	0.135	0.136 ± 0.008	0.154	0.145	**0.157**
6	0.134	0.136 ± 0.007	0.152	0.144	**0.159**
7	0.137	0.141 ± 0.007	0.156	0.146	**0.158**
8	0.140	0.143 ± 0.007	0.159	0.149	**0.164**
9	0.134	0.143 ± 0.009	**0.162**	0.149	**0.162**
10	0.138	0.143 ± 0.007	0.157	0.150	**0.166**
15	0.138	0.144 ± 0.007	0.159	0.151	**0.166**
20	0.136	0.145 ± 0.007	0.160	0.152	**0.166**
25	0.134	0.144 ± 0.008	**0.159**	0.152	**0.158**
30	0.136	0.145 ± 0.008	**0.161**	0.150	0.157
40	0.136	0.145 ± 0.008	**0.160**	0.150	**0.161**
50	0.138	0.145 ± 0.008	**0.161**	0.150	0.157
60	0.133	0.143 ± 0.009	**0.161**	0.151	0.156
70	0.135	0.145 ± 0.008	**0.159**	0.150	0.151
80	0.135	0.143 ± 0.008	**0.159**	0.148	0.148
90	0.136	0.144 ± 0.008	**0.161**	0.148	0.151
100	0.137	0.144 ± 0.007	**0.157**	0.147	**0.155**

Table B.23: Clustering normalized mutual information on the animal with attributes (AWA) dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	**1.993** ± 0.060	**2.000** ± 0.079	**1.994** ± 0.008	**1.991** ± 0.006	**1.984** ± 0.014
3	**1.986** ± 0.061	**1.987** ± 0.099	**1.993** ± 0.017	**1.988** ± 0.009	**1.975** ± 0.017
4	**1.982** ± 0.062	**1.980** ± 0.109	**1.978** ± 0.023	**1.986** ± 0.011	**1.968** ± 0.019
5	**1.979** ± 0.060	**1.979** ± 0.113	**1.983** ± 0.026	**1.986** ± 0.011	**1.983** ± 0.033
6	1.979 ± 0.083	1.974 ± 0.140	**1.947** ± 0.028	1.984 ± 0.011	1.981 ± 0.030
7	1.976 ± 0.100	1.966 ± 0.162	**1.921** ± 0.030	1.983 ± 0.012	1.976 ± 0.037
8	1.976 ± 0.109	1.967 ± 0.171	**1.912** ± 0.032	1.979 ± 0.015	1.956 ± 0.025
9	1.975 ± 0.119	1.961 ± 0.182	**1.906** ± 0.034	1.978 ± 0.020	1.952 ± 0.031
10	1.974 ± 0.126	1.962 ± 0.188	**1.906** ± 0.034	1.978 ± 0.018	1.949 ± 0.033
15	1.972 ± 0.140	1.959 ± 0.211	**1.906** ± 0.037	1.975 ± 0.019	1.947 ± 0.038
20	1.972 ± 0.140	1.954 ± 0.218	**1.905** ± 0.039	1.973 ± 0.021	1.949 ± 0.033
25	1.972 ± 0.143	1.955 ± 0.228	**1.904** ± 0.040	1.971 ± 0.029	**1.923** ± 0.044
30	1.972 ± 0.140	1.951 ± 0.224	**1.905** ± 0.040	1.968 ± 0.031	**1.919** ± 0.047
40	1.970 ± 0.141	1.954 ± 0.231	1.904 ± 0.042	1.968 ± 0.038	**1.873** ± 0.052
50	1.969 ± 0.142	1.945 ± 0.231	**1.903** ± 0.044	1.967 ± 0.041	**1.895** ± 0.049
60	1.971 ± 0.141	1.946 ± 0.234	**1.905** ± 0.045	1.966 ± 0.039	**1.918** ± 0.055
70	1.971 ± 0.140	1.946 ± 0.231	**1.905** ± 0.046	1.963 ± 0.042	**1.905** ± 0.053
80	1.972 ± 0.157	1.948 ± 0.241	1.908 ± 0.046	1.963 ± 0.045	**1.877** ± 0.054
90	1.972 ± 0.147	1.943 ± 0.239	1.906 ± 0.046	1.964 ± 0.054	**1.881** ± 0.074
100	1.972 ± 0.141	1.947 ± 0.231	**1.904** ± 0.045	1.963 ± 0.047	**1.902** ± 0.064

Table B.24: Davies-Boulding index on the animal with attributes (AWA) dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
1	0.713 ± 0.023	0.740 ± 0.077	0.708 ± 0.016	0.703 ± 0.012	**0.793** ± 0.011
2	0.718 ± 0.024	0.746 ± 0.086	0.708 ± 0.014	0.703 ± 0.013	**0.808** ± 0.014
3	0.721 ± 0.026	0.755 ± 0.087	0.723 ± 0.012	0.701 ± 0.011	**0.816** ± 0.015
4	0.000 ± 0.021	0.239 ± 0.587	0.718 ± 0.015	0.704 ± 0.012	**0.821** ± 0.015
5	0.000 ± 0.023	0.251 ± 0.617	0.754 ± 0.012	0.712 ± 0.010	**0.838** ± 0.012
6	0.000 ± 0.020	0.248 ± 0.607	0.743 ± 0.013	0.710 ± 0.009	**0.834** ± 0.012
7	0.000 ± 0.021	0.244 ± 0.599	0.733 ± 0.015	0.710 ± 0.009	**0.841** ± 0.007
8	0.000 ± 0.017	0.245 ± 0.602	0.736 ± 0.018	0.709 ± 0.011	**0.838** ± 0.006
9	0.000 ± 0.023	0.247 ± 0.607	0.742 ± 0.016	0.709 ± 0.012	**0.840** ± 0.008
10	0.000 ± 0.019	0.250 ± 0.613	0.750 ± 0.013	0.714 ± 0.009	**0.856** ± 0.012
11	0.000 ± 0.016	0.250 ± 0.614	0.751 ± 0.015	0.727 ± 0.013	**0.855** ± 0.014
12	0.000 ± 0.017	0.252 ± 0.618	0.756 ± 0.022	0.738 ± 0.010	**0.832** ± 0.012
13	0.000 ± 0.016	0.253 ± 0.621	0.760 ± 0.023	0.745 ± 0.016	**0.795** ± 0.012
14	0.000 ± 0.014	0.256 ± 0.627	**0.768** ± 0.019	**0.770** ± 0.020	0.739 ± 0.020
15	0.000 ± 0.016	0.257 ± 0.630	0.771 ± 0.021	**0.790** ± 0.017	0.709 ± 0.018
16	0.000 ± 0.014	0.256 ± 0.628	0.768 ± 0.012	**0.795** ± 0.015	0.705 ± 0.016
17	0.000 ± 0.014	0.256 ± 0.627	0.767 ± 0.012	**0.816** ± 0.015	0.705 ± 0.018
18	0.000 ± 0.015	0.255 ± 0.624	0.764 ± 0.018	**0.812** ± 0.015	0.703 ± 0.017
19	0.000 ± 0.015	0.255 ± 0.625	0.764 ± 0.022	**0.808** ± 0.015	0.703 ± 0.017
20	0.000 ± 0.015	0.255 ± 0.624	0.764 ± 0.019	**0.801** ± 0.015	0.703 ± 0.017

Table B.25: One-vs-one SVM classification accuracy on the Berkeley protein dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.215 ± 0.293	0.308 ± 0.655	0.382 ± 0.413	**0.438** ± 0.304	0.408 ± 0.281
3	0.288 ± 0.353	0.326 ± 0.665	0.370 ± 0.408	**0.468** ± 0.311	0.406 ± 0.273
4	0.292 ± 0.362	0.324 ± 0.669	0.367 ± 0.409	**0.472** ± 0.307	0.407 ± 0.267
5	0.279 ± 0.357	0.323 ± 0.664	0.365 ± 0.409	**0.466** ± 0.304	0.354 ± 0.236
6	0.303 ± 0.363	0.329 ± 0.664	0.364 ± 0.410	**0.463** ± 0.306	0.307 ± 0.215
7	0.299 ± 0.373	0.329 ± 0.666	0.363 ± 0.409	**0.460** ± 0.307	0.279 ± 0.220
8	0.297 ± 0.372	0.327 ± 0.664	0.361 ± 0.408	**0.459** ± 0.309	0.247 ± 0.210
9	0.300 ± 0.369	0.327 ± 0.661	0.359 ± 0.408	**0.456** ± 0.309	0.228 ± 0.211
10	0.300 ± 0.369	0.327 ± 0.661	0.358 ± 0.407	**0.455** ± 0.309	0.228 ± 0.215
15	0.307 ± 0.403	0.325 ± 0.675	0.348 ± 0.404	**0.448** ± 0.313	0.165 ± 0.220
20	0.284 ± 0.392	0.314 ± 0.666	0.339 ± 0.400	**0.442** ± 0.315	0.130 ± 0.213
25	0.271 ± 0.385	0.306 ± 0.659	0.331 ± 0.397	**0.436** ± 0.318	0.108 ± 0.210
30	0.267 ± 0.370	0.301 ± 0.648	0.325 ± 0.394	**0.432** ± 0.319	0.091 ± 0.206
40	0.244 ± 0.354	0.289 ± 0.633	0.314 ± 0.389	**0.423** ± 0.320	0.065 ± 0.202
50	0.242 ± 0.326	0.286 ± 0.615	0.308 ± 0.385	**0.416** ± 0.319	0.050 ± 0.203
60	0.238 ± 0.312	0.282 ± 0.605	0.302 ± 0.382	**0.415** ± 0.311	0.040 ± 0.203
70	0.233 ± 0.294	0.279 ± 0.593	0.297 ± 0.378	**0.410** ± 0.308	0.040 ± 0.206
80	0.234 ± 0.284	0.278 ± 0.586	0.292 ± 0.374	**0.407** ± 0.304	0.041 ± 0.207
90	0.243 ± 0.271	0.280 ± 0.578	0.287 ± 0.370	**0.403** ± 0.302	0.043 ± 0.208
100	0.250 ± 0.262	0.282 ± 0.572	0.284 ± 0.369	**0.399** ± 0.301	0.046 ± 0.210

Table B.26: Cophenetic correlation on the Berkeley protein dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.143 ± 0.180	0.171 ± 0.293	**0.223** ± 0.204	0.164 ± 0.053	0.150 ± 0.040
3	0.157 ± 0.190	0.188 ± 0.320	**0.242** ± 0.226	0.185 ± 0.060	0.157 ± 0.043
4	0.158 ± 0.193	0.189 ± 0.316	**0.240** ± 0.219	0.195 ± 0.053	0.179 ± 0.052
5	0.170 ± 0.202	0.194 ± 0.319	**0.240** ± 0.216	0.203 ± 0.054	0.174 ± 0.047
6	0.173 ± 0.203	0.196 ± 0.318	**0.237** ± 0.211	0.205 ± 0.054	0.165 ± 0.043
7	0.179 ± 0.202	0.198 ± 0.315	**0.235** ± 0.208	0.206 ± 0.057	0.159 ± 0.043
8	0.182 ± 0.205	0.199 ± 0.318	**0.235** ± 0.208	0.206 ± 0.058	0.153 ± 0.041
9	0.180 ± 0.202	0.200 ± 0.316	**0.235** ± 0.206	0.208 ± 0.060	0.150 ± 0.042
10	0.183 ± 0.204	0.201 ± 0.317	**0.232** ± 0.203	0.209 ± 0.062	0.150 ± 0.042
15	0.180 ± 0.196	0.204 ± 0.313	**0.227** ± 0.196	0.212 ± 0.062	0.136 ± 0.044
20	0.176 ± 0.191	0.204 ± 0.312	**0.222** ± 0.187	0.213 ± 0.061	0.125 ± 0.041
25	0.173 ± 0.186	0.203 ± 0.312	**0.218** ± 0.182	0.214 ± 0.058	0.115 ± 0.039
30	0.171 ± 0.184	0.201 ± 0.309	**0.215** ± 0.178	**0.214** ± 0.057	0.105 ± 0.037
40	0.167 ± 0.178	0.199 ± 0.310	0.206 ± 0.168	**0.212** ± 0.054	0.090 ± 0.032
50	0.163 ± 0.173	0.196 ± 0.303	0.202 ± 0.161	**0.211** ± 0.053	0.080 ± 0.030
60	0.161 ± 0.169	0.190 ± 0.293	0.197 ± 0.156	**0.210** ± 0.051	0.072 ± 0.027
70	0.160 ± 0.166	0.187 ± 0.288	0.194 ± 0.152	**0.209** ± 0.049	0.066 ± 0.025
80	0.159 ± 0.164	0.179 ± 0.276	0.189 ± 0.147	**0.208** ± 0.047	0.063 ± 0.024
90	0.157 ± 0.162	0.175 ± 0.269	0.186 ± 0.143	**0.207** ± 0.045	0.060 ± 0.021
100	0.156 ± 0.160	0.170 ± 0.262	0.183 ± 0.140	**0.205** ± 0.044	0.058 ± 0.020

Table B.27: Area under the curve of the R_{NX} index on the Berkeley protein dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.700	0.720 ± 0.028	0.759	**0.781**	0.771
3	0.700	0.720 ± 0.028	0.759	**0.784**	0.770
4	0.700	0.720 ± 0.028	0.760	**0.785**	0.765
5	0.700	0.720 ± 0.028	0.759	**0.785**	0.700
6	0.700	0.720 ± 0.029	0.761	**0.785**	0.700
7	0.700	0.721 ± 0.029	0.762	**0.787**	0.700
8	0.700	0.721 ± 0.029	0.762	**0.788**	0.700
9	0.700	0.721 ± 0.029	0.762	**0.788**	0.700
10	0.700	0.721 ± 0.029	0.762	**0.788**	0.700
15	0.700	0.721 ± 0.030	0.764	**0.788**	0.700
20	0.700	0.721 ± 0.030	0.764	**0.788**	0.700
25	0.700	0.720 ± 0.029	0.761	**0.788**	0.700
30	0.700	0.720 ± 0.029	0.761	**0.788**	0.700
40	0.700	0.720 ± 0.029	0.761	**0.788**	0.700
50	0.700	0.720 ± 0.029	0.761	**0.788**	0.700
60	0.700	0.720 ± 0.029	0.761	0.788	**0.812**
70	0.700	0.720 ± 0.029	0.761	**0.788**	0.700
80	0.700	0.720 ± 0.029	0.761	**0.788**	0.700
90	0.700	0.720 ± 0.029	0.761	**0.788**	0.700
100	0.700	0.720 ± 0.029	0.761	0.788	**0.807**

Table B.28: Clustering purity on the Berkeley protein dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.091	0.163 ± 0.076	0.268	**0.311**	0.306
3	0.021	0.140 ± 0.102	0.270	**0.315**	0.305
4	0.019	0.140 ± 0.103	0.271	**0.319**	0.292
5	0.017	0.138 ± 0.103	0.268	**0.320**	0.116
6	0.017	0.139 ± 0.104	0.270	**0.314**	0.117
7	0.017	0.139 ± 0.104	0.271	**0.319**	0.109
8	0.017	0.139 ± 0.104	0.271	**0.320**	0.109
9	0.017	0.139 ± 0.104	0.271	**0.320**	0.116
10	0.017	0.139 ± 0.104	0.271	**0.320**	0.125
15	0.017	0.140 ± 0.105	0.273	**0.320**	0.094
20	0.017	0.140 ± 0.105	0.273	**0.320**	0.106
25	0.017	0.138 ± 0.104	0.270	**0.320**	0.085
30	0.017	0.139 ± 0.104	0.270	**0.320**	0.156
40	0.017	0.139 ± 0.104	0.270	**0.320**	0.121
50	0.017	0.139 ± 0.103	0.270	**0.320**	0.118
60	0.017	0.138 ± 0.103	0.269	0.320	**0.392**
70	0.017	0.139 ± 0.104	0.270	**0.320**	0.045
80	0.017	0.139 ± 0.104	0.270	**0.320**	0.024
90	0.017	0.139 ± 0.103	0.270	**0.320**	0.040
100	0.017	0.139 ± 0.103	0.269	0.320	**0.371**

Table B.29: Clustering normalized mutual information on the Berkeley protein dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	1.747 ± 0.212	**1.648** ± 0.599	**1.636** ± 0.258	1.693 ± 0.161	1.713 ± 0.146
3	1.750 ± 0.309	1.614 ± 0.645	**1.531** ± 0.251	1.693 ± 0.159	1.713 ± 0.145
4	1.751 ± 0.489	1.564 ± 0.774	**1.382** ± 0.250	1.695 ± 0.158	1.715 ± 0.148
5	1.751 ± 0.645	1.512 ± 0.921	**1.225** ± 0.245	1.695 ± 0.160	1.637 ± 0.385
6	1.752 ± 0.645	1.512 ± 0.921	**1.225** ± 0.246	1.694 ± 0.161	1.623 ± 0.406
7	1.753 ± 0.645	1.513 ± 0.921	**1.225** ± 0.245	1.690 ± 0.163	1.628 ± 0.400
8	1.753 ± 0.645	1.512 ± 0.922	**1.225** ± 0.246	1.689 ± 0.163	1.633 ± 0.393
9	1.754 ± 0.645	1.513 ± 0.922	**1.225** ± 0.245	1.689 ± 0.163	1.632 ± 0.394
10	1.754 ± 0.645	1.512 ± 0.921	**1.225** ± 0.245	1.689 ± 0.163	1.649 ± 0.372
15	1.754 ± 0.645	1.513 ± 0.922	**1.225** ± 0.246	1.690 ± 0.163	1.637 ± 0.366
20	1.754 ± 0.645	1.513 ± 0.922	**1.225** ± 0.247	1.690 ± 0.163	1.725 ± 0.229
25	1.751 ± 0.645	1.512 ± 0.921	**1.225** ± 0.248	1.690 ± 0.163	1.631 ± 0.379
30	1.752 ± 0.645	1.512 ± 0.920	**1.225** ± 0.246	1.690 ± 0.163	1.719 ± 0.165
40	1.751 ± 0.645	1.512 ± 0.920	**1.225** ± 0.246	1.690 ± 0.163	1.753 ± 0.189
50	1.752 ± 0.645	1.512 ± 0.920	**1.225** ± 0.246	1.690 ± 0.163	1.740 ± 0.182
60	1.752 ± 0.645	1.512 ± 0.921	**1.225** ± 0.245	1.689 ± 0.163	1.710 ± 0.131
70	1.751 ± 0.645	1.512 ± 0.921	**1.225** ± 0.246	1.690 ± 0.163	1.765 ± 0.213
80	1.751 ± 0.645	1.512 ± 0.921	**1.225** ± 0.246	1.690 ± 0.163	1.818 ± 0.153
90	1.751 ± 0.645	1.511 ± 0.920	**1.225** ± 0.246	1.690 ± 0.163	1.817 ± 0.159
100	1.752 ± 0.645	1.512 ± 0.920	**1.225** ± 0.245	1.690 ± 0.163	1.702 ± 0.139

Table B.30: Davies-Boulding index on the Berkeley protein dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.431 ± 0.006	0.446 ± 0.025	0.460 ± 0.013	0.455 ± 0.006	**0.487** ± 0.010
3	0.453 ± 0.022	0.457 ± 0.029	0.462 ± 0.018	0.490 ± 0.013	**0.574** ± 0.017
4	0.488 ± 0.012	0.508 ± 0.033	0.528 ± 0.012	0.516 ± 0.014	**0.621** ± 0.014
5	0.528 ± 0.013	0.531 ± 0.023	0.534 ± 0.018	0.559 ± 0.019	**0.653** ± 0.014
6	0.559 ± 0.012	0.546 ± 0.030	0.533 ± 0.020	0.590 ± 0.014	**0.664** ± 0.017
7	0.564 ± 0.022	0.552 ± 0.028	0.541 ± 0.007	0.601 ± 0.007	**0.692** ± 0.010
8	0.570 ± 0.007	0.561 ± 0.022	0.551 ± 0.017	0.600 ± 0.012	**0.699** ± 0.018
9	0.584 ± 0.009	0.575 ± 0.018	0.567 ± 0.011	0.614 ± 0.010	**0.700** ± 0.014
10	0.590 ± 0.012	0.577 ± 0.026	0.565 ± 0.014	0.629 ± 0.017	**0.710** ± 0.016
15	0.653 ± 0.015	0.631 ± 0.037	0.609 ± 0.012	0.683 ± 0.013	**0.738** ± 0.012
20	0.656 ± 0.017	0.637 ± 0.034	0.619 ± 0.014	0.691 ± 0.015	**0.734** ± 0.020
25	0.627 ± 0.010	0.630 ± 0.018	0.633 ± 0.014	0.654 ± 0.012	**0.745** ± 0.018
30	0.582 ± 0.006	0.608 ± 0.041	0.635 ± 0.016	0.614 ± 0.011	**0.734** ± 0.012
40	0.471 ± 0.013	0.563 ± 0.132	0.655 ± 0.012	0.500 ± 0.012	**0.702** ± 0.014
50	0.391 ± 0.009	0.520 ± 0.182	**0.648** ± 0.009	0.406 ± 0.013	0.628 ± 0.023
60	0.356 ± 0.010	0.505 ± 0.210	**0.653** ± 0.011	0.363 ± 0.012	0.525 ± 0.019
70	0.341 ± 0.011	0.489 ± 0.210	**0.637** ± 0.015	0.345 ± 0.011	0.454 ± 0.020
80	0.335 ± 0.011	0.480 ± 0.206	**0.626** ± 0.011	0.340 ± 0.010	0.393 ± 0.014
90	0.326 ± 0.016	0.477 ± 0.215	**0.628** ± 0.012	0.330 ± 0.018	0.366 ± 0.011
100	0.326 ± 0.015	0.472 ± 0.208	**0.618** ± 0.022	0.324 ± 0.017	0.346 ± 0.010

Table B.31: One-vs-one SVM classification accuracy on the Cora dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.429	0.397 ± 0.032	0.366	0.372	**0.464**
3	0.444	0.407 ± 0.037	0.370	0.384	**0.460**
4	0.385	0.387 ± 0.002	0.389	0.405	**0.448**
5	0.408	0.406 ± 0.002	0.404	0.409	**0.431**
6	0.383	0.389 ± 0.006	0.395	0.404	**0.440**
7	0.383	0.390 ± 0.007	0.397	0.414	**0.481**
8	0.383	0.389 ± 0.006	0.395	0.403	**0.467**
9	0.386	0.389 ± 0.004	0.393	0.399	**0.448**
10	0.388	0.390 ± 0.003	0.393	0.398	**0.442**
15	**0.399**	0.394 ± 0.005	0.389	**0.396**	0.389
20	**0.395**	**0.392** ± 0.003	0.389	**0.395**	0.350
25	0.395	0.391 ± 0.004	0.387	**0.434**	0.333
30	0.391	0.375 ± 0.016	0.359	**0.442**	0.339
40	0.383	0.372 ± 0.012	0.360	**0.407**	0.353
50	0.329	0.345 ± 0.016	0.360	**0.416**	0.311
60	0.333	0.348 ± 0.015	**0.363**	0.359	0.307
70	0.308	0.334 ± 0.026	0.360	**0.412**	0.336
80	0.308	0.319 ± 0.011	0.330	0.350	**0.373**
90	0.333	0.336 ± 0.004	0.340	0.314	**0.374**
100	0.322	0.352 ± 0.030	**0.381**	0.317	0.377

Table B.32: Clustering purity on the Cora dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

K	Single view			Stacked views	MV-MDS
	Worst	Average	Best		
2	0.119	0.177 ± 0.058	0.236	0.125	**0.239**
3	0.125	0.193 ± 0.068	**0.260**	0.138	0.236
4	0.137	0.169 ± 0.032	0.201	0.151	**0.223**
5	0.151	0.190 ± 0.038	**0.228**	0.158	0.195
6	0.144	0.171 ± 0.026	**0.197**	0.151	**0.196**
7	0.145	0.158 ± 0.013	0.170	0.157	**0.225**
8	0.144	0.169 ± 0.026	0.195	0.150	**0.216**
9	0.141	0.167 ± 0.026	0.192	0.147	**0.202**
10	0.140	0.167 ± 0.027	0.193	0.144	**0.209**
15	0.143	0.173 ± 0.030	**0.203**	0.148	0.157
20	0.143	0.177 ± 0.034	**0.211**	0.149	0.118
25	0.141	0.173 ± 0.032	0.205	**0.211**	0.079
30	0.099	0.146 ± 0.047	0.193	**0.209**	0.058
40	0.097	0.140 ± 0.044	**0.184**	0.165	0.125
50	0.099	0.106 ± 0.007	0.114	**0.194**	0.074
60	**0.114**	0.112 ± 0.002	0.110	0.100	0.017
70	0.093	0.101 ± 0.009	0.110	**0.191**	0.093
80	0.081	0.095 ± 0.014	**0.109**	0.088	0.098
90	0.109	0.111 ± 0.002	**0.113**	0.071	0.107
100	**0.121**	0.114 ± 0.007	0.107	0.074	0.113

Table B.33: Clustering normalized mutual information on the Cora dataset of MV-MDS compared with single view and stacked views MDS. K is the dimensionality of the projection.

Appendix C

Results of MVSC-CEV experiments

In this appendix, the detailed results of the experiments with the MV-SC-CEV method are presented. There is a results table for each combination of dataset and evaluation method, yielding a total of 36 tables. The methods are compared with the counterpart single view method, either applied to single views individually or to all views stacked on a single matrix. For single views, the worst, average and best views (on average) are given.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.515 ± 0.026	0.450 ± 0.184	**0.565** ± 0.013	0.508 ± 0.018	0.551 ± 0.023
3	0.500 ± 0.024	0.576 ± 0.179	0.685 ± 0.014	0.706 ± 0.019	**0.776** ± 0.009
4	0.511 ± 0.014	0.661 ± 0.305	0.840 ± 0.010	0.850 ± 0.013	**0.905** ± 0.007
5	0.514 ± 0.015	0.713 ± 0.329	0.877 ± 0.010	0.901 ± 0.007	**0.946** ± 0.008
6	0.534 ± 0.024	0.746 ± 0.351	0.910 ± 0.010	0.926 ± 0.006	**0.957** ± 0.005
7	0.547 ± 0.023	0.766 ± 0.358	0.926 ± 0.007	0.947 ± 0.008	**0.965** ± 0.007
8	0.547 ± 0.023	0.782 ± 0.364	0.926 ± 0.009	**0.961** ± 0.005	**0.968** ± 0.007
9	0.547 ± 0.023	0.798 ± 0.354	0.928 ± 0.009	**0.964** ± 0.004	**0.971** ± 0.005
10	0.597 ± 0.036	0.813 ± 0.333	0.932 ± 0.009	**0.966** ± 0.005	**0.974** ± 0.005
15	0.655 ± 0.020	0.841 ± 0.299	0.950 ± 0.005	**0.975** ± 0.005	**0.977** ± 0.004
20	0.663 ± 0.024	0.848 ± 0.295	0.952 ± 0.005	**0.976** ± 0.005	**0.974** ± 0.005
25	0.690 ± 0.019	0.858 ± 0.274	0.955 ± 0.006	**0.976** ± 0.005	**0.974** ± 0.004
30	0.689 ± 0.016	0.859 ± 0.269	0.955 ± 0.006	**0.980** ± 0.004	**0.974** ± 0.005
40	0.685 ± 0.021	0.849 ± 0.272	0.953 ± 0.006	**0.978** ± 0.003	**0.974** ± 0.004
50	0.682 ± 0.027	0.823 ± 0.268	0.956 ± 0.007	**0.974** ± 0.005	**0.970** ± 0.006
60	0.685 ± 0.017	0.763 ± 0.348	**0.949** ± 0.009	**0.952** ± 0.011	**0.943** ± 0.009
70	0.688 ± 0.016	0.683 ± 0.458	**0.940** ± 0.009	0.873 ± 0.027	0.889 ± 0.027
80	0.674 ± 0.018	0.590 ± 0.519	**0.901** ± 0.018	0.692 ± 0.063	0.786 ± 0.042
90	0.677 ± 0.018	0.506 ± 0.523	**0.814** ± 0.023	0.495 ± 0.071	0.654 ± 0.047
100	0.669 ± 0.014	0.429 ± 0.515	**0.705** ± 0.041	0.350 ± 0.072	0.530 ± 0.066

Table C.1: One-vs-one SVM classification accuracy on the digits dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.048 ± 0.141	0.175 ± 0.392	0.242 ± 0.195	**0.257** ± 0.153	0.166 ± 0.131
3	0.047 ± 0.145	0.209 ± 0.430	**0.310** ± 0.182	**0.310** ± 0.137	0.235 ± 0.090
4	0.047 ± 0.147	0.231 ± 0.448	0.323 ± 0.215	**0.335** ± 0.193	0.245 ± 0.084
5	0.051 ± 0.147	0.241 ± 0.451	0.345 ± 0.213	**0.359** ± 0.188	0.269 ± 0.087
6	0.106 ± 0.110	0.244 ± 0.400	0.349 ± 0.192	**0.362** ± 0.136	0.282 ± 0.086
7	0.114 ± 0.114	0.236 ± 0.376	0.335 ± 0.172	**0.345** ± 0.145	0.280 ± 0.084
8	0.114 ± 0.113	0.239 ± 0.367	0.338 ± 0.160	**0.344** ± 0.134	0.243 ± 0.096
9	0.115 ± 0.113	0.244 ± 0.366	0.344 ± 0.160	**0.359** ± 0.140	0.245 ± 0.090
10	0.115 ± 0.113	0.237 ± 0.356	0.326 ± 0.142	**0.348** ± 0.142	0.248 ± 0.090
15	0.116 ± 0.111	0.225 ± 0.327	0.320 ± 0.136	**0.344** ± 0.123	0.263 ± 0.058
20	0.112 ± 0.113	0.218 ± 0.308	0.305 ± 0.115	**0.331** ± 0.115	0.265 ± 0.065
25	0.139 ± 0.105	0.215 ± 0.275	0.296 ± 0.103	**0.306** ± 0.109	0.254 ± 0.067
30	0.147 ± 0.096	0.210 ± 0.262	0.285 ± 0.104	**0.295** ± 0.113	0.241 ± 0.055
40	0.137 ± 0.064	0.198 ± 0.237	0.274 ± 0.113	**0.281** ± 0.114	0.228 ± 0.057
50	0.141 ± 0.061	0.190 ± 0.218	0.257 ± 0.102	**0.264** ± 0.095	0.233 ± 0.050
60	0.132 ± 0.047	0.190 ± 0.213	0.256 ± 0.101	**0.263** ± 0.090	0.240 ± 0.053
70	0.127 ± 0.053	0.194 ± 0.218	0.253 ± 0.103	**0.259** ± 0.080	0.254 ± 0.066
80	0.129 ± 0.060	0.206 ± 0.235	0.259 ± 0.110	0.264 ± 0.086	**0.272** ± 0.069
90	0.124 ± 0.054	0.216 ± 0.253	0.264 ± 0.110	0.278 ± 0.090	**0.293** ± 0.070
100	0.153 ± 0.037	0.232 ± 0.266	0.272 ± 0.111	0.293 ± 0.091	**0.308** ± 0.071

Table C.2: Cophenetic correlation on the digits dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.096 ± 0.063	0.087 ± 0.114	0.098 ± 0.048	**0.103** ± 0.039	**0.104** ± 0.019
3	0.095 ± 0.063	0.120 ± 0.155	0.152 ± 0.066	**0.159** ± 0.058	0.152 ± 0.029
4	0.098 ± 0.065	0.145 ± 0.194	0.199 ± 0.093	**0.211** ± 0.092	0.186 ± 0.037
5	0.098 ± 0.065	0.163 ± 0.223	0.224 ± 0.106	**0.239** ± 0.109	0.207 ± 0.040
6	0.099 ± 0.058	0.172 ± 0.246	0.243 ± 0.121	**0.256** ± 0.109	0.219 ± 0.042
7	0.100 ± 0.059	0.178 ± 0.260	0.255 ± 0.124	**0.271** ± 0.117	0.226 ± 0.048
8	0.100 ± 0.059	0.185 ± 0.273	0.270 ± 0.130	**0.279** ± 0.120	0.225 ± 0.050
9	0.100 ± 0.059	0.191 ± 0.284	0.276 ± 0.134	**0.287** ± 0.127	0.228 ± 0.050
10	0.100 ± 0.059	0.195 ± 0.294	0.281 ± 0.134	**0.289** ± 0.130	0.230 ± 0.053
15	0.101 ± 0.061	0.205 ± 0.319	0.291 ± 0.145	**0.306** ± 0.138	0.245 ± 0.061
20	0.103 ± 0.063	0.211 ± 0.336	0.296 ± 0.150	**0.313** ± 0.146	0.257 ± 0.072
25	0.125 ± 0.088	0.219 ± 0.345	0.301 ± 0.154	**0.312** ± 0.147	0.258 ± 0.074
30	0.125 ± 0.090	0.220 ± 0.353	0.300 ± 0.156	**0.310** ± 0.149	0.260 ± 0.073
40	0.129 ± 0.102	0.222 ± 0.368	0.299 ± 0.161	**0.308** ± 0.151	0.262 ± 0.077
50	0.127 ± 0.105	0.221 ± 0.375	0.294 ± 0.160	**0.304** ± 0.146	0.265 ± 0.076
60	0.125 ± 0.105	0.223 ± 0.382	0.294 ± 0.163	**0.302** ± 0.144	0.267 ± 0.077
70	0.122 ± 0.104	0.225 ± 0.392	0.294 ± 0.164	**0.301** ± 0.142	0.271 ± 0.079
80	0.121 ± 0.105	0.228 ± 0.403	0.296 ± 0.167	**0.302** ± 0.141	0.276 ± 0.079
90	0.134 ± 0.129	0.233 ± 0.415	0.297 ± 0.169	**0.303** ± 0.139	0.282 ± 0.079
100	0.147 ± 0.128	0.238 ± 0.421	0.299 ± 0.170	**0.305** ± 0.137	0.286 ± 0.079

Table C.3: Area under the curve of the R_{NX} index on the digits dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.329	0.364 ± 0.047	0.417	0.421	**0.466**
3	0.317	0.444 ± 0.072	0.556	0.570	**0.596**
4	0.315	0.495 ± 0.097	0.621	0.715	**0.762**
5	0.289	0.555 ± 0.148	0.676	0.708	**0.875**
6	0.429	0.559 ± 0.103	0.622	0.677	**0.802**
7	0.407	0.576 ± 0.110	0.679	0.781	**0.795**
8	0.407	0.588 ± 0.121	0.734	**0.875**	0.771
9	0.407	0.582 ± 0.111	0.682	**0.885**	0.817
10	0.407	0.602 ± 0.135	0.751	0.787	**0.944**
15	0.408	0.653 ± 0.171	0.805	**0.902**	0.809
20	0.409	0.667 ± 0.178	0.895	**0.907**	0.791
25	0.516	0.675 ± 0.121	0.806	**0.898**	0.810
30	0.472	0.668 ± 0.126	0.738	**0.799**	0.791
40	0.409	0.637 ± 0.137	0.731	0.800	**0.822**
50	0.379	0.625 ± 0.138	0.765	**0.806**	**0.799**
60	0.367	0.592 ± 0.127	0.754	0.753	**0.786**
70	0.342	0.577 ± 0.121	0.695	**0.751**	0.736
80	0.332	0.592 ± 0.138	0.716	**0.751**	0.726
90	0.390	0.603 ± 0.143	0.736	**0.755**	0.719
100	0.355	0.579 ± 0.125	0.755	**0.784**	0.752

Table C.4: Clustering purity on the digits dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.439	0.360 ± 0.077	0.415	0.429	**0.468**
3	0.430	0.415 ± 0.055	0.491	0.553	**0.602**
4	0.416	0.473 ± 0.070	0.583	0.634	**0.730**
5	0.391	0.516 ± 0.092	0.622	0.650	**0.829**
6	0.402	0.514 ± 0.093	0.590	0.681	**0.821**
7	0.398	0.529 ± 0.105	0.636	0.739	**0.824**
8	0.398	0.537 ± 0.109	0.660	0.791	**0.805**
9	0.398	0.536 ± 0.101	0.640	0.807	**0.840**
10	0.398	0.543 ± 0.115	0.684	0.750	**0.893**
15	0.400	0.595 ± 0.151	0.758	**0.835**	0.815
20	0.400	0.601 ± 0.155	0.823	**0.847**	0.797
25	0.478	0.615 ± 0.115	0.762	**0.841**	0.808
30	0.439	0.615 ± 0.117	0.734	**0.793**	**0.792**
40	0.400	0.597 ± 0.123	0.729	0.803	**0.840**
50	0.383	0.584 ± 0.118	0.720	**0.811**	**0.814**
60	0.390	0.570 ± 0.112	0.719	0.753	**0.782**
70	0.348	0.557 ± 0.111	0.658	0.713	**0.728**
80	0.346	0.562 ± 0.132	0.676	0.718	**0.738**
90	0.379	0.568 ± 0.129	0.705	0.719	**0.735**
100	0.417	0.561 ± 0.092	0.693	**0.744**	**0.743**

Table C.5: Clustering normalized mutual information on the digits dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	1.977 ± 0.028	1.940 ± 0.121	1.922 ± 0.013	**1.906** ± 0.036	**1.895** ± 0.061
3	1.897 ± 0.051	1.900 ± 0.172	1.870 ± 0.076	**1.848** ± 0.072	1.868 ± 0.066
4	1.873 ± 0.078	1.877 ± 0.205	1.813 ± 0.081	1.788 ± 0.097	**1.766** ± 0.143
5	1.857 ± 0.091	1.819 ± 0.236	1.779 ± 0.080	1.777 ± 0.091	**1.746** ± 0.123
6	1.863 ± 0.088	1.834 ± 0.245	1.784 ± 0.081	1.802 ± 0.081	**1.746** ± 0.140
7	1.864 ± 0.090	1.817 ± 0.276	1.770 ± 0.078	**1.718** ± 0.117	1.741 ± 0.145
8	1.892 ± 0.100	1.819 ± 0.288	1.754 ± 0.067	**1.701** ± 0.130	1.739 ± 0.134
9	1.869 ± 0.106	1.808 ± 0.301	1.750 ± 0.085	**1.709** ± 0.128	1.728 ± 0.150
10	1.854 ± 0.112	1.805 ± 0.298	1.736 ± 0.094	1.737 ± 0.102	**1.696** ± 0.141
15	1.843 ± 0.116	1.790 ± 0.314	1.719 ± 0.103	**1.699** ± 0.124	**1.705** ± 0.181
20	1.857 ± 0.119	1.785 ± 0.316	**1.705** ± 0.072	**1.700** ± 0.123	**1.695** ± 0.180
25	1.856 ± 0.125	1.768 ± 0.384	1.719 ± 0.069	**1.699** ± 0.122	**1.716** ± 0.173
30	1.848 ± 0.117	1.762 ± 0.380	1.729 ± 0.067	1.724 ± 0.107	**1.705** ± 0.184
40	1.857 ± 0.126	1.759 ± 0.398	**1.712** ± 0.067	**1.724** ± 0.111	**1.717** ± 0.145
50	1.836 ± 0.128	1.757 ± 0.390	**1.696** ± 0.071	1.723 ± 0.114	1.718 ± 0.133
60	1.814 ± 0.137	1.752 ± 0.386	**1.687** ± 0.080	1.726 ± 0.108	1.723 ± 0.134
70	1.838 ± 0.132	1.763 ± 0.389	**1.686** ± 0.078	**1.679** ± 0.146	1.733 ± 0.144
80	1.843 ± 0.113	1.753 ± 0.390	1.708 ± 0.075	**1.682** ± 0.145	1.712 ± 0.160
90	1.881 ± 0.134	1.769 ± 0.379	1.714 ± 0.054	**1.686** ± 0.138	1.705 ± 0.180
100	1.841 ± 0.131	1.784 ± 0.314	**1.704** ± 0.080	1.724 ± 0.121	1.732 ± 0.150

Table C.6: Davies-Boulding index on the digits dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.505 ± 0.018	0.476 ± 0.053	0.497 ± 0.007	0.482 ± 0.016	**0.523** ± 0.015
3	0.536 ± 0.021	0.522 ± 0.058	0.517 ± 0.008	0.574 ± 0.021	**0.601** ± 0.019
4	0.525 ± 0.018	0.528 ± 0.055	0.535 ± 0.005	0.585 ± 0.016	**0.592** ± 0.014
5	0.546 ± 0.023	0.568 ± 0.058	0.552 ± 0.011	0.604 ± 0.028	**0.650** ± 0.021
6	0.560 ± 0.006	0.589 ± 0.054	0.624 ± 0.009	0.611 ± 0.029	**0.650** ± 0.014
7	0.583 ± 0.020	0.609 ± 0.079	**0.666** ± 0.006	0.612 ± 0.029	**0.664** ± 0.016
8	0.594 ± 0.027	0.618 ± 0.087	**0.684** ± 0.004	0.624 ± 0.027	0.674 ± 0.022
9	0.622 ± 0.026	0.638 ± 0.081	**0.699** ± 0.007	0.669 ± 0.026	**0.703** ± 0.021
10	0.626 ± 0.023	0.644 ± 0.091	**0.716** ± 0.009	0.673 ± 0.027	**0.717** ± 0.024
15	0.665 ± 0.028	0.682 ± 0.097	0.760 ± 0.006	0.717 ± 0.017	**0.783** ± 0.010
20	0.672 ± 0.021	0.697 ± 0.113	**0.793** ± 0.006	0.733 ± 0.012	0.784 ± 0.016
25	0.685 ± 0.017	0.712 ± 0.108	**0.799** ± 0.006	0.741 ± 0.013	**0.796** ± 0.012
30	0.683 ± 0.022	0.717 ± 0.118	**0.815** ± 0.003	0.742 ± 0.007	0.799 ± 0.013
40	0.681 ± 0.021	0.726 ± 0.120	**0.821** ± 0.004	0.757 ± 0.013	0.801 ± 0.015
50	0.672 ± 0.007	0.720 ± 0.119	**0.815** ± 0.008	0.745 ± 0.015	0.792 ± 0.011
60	0.640 ± 0.016	0.704 ± 0.134	**0.811** ± 0.007	0.719 ± 0.025	0.766 ± 0.017
70	0.567 ± 0.026	0.657 ± 0.161	**0.783** ± 0.008	0.656 ± 0.017	0.698 ± 0.016
80	0.497 ± 0.032	0.595 ± 0.175	**0.725** ± 0.012	0.562 ± 0.036	0.591 ± 0.032
90	0.436 ± 0.034	0.524 ± 0.160	**0.629** ± 0.031	0.465 ± 0.029	0.504 ± 0.036
100	0.391 ± 0.027	0.455 ± 0.126	**0.518** ± 0.035	0.412 ± 0.032	0.453 ± 0.029

Table C.7: One-vs-one SVM classification accuracy on the Reuters multilingual corpus dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	**0.720** ± 0.064	0.699 ± 0.165	**0.723** ± 0.066	**0.723** ± 0.057	**0.720** ± 0.058
3	**0.719** ± 0.064	0.659 ± 0.295	**0.726** ± 0.064	**0.725** ± 0.058	**0.722** ± 0.059
4	0.574 ± 0.074	0.613 ± 0.299	**0.732** ± 0.065	0.514 ± 0.151	0.698 ± 0.058
5	0.545 ± 0.089	0.612 ± 0.245	**0.623** ± 0.044	0.567 ± 0.120	0.528 ± 0.027
6	0.410 ± 0.085	0.554 ± 0.254	**0.561** ± 0.073	0.501 ± 0.088	0.472 ± 0.062
7	0.478 ± 0.081	**0.567** ± 0.228	0.512 ± 0.071	**0.571** ± 0.076	0.541 ± 0.052
8	0.466 ± 0.085	0.563 ± 0.201	**0.591** ± 0.057	0.480 ± 0.062	0.483 ± 0.055
9	0.443 ± 0.084	0.541 ± 0.197	**0.604** ± 0.054	0.493 ± 0.057	0.494 ± 0.050
10	0.378 ± 0.086	0.516 ± 0.226	**0.564** ± 0.049	0.537 ± 0.049	0.528 ± 0.045
15	0.467 ± 0.073	0.508 ± 0.151	**0.551** ± 0.051	0.510 ± 0.058	0.512 ± 0.046
20	0.435 ± 0.064	0.516 ± 0.152	**0.571** ± 0.034	0.414 ± 0.052	0.378 ± 0.046
25	0.411 ± 0.057	0.492 ± 0.140	**0.529** ± 0.036	0.376 ± 0.045	0.372 ± 0.038
30	0.433 ± 0.055	0.478 ± 0.140	**0.562** ± 0.029	0.421 ± 0.035	0.369 ± 0.036
40	0.379 ± 0.044	0.441 ± 0.121	**0.501** ± 0.029	0.379 ± 0.033	0.400 ± 0.027
50	0.336 ± 0.046	0.390 ± 0.188	**0.523** ± 0.025	0.374 ± 0.032	0.388 ± 0.023
60	0.325 ± 0.041	0.366 ± 0.097	**0.402** ± 0.021	0.299 ± 0.030	0.359 ± 0.024
70	0.288 ± 0.039	0.338 ± 0.104	**0.387** ± 0.020	0.293 ± 0.023	0.304 ± 0.021
80	0.258 ± 0.039	0.301 ± 0.093	**0.321** ± 0.019	0.273 ± 0.023	0.256 ± 0.022
90	0.237 ± 0.039	0.268 ± 0.103	**0.338** ± 0.017	0.258 ± 0.025	0.231 ± 0.024
100	0.184 ± 0.044	0.200 ± 0.083	0.216 ± 0.025	0.228 ± 0.029	**0.233** ± 0.029

Table C.8: Cophenetic correlation on the Reuters multilingual corpus dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.101 ± 0.010	0.095 ± 0.026	0.101 ± 0.009	**0.106** ± 0.005	0.103 ± 0.005
3	**0.117** ± 0.015	0.112 ± 0.074	0.110 ± 0.011	0.115 ± 0.005	0.111 ± 0.005
4	0.134 ± 0.039	0.139 ± 0.092	0.134 ± 0.016	**0.172** ± 0.051	0.152 ± 0.010
5	0.169 ± 0.056	0.170 ± 0.120	0.184 ± 0.042	**0.216** ± 0.052	0.196 ± 0.023
6	0.168 ± 0.060	0.188 ± 0.132	0.217 ± 0.055	**0.232** ± 0.042	0.228 ± 0.031
7	0.175 ± 0.063	0.197 ± 0.136	0.228 ± 0.058	**0.236** ± 0.042	**0.234** ± 0.031
8	0.180 ± 0.070	0.202 ± 0.140	0.231 ± 0.059	**0.242** ± 0.036	**0.243** ± 0.032
9	0.187 ± 0.076	0.209 ± 0.147	0.240 ± 0.063	**0.251** ± 0.035	0.248 ± 0.031
10	0.187 ± 0.081	0.213 ± 0.153	0.241 ± 0.063	**0.255** ± 0.035	**0.256** ± 0.032
15	0.213 ± 0.090	0.227 ± 0.161	0.248 ± 0.067	**0.274** ± 0.038	**0.277** ± 0.030
20	0.211 ± 0.091	0.234 ± 0.169	0.253 ± 0.070	0.276 ± 0.030	**0.280** ± 0.028
25	0.212 ± 0.098	0.239 ± 0.183	0.263 ± 0.077	**0.278** ± 0.031	**0.281** ± 0.028
30	0.216 ± 0.102	0.240 ± 0.190	0.267 ± 0.079	**0.283** ± 0.030	**0.282** ± 0.027
40	0.212 ± 0.106	0.243 ± 0.205	0.273 ± 0.084	0.281 ± 0.029	**0.286** ± 0.026
50	0.211 ± 0.111	0.239 ± 0.216	0.272 ± 0.085	**0.284** ± 0.028	**0.285** ± 0.024
60	0.212 ± 0.112	0.239 ± 0.222	0.265 ± 0.088	0.280 ± 0.026	**0.287** ± 0.025
70	0.214 ± 0.115	0.240 ± 0.229	0.263 ± 0.090	0.283 ± 0.025	**0.287** ± 0.023
80	0.218 ± 0.119	0.240 ± 0.234	0.261 ± 0.092	0.286 ± 0.027	**0.290** ± 0.024
90	0.222 ± 0.121	0.239 ± 0.236	0.261 ± 0.094	0.288 ± 0.027	**0.291** ± 0.024
100	0.226 ± 0.122	0.240 ± 0.241	0.257 ± 0.096	0.288 ± 0.026	**0.293** ± 0.024

Table C.9: Area under the curve of the R_{NX} index on the Reuters multilingual corpus dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.463	0.458 ± 0.017	0.427	0.469	**0.509**
3	0.458	0.457 ± 0.025	0.426	0.456	**0.494**
4	0.475	0.476 ± 0.035	0.520	0.529	**0.562**
5	0.522	0.491 ± 0.025	0.507	0.530	**0.566**
6	0.495	0.475 ± 0.020	0.439	0.494	**0.567**
7	0.504	0.487 ± 0.025	0.442	0.497	**0.567**
8	0.505	0.509 ± 0.011	0.529	0.487	**0.573**
9	0.508	0.508 ± 0.007	0.518	0.486	**0.579**
10	0.478	0.489 ± 0.021	0.463	0.488	**0.581**
15	0.491	0.517 ± 0.032	**0.570**	0.537	0.551
20	0.475	0.537 ± 0.045	**0.613**	0.537	0.599
25	0.478	0.534 ± 0.061	**0.648**	0.566	0.607
30	0.383	0.512 ± 0.096	**0.672**	0.525	0.579
40	0.386	0.504 ± 0.086	**0.640**	0.529	0.555
50	0.491	0.517 ± 0.061	**0.634**	0.493	0.541
60	0.513	0.498 ± 0.020	0.526	0.511	**0.583**
70	0.510	0.505 ± 0.040	**0.579**	0.505	0.544
80	0.492	0.495 ± 0.019	0.531	0.482	**0.568**
90	0.490	0.503 ± 0.036	0.574	0.482	**0.606**
100	0.512	0.503 ± 0.014	0.525	0.497	**0.593**

Table C.10: Clustering purity on the Reuters multilingual corpus dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.163	0.156 ± 0.012	0.136	0.168	**0.270**
3	0.183	0.159 ± 0.016	0.135	0.170	**0.270**
4	0.169	0.175 ± 0.033	0.225	0.204	**0.321**
5	0.202	0.194 ± 0.024	0.224	0.252	**0.320**
6	0.163	0.189 ± 0.023	0.162	0.196	**0.313**
7	0.163	0.182 ± 0.019	0.164	0.198	**0.311**
8	0.180	0.203 ± 0.030	0.263	0.233	**0.339**
9	0.187	0.214 ± 0.029	0.268	0.233	**0.341**
10	0.190	0.207 ± 0.025	0.230	0.207	**0.338**
15	0.217	0.238 ± 0.048	0.329	0.224	**0.338**
20	0.241	0.236 ± 0.049	0.321	0.256	**0.350**
25	0.190	0.237 ± 0.074	**0.377**	0.260	0.361
30	0.176	0.221 ± 0.090	**0.381**	0.259	0.339
40	0.183	0.215 ± 0.076	**0.350**	0.219	0.330
50	0.186	0.229 ± 0.058	**0.340**	0.198	0.315
60	0.189	0.220 ± 0.042	0.300	0.207	**0.319**
70	0.190	0.212 ± 0.036	0.276	0.208	**0.307**
80	0.174	0.213 ± 0.039	0.288	0.192	**0.311**
90	0.185	0.210 ± 0.033	0.274	0.180	**0.350**
100	0.156	0.204 ± 0.039	0.274	0.192	**0.330**

Table C.11: Clustering normalized mutual information on the Reuters multilingual corpus dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	1.902 ± 0.013	1.834 ± 0.082	**1.817** ± 0.016	**1.811** ± 0.012	**1.810** ± 0.013
3	**1.623** ± 0.018	1.675 ± 0.154	1.798 ± 0.020	**1.622** ± 0.019	**1.619** ± 0.017
4	**1.532** ± 0.048	1.633 ± 0.205	1.623 ± 0.028	1.600 ± 0.038	**1.530** ± 0.018
5	1.660 ± 0.076	1.608 ± 0.209	**1.507** ± 0.022	1.708 ± 0.054	1.753 ± 0.014
6	**1.661** ± 0.074	1.686 ± 0.142	1.701 ± 0.026	**1.654** ± 0.063	**1.656** ± 0.059
7	**1.663** ± 0.073	1.700 ± 0.133	1.701 ± 0.026	**1.652** ± 0.063	**1.655** ± 0.058
8	1.663 ± 0.079	1.702 ± 0.142	1.687 ± 0.014	**1.646** ± 0.065	1.688 ± 0.056
9	**1.628** ± 0.076	1.670 ± 0.159	1.659 ± 0.016	1.647 ± 0.064	1.685 ± 0.051
10	1.622 ± 0.075	1.654 ± 0.149	1.660 ± 0.010	**1.592** ± 0.059	1.652 ± 0.061
15	**1.680** ± 0.066	1.691 ± 0.173	**1.674** ± 0.017	1.760 ± 0.067	1.759 ± 0.080
20	1.727 ± 0.062	1.702 ± 0.122	**1.671** ± 0.015	1.708 ± 0.050	1.780 ± 0.066
25	**1.726** ± 0.071	1.752 ± 0.183	1.848 ± 0.016	1.819 ± 0.060	1.807 ± 0.069
30	1.880 ± 0.059	1.759 ± 0.219	**1.693** ± 0.016	1.830 ± 0.074	1.846 ± 0.070
40	1.880 ± 0.057	1.766 ± 0.190	**1.693** ± 0.009	1.830 ± 0.062	1.833 ± 0.079
50	1.878 ± 0.056	1.746 ± 0.189	**1.692** ± 0.011	1.837 ± 0.056	1.717 ± 0.047
60	1.721 ± 0.054	1.706 ± 0.106	1.695 ± 0.018	**1.670** ± 0.047	1.725 ± 0.038
70	1.746 ± 0.054	1.708 ± 0.118	1.697 ± 0.016	**1.668** ± 0.046	**1.670** ± 0.044
80	1.716 ± 0.055	**1.692** ± 0.111	**1.694** ± 0.018	**1.702** ± 0.054	1.725 ± 0.043
90	1.745 ± 0.057	1.740 ± 0.166	**1.716** ± 0.018	1.851 ± 0.058	**1.716** ± 0.042
100	**1.884** ± 0.038	1.889 ± 0.092	1.910 ± 0.008	**1.870** ± 0.046	**1.875** ± 0.053

Table C.12: Davies-Boulding index on the Reuters multilingual corpus dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.584 ± 0.018	0.588 ± 0.024	0.591 ± 0.015	**0.619** ± 0.012	**0.620** ± 0.015
3	0.816 ± 0.012	0.818 ± 0.016	0.819 ± 0.010	**0.860** ± 0.013	**0.862** ± 0.014
4	0.920 ± 0.009	0.913 ± 0.017	0.906 ± 0.011	**0.943** ± 0.009	**0.943** ± 0.007
5	0.939 ± 0.012	0.937 ± 0.014	0.936 ± 0.006	**0.958** ± 0.006	**0.958** ± 0.005
6	0.938 ± 0.012	0.940 ± 0.014	0.941 ± 0.007	**0.958** ± 0.006	**0.958** ± 0.006
7	0.948 ± 0.010	0.944 ± 0.014	0.940 ± 0.009	**0.959** ± 0.007	**0.960** ± 0.008
8	0.950 ± 0.010	0.949 ± 0.013	0.949 ± 0.009	**0.967** ± 0.007	**0.967** ± 0.007
9	0.950 ± 0.008	0.950 ± 0.012	0.951 ± 0.009	**0.966** ± 0.006	**0.966** ± 0.006
10	0.951 ± 0.008	0.951 ± 0.011	0.951 ± 0.008	**0.964** ± 0.008	**0.964** ± 0.008
15	0.956 ± 0.008	0.952 ± 0.012	0.949 ± 0.008	**0.968** ± 0.007	**0.969** ± 0.006
20	0.953 ± 0.009	0.950 ± 0.013	0.947 ± 0.009	**0.967** ± 0.005	**0.967** ± 0.005
25	0.949 ± 0.006	0.947 ± 0.010	0.946 ± 0.008	**0.963** ± 0.004	**0.963** ± 0.004
30	0.944 ± 0.007	0.945 ± 0.011	0.946 ± 0.009	**0.959** ± 0.007	**0.959** ± 0.006
40	0.934 ± 0.007	0.936 ± 0.013	0.937 ± 0.010	**0.952** ± 0.005	**0.952** ± 0.005
50	0.907 ± 0.010	0.913 ± 0.018	0.918 ± 0.012	**0.939** ± 0.010	**0.939** ± 0.010
60	0.809 ± 0.021	0.827 ± 0.040	0.846 ± 0.022	**0.889** ± 0.018	**0.889** ± 0.018
70	0.638 ± 0.034	0.653 ± 0.050	0.669 ± 0.030	**0.767** ± 0.025	**0.768** ± 0.026
80	0.523 ± 0.030	0.528 ± 0.035	0.533 ± 0.018	**0.641** ± 0.028	**0.644** ± 0.029
90	0.448 ± 0.032	0.456 ± 0.037	0.464 ± 0.015	**0.553** ± 0.027	**0.555** ± 0.026
100	0.401 ± 0.023	0.414 ± 0.034	0.428 ± 0.016	**0.498** ± 0.021	**0.498** ± 0.022

Table C.13: One-vs-one SVM classification accuracy on the BBC segmented news dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	**0.139** ± 0.007	0.127 ± 0.018	0.115 ± 0.003	0.133 ± 0.001	0.133 ± 0.001
3	0.195 ± 0.008	0.193 ± 0.009	0.191 ± 0.003	**0.201** ± 0.002	**0.201** ± 0.002
4	0.207 ± 0.006	0.208 ± 0.007	0.209 ± 0.003	**0.230** ± 0.001	**0.229** ± 0.000
5	0.221 ± 0.008	0.225 ± 0.011	0.229 ± 0.004	**0.237** ± 0.001	**0.237** ± 0.001
6	0.221 ± 0.010	0.224 ± 0.012	0.227 ± 0.005	**0.235** ± 0.002	**0.235** ± 0.002
7	0.217 ± 0.010	0.219 ± 0.012	0.220 ± 0.006	**0.226** ± 0.002	**0.226** ± 0.002
8	0.195 ± 0.009	0.202 ± 0.016	0.210 ± 0.007	**0.214** ± 0.001	**0.213** ± 0.001
9	0.192 ± 0.010	0.197 ± 0.014	0.202 ± 0.008	**0.207** ± 0.001	**0.207** ± 0.000
10	0.182 ± 0.010	0.188 ± 0.016	0.195 ± 0.009	**0.199** ± 0.000	**0.199** ± 0.000
15	0.182 ± 0.016	0.182 ± 0.022	0.182 ± 0.015	**0.211** ± 0.000	**0.211** ± 0.000
20	0.180 ± 0.021	0.181 ± 0.028	0.182 ± 0.018	**0.198** ± 0.001	**0.198** ± 0.000
25	0.189 ± 0.028	0.192 ± 0.036	0.194 ± 0.023	**0.211** ± 0.002	**0.211** ± 0.002
30	0.197 ± 0.031	0.199 ± 0.040	0.201 ± 0.025	**0.221** ± 0.002	**0.221** ± 0.002
40	0.221 ± 0.040	0.218 ± 0.052	0.215 ± 0.032	**0.238** ± 0.002	**0.238** ± 0.002
50	0.244 ± 0.049	0.241 ± 0.065	0.239 ± 0.043	**0.257** ± 0.002	**0.257** ± 0.002
60	0.270 ± 0.059	0.270 ± 0.081	0.270 ± 0.056	**0.289** ± 0.003	**0.289** ± 0.003
70	0.314 ± 0.074	0.319 ± 0.105	0.325 ± 0.073	**0.347** ± 0.003	**0.347** ± 0.003
80	0.385 ± 0.098	0.388 ± 0.136	0.390 ± 0.095	**0.429** ± 0.002	**0.429** ± 0.002
90	0.467 ± 0.124	0.465 ± 0.172	0.462 ± 0.120	**0.513** ± 0.001	**0.513** ± 0.000
100	0.535 ± 0.144	0.534 ± 0.205	0.534 ± 0.146	**0.599** ± 0.000	**0.599** ± 0.001

Table C.14: Cophenetic correlation on the BBC segmented news dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.081 ± 0.006	0.080 ± 0.008	0.078 ± 0.005	**0.089** ± 0.000	**0.089** ± 0.000
3	0.121 ± 0.010	0.120 ± 0.013	0.119 ± 0.008	**0.140** ± 0.000	**0.140** ± 0.000
4	0.143 ± 0.012	0.142 ± 0.017	0.141 ± 0.012	**0.170** ± 0.000	**0.170** ± 0.000
5	0.155 ± 0.014	0.155 ± 0.019	0.155 ± 0.014	**0.177** ± 0.001	**0.177** ± 0.000
6	0.173 ± 0.018	0.173 ± 0.025	0.172 ± 0.017	**0.196** ± 0.000	**0.196** ± 0.000
7	0.180 ± 0.021	0.182 ± 0.028	0.183 ± 0.019	**0.206** ± 0.001	**0.206** ± 0.001
8	0.183 ± 0.022	0.187 ± 0.032	0.192 ± 0.022	**0.213** ± 0.001	**0.213** ± 0.001
9	0.192 ± 0.025	0.194 ± 0.035	0.197 ± 0.024	**0.217** ± 0.001	**0.217** ± 0.001
10	0.193 ± 0.026	0.196 ± 0.037	0.198 ± 0.026	**0.220** ± 0.002	**0.220** ± 0.001
15	0.212 ± 0.034	0.213 ± 0.048	0.214 ± 0.034	**0.243** ± 0.001	**0.243** ± 0.001
20	0.223 ± 0.042	0.222 ± 0.058	0.222 ± 0.039	**0.246** ± 0.002	**0.247** ± 0.002
25	0.233 ± 0.050	0.232 ± 0.067	0.231 ± 0.045	**0.257** ± 0.002	**0.257** ± 0.002
30	0.236 ± 0.053	0.237 ± 0.073	0.239 ± 0.051	**0.264** ± 0.003	**0.265** ± 0.003
40	0.249 ± 0.063	0.248 ± 0.086	0.248 ± 0.058	**0.271** ± 0.003	**0.271** ± 0.003
50	0.258 ± 0.071	0.258 ± 0.098	0.258 ± 0.068	**0.279** ± 0.001	**0.279** ± 0.001
60	0.269 ± 0.079	0.271 ± 0.112	0.272 ± 0.080	**0.290** ± 0.002	**0.289** ± 0.001
70	0.286 ± 0.091	0.289 ± 0.130	0.292 ± 0.092	**0.306** ± 0.002	**0.306** ± 0.001
80	0.308 ± 0.105	0.309 ± 0.149	0.310 ± 0.105	**0.329** ± 0.000	**0.329** ± 0.000
90	0.332 ± 0.121	0.331 ± 0.170	0.330 ± 0.120	**0.352** ± 0.001	**0.352** ± 0.000
100	0.353 ± 0.136	0.353 ± 0.193	0.353 ± 0.137	**0.380** ± 0.001	**0.380** ± 0.000

Table C.15: Area under the curve of the R_{NX} index on the BBC segmented news dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.535	0.542 ± 0.007	0.550	0.547	**0.567**
3	0.552	0.575 ± 0.023	0.597	0.594	**0.616**
4	0.630	0.663 ± 0.033	0.696	0.837	**0.860**
5	0.861	0.862 ± 0.001	0.863	0.885	**0.911**
6	0.635	0.717 ± 0.082	0.800	0.919	**0.940**
7	0.684	0.746 ± 0.061	**0.807**	0.713	0.733
8	0.591	0.628 ± 0.037	**0.665**	0.613	0.637
9	0.551	0.568 ± 0.017	0.585	0.598	**0.624**
10	0.527	0.551 ± 0.025	0.576	0.575	**0.591**
15	0.443	0.518 ± 0.075	**0.593**	0.456	0.478
20	0.429	0.446 ± 0.017	**0.463**	0.435	0.457
25	**0.505**	0.497 ± 0.009	0.488	0.428	0.452
30	0.388	0.494 ± 0.106	**0.600**	0.421	0.443
40	**0.532**	**0.530** ± 0.002	**0.528**	0.414	0.469
50	0.484	0.439 ± 0.045	0.393	0.503	**0.525**
60	**0.498**	0.460 ± 0.038	0.423	0.394	0.415
70	0.380	0.443 ± 0.063	**0.505**	0.397	0.409
80	0.341	0.374 ± 0.033	0.407	0.413	**0.510**
90	0.383	0.383 ± 0.000	0.383	0.435	**0.509**
100	**0.582**	0.524 ± 0.058	0.466	0.478	0.523

Table C.16: Clustering purity on the BBC segmented news dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.367	0.377 ± 0.010	0.387	0.399	**0.433**
3	0.498	0.497 ± 0.001	0.496	0.510	**0.547**
4	0.548	0.563 ± 0.015	0.578	0.710	**0.747**
5	0.694	0.700 ± 0.006	0.706	0.757	**0.798**
6	0.493	0.560 ± 0.067	0.627	0.790	**0.826**
7	0.584	0.602 ± 0.019	0.621	0.630	**0.666**
8	0.493	0.530 ± 0.037	**0.567**	0.515	0.556
9	0.458	0.464 ± 0.005	0.469	0.522	**0.562**
10	0.353	0.408 ± 0.054	0.462	0.490	**0.522**
15	0.266	0.390 ± 0.124	**0.514**	0.315	0.352
20	0.247	0.295 ± 0.049	**0.344**	0.297	0.334
25	0.332	0.361 ± 0.028	**0.389**	0.288	0.328
30	0.151	0.349 ± 0.198	**0.547**	0.306	0.343
40	**0.479**	0.445 ± 0.034	0.412	0.299	0.385
50	0.298	0.237 ± 0.062	0.175	0.357	**0.394**
60	**0.403**	0.318 ± 0.085	0.233	0.237	0.274
70	0.222	0.294 ± 0.072	**0.366**	0.240	0.257
80	0.132	0.165 ± 0.033	0.198	0.215	**0.416**
90	0.172	0.183 ± 0.012	0.195	0.267	**0.327**
100	**0.364**	0.313 ± 0.050	0.263	0.331	0.342

Table C.17: Clustering normalized mutual information on the BBC segmented news dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	**1.984** ± 0.001	**1.983** ± 0.003	**1.983** ± 0.002	**1.983** ± 0.000	**1.983** ± 0.000
3	**1.982** ± 0.002	**1.982** ± 0.003	**1.982** ± 0.002	**1.982** ± 0.000	**1.982** ± 0.001
4	**1.978** ± 0.002	**1.978** ± 0.003	**1.978** ± 0.002	**1.972** ± 0.000	**1.972** ± 0.000
5	**1.972** ± 0.001	**1.972** ± 0.002	**1.972** ± 0.001	**1.971** ± 0.000	**1.971** ± 0.000
6	**1.973** ± 0.002	**1.975** ± 0.003	**1.976** ± 0.001	**1.970** ± 0.000	**1.970** ± 0.000
7	**1.973** ± 0.002	**1.971** ± 0.003	**1.970** ± 0.001	**1.970** ± 0.000	**1.970** ± 0.000
8	**1.972** ± 0.005	**1.968** ± 0.007	**1.964** ± 0.001	**1.967** ± 0.000	**1.967** ± 0.000
9	**1.968** ± 0.004	**1.964** ± 0.007	**1.960** ± 0.002	**1.966** ± 0.001	**1.967** ± 0.001
10	**1.961** ± 0.005	**1.963** ± 0.006	**1.965** ± 0.003	**1.962** ± 0.001	**1.961** ± 0.001
15	**1.964** ± 0.005	**1.964** ± 0.006	**1.965** ± 0.003	**1.952** ± 0.003	**1.952** ± 0.003
20	**1.955** ± 0.005	**1.959** ± 0.008	**1.964** ± 0.003	**1.948** ± 0.002	**1.948** ± 0.002
25	**1.959** ± 0.006	**1.960** ± 0.006	**1.961** ± 0.003	**1.946** ± 0.002	**1.946** ± 0.002
30	**1.962** ± 0.006	**1.963** ± 0.007	**1.965** ± 0.003	**1.959** ± 0.001	**1.959** ± 0.001
40	**1.960** ± 0.005	**1.958** ± 0.007	**1.956** ± 0.004	**1.956** ± 0.001	**1.955** ± 0.001
50	**1.965** ± 0.005	**1.965** ± 0.007	**1.965** ± 0.004	**1.956** ± 0.001	**1.956** ± 0.001
60	**1.967** ± 0.005	**1.964** ± 0.007	**1.961** ± 0.004	**1.959** ± 0.001	**1.959** ± 0.001
70	**1.959** ± 0.005	**1.963** ± 0.009	**1.967** ± 0.004	**1.961** ± 0.001	**1.961** ± 0.001
80	**1.970** ± 0.005	**1.970** ± 0.007	**1.970** ± 0.005	**1.961** ± 0.001	**1.955** ± 0.001
90	**1.974** ± 0.005	**1.972** ± 0.006	**1.970** ± 0.004	**1.971** ± 0.000	**1.971** ± 0.000
100	**1.983** ± 0.002	**1.984** ± 0.003	**1.985** ± 0.002	**1.984** ± 0.000	**1.986** ± 0.000

Table C.18: Davies-Boulding index on the BBC segmented news dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.044 ± 0.003	0.054 ± 0.024	0.053 ± 0.006	**0.071** ± 0.008	0.069 ± 0.008
3	0.044 ± 0.003	0.057 ± 0.029	0.053 ± 0.006	**0.076** ± 0.007	**0.076** ± 0.008
4	0.044 ± 0.004	0.058 ± 0.033	0.052 ± 0.006	**0.078** ± 0.008	**0.077** ± 0.006
5	0.044 ± 0.004	0.051 ± 0.029	0.052 ± 0.006	0.047 ± 0.013	**0.079** ± 0.004
6	0.044 ± 0.003	0.048 ± 0.019	0.053 ± 0.006	0.042 ± 0.005	**0.067** ± 0.020
7	0.044 ± 0.003	0.048 ± 0.021	**0.056** ± 0.006	0.041 ± 0.005	0.052 ± 0.024
8	0.044 ± 0.003	0.048 ± 0.021	**0.059** ± 0.005	0.039 ± 0.004	0.037 ± 0.003
9	0.044 ± 0.003	0.048 ± 0.022	**0.060** ± 0.005	0.039 ± 0.004	0.043 ± 0.014
10	0.043 ± 0.003	0.048 ± 0.025	**0.063** ± 0.008	0.041 ± 0.010	0.043 ± 0.015
15	0.044 ± 0.006	0.051 ± 0.031	**0.073** ± 0.008	0.041 ± 0.011	0.039 ± 0.007
20	0.047 ± 0.005	0.046 ± 0.019	**0.048** ± 0.005	0.043 ± 0.011	0.042 ± 0.007
25	0.053 ± 0.008	0.046 ± 0.024	0.038 ± 0.004	**0.060** ± 0.037	0.044 ± 0.007
30	**0.055** ± 0.007	0.050 ± 0.033	0.046 ± 0.013	0.047 ± 0.027	0.045 ± 0.006
40	0.053 ± 0.007	0.054 ± 0.052	0.068 ± 0.039	**0.087** ± 0.019	0.040 ± 0.005
50	0.047 ± 0.005	0.056 ± 0.079	**0.114** ± 0.042	0.061 ± 0.008	0.072 ± 0.025
60	0.043 ± 0.004	0.055 ± 0.070	**0.117** ± 0.008	0.050 ± 0.006	0.068 ± 0.009
70	0.039 ± 0.004	0.050 ± 0.044	**0.085** ± 0.011	0.042 ± 0.004	0.061 ± 0.006
80	0.038 ± 0.004	0.046 ± 0.030	**0.064** ± 0.007	0.038 ± 0.005	0.050 ± 0.004
90	0.037 ± 0.004	0.046 ± 0.031	**0.052** ± 0.005	0.038 ± 0.004	0.044 ± 0.004
100	0.036 ± 0.004	**0.046** ± 0.037	0.042 ± 0.005	0.038 ± 0.005	0.040 ± 0.004

Table C.19: One-vs-one SVM classification accuracy on the animal with attributes (AWA) dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	-0.025 ± 0.085	0.142 ± 0.417	**0.272** ± 0.127	0.013 ± 0.055	0.146 ± 0.064
3	0.014 ± 0.093	0.132 ± 0.408	**0.272** ± 0.129	0.089 ± 0.078	0.146 ± 0.068
4	-0.004 ± 0.095	0.129 ± 0.426	**0.274** ± 0.132	0.028 ± 0.083	0.103 ± 0.054
5	-0.007 ± 0.098	0.122 ± 0.444	**0.275** ± 0.134	0.030 ± 0.091	0.111 ± 0.058
6	-0.038 ± 0.111	0.118 ± 0.465	**0.275** ± 0.135	-0.001 ± 0.098	0.123 ± 0.055
7	-0.031 ± 0.113	0.121 ± 0.466	**0.276** ± 0.136	-0.020 ± 0.096	0.125 ± 0.050
8	-0.031 ± 0.109	0.123 ± 0.469	**0.277** ± 0.137	-0.014 ± 0.093	0.126 ± 0.039
9	-0.029 ± 0.105	0.124 ± 0.473	**0.277** ± 0.137	-0.018 ± 0.090	0.125 ± 0.030
10	-0.015 ± 0.101	0.126 ± 0.476	**0.280** ± 0.141	-0.020 ± 0.087	0.129 ± 0.030
15	-0.024 ± 0.101	0.130 ± 0.502	**0.286** ± 0.148	-0.045 ± 0.087	0.094 ± 0.027
20	-0.028 ± 0.100	0.107 ± 0.491	**0.291** ± 0.152	-0.040 ± 0.098	0.089 ± 0.023
25	-0.032 ± 0.100	0.104 ± 0.503	**0.293** ± 0.156	-0.032 ± 0.093	0.081 ± 0.021
30	-0.037 ± 0.101	0.109 ± 0.502	**0.295** ± 0.160	-0.025 ± 0.091	0.074 ± 0.017
40	-0.059 ± 0.095	0.086 ± 0.462	**0.300** ± 0.165	-0.028 ± 0.076	0.050 ± 0.015
50	-0.043 ± 0.079	0.069 ± 0.393	**0.307** ± 0.173	-0.034 ± 0.069	0.062 ± 0.020
60	-0.043 ± 0.065	0.077 ± 0.391	**0.310** ± 0.178	-0.037 ± 0.059	0.079 ± 0.029
70	-0.030 ± 0.043	0.086 ± 0.372	**0.314** ± 0.181	-0.026 ± 0.044	0.086 ± 0.032
80	0.004 ± 0.023	0.099 ± 0.353	**0.320** ± 0.170	-0.008 ± 0.024	0.099 ± 0.033
90	0.042 ± 0.042	0.114 ± 0.339	**0.320** ± 0.172	0.041 ± 0.037	0.119 ± 0.048
100	0.088 ± 0.087	0.116 ± 0.253	**0.231** ± 0.076	0.087 ± 0.069	0.144 ± 0.064

Table C.20: Cophenetic correlation on the animal with attributes (AWA) dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.010 ± 0.016	0.018 ± 0.040	**0.031** ± 0.022	0.017 ± 0.011	0.027 ± 0.013
3	0.010 ± 0.016	0.017 ± 0.043	0.031 ± 0.022	0.024 ± 0.012	**0.032** ± 0.015
4	0.011 ± 0.016	0.017 ± 0.042	**0.031** ± 0.022	0.022 ± 0.013	0.030 ± 0.013
5	0.011 ± 0.016	0.017 ± 0.044	0.031 ± 0.022	0.024 ± 0.015	**0.034** ± 0.014
6	0.011 ± 0.016	0.017 ± 0.046	0.031 ± 0.022	0.021 ± 0.018	**0.037** ± 0.016
7	0.011 ± 0.016	0.017 ± 0.046	0.031 ± 0.022	0.020 ± 0.020	**0.038** ± 0.016
8	0.011 ± 0.016	0.018 ± 0.046	0.031 ± 0.022	0.023 ± 0.018	**0.039** ± 0.015
9	0.011 ± 0.016	0.018 ± 0.048	0.031 ± 0.022	0.023 ± 0.019	**0.041** ± 0.015
10	0.011 ± 0.016	0.018 ± 0.050	0.031 ± 0.022	0.023 ± 0.022	**0.043** ± 0.015
15	0.011 ± 0.017	0.019 ± 0.055	0.031 ± 0.022	0.024 ± 0.018	**0.044** ± 0.013
20	0.011 ± 0.017	0.023 ± 0.067	0.032 ± 0.023	0.024 ± 0.022	**0.046** ± 0.013
25	0.012 ± 0.018	0.023 ± 0.072	0.032 ± 0.023	0.026 ± 0.025	**0.047** ± 0.013
30	0.012 ± 0.018	0.025 ± 0.077	0.032 ± 0.023	0.027 ± 0.028	**0.047** ± 0.013
40	0.015 ± 0.021	0.028 ± 0.090	0.032 ± 0.024	0.029 ± 0.029	**0.047** ± 0.013
50	0.020 ± 0.023	0.033 ± 0.105	0.033 ± 0.024	0.032 ± 0.033	**0.050** ± 0.015
60	0.026 ± 0.029	0.038 ± 0.116	0.033 ± 0.025	0.034 ± 0.036	**0.055** ± 0.018
70	0.045 ± 0.044	0.045 ± 0.131	0.037 ± 0.028	0.040 ± 0.042	**0.058** ± 0.019
80	0.048 ± 0.048	0.051 ± 0.144	0.044 ± 0.035	0.046 ± 0.047	**0.061** ± 0.020
90	0.052 ± 0.054	0.056 ± 0.157	0.044 ± 0.035	0.056 ± 0.054	**0.067** ± 0.023
100	0.065 ± 0.068	0.066 ± 0.178	0.058 ± 0.048	0.064 ± 0.060	**0.075** ± 0.027

Table C.21: Area under the curve of the R_{NX} index on the animal with attributes (AWA) dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.079	0.085 ± 0.007	0.081	0.096	**0.117**
3	0.079	0.086 ± 0.009	0.081	0.102	**0.121**
4	0.079	0.084 ± 0.008	0.080	0.099	**0.121**
5	0.080	0.086 ± 0.009	0.081	0.102	**0.121**
6	0.079	0.085 ± 0.008	0.082	0.100	**0.122**
7	0.079	0.086 ± 0.009	0.081	0.101	**0.122**
8	0.079	0.087 ± 0.010	0.081	0.105	**0.122**
9	0.079	0.085 ± 0.010	0.079	0.110	**0.119**
10	0.079	0.087 ± 0.011	0.079	0.110	**0.120**
15	0.079	0.088 ± 0.014	0.080	0.115	**0.128**
20	0.079	0.093 ± 0.016	0.113	0.113	**0.130**
25	0.080	0.093 ± 0.015	0.115	0.108	**0.131**
30	0.080	0.094 ± 0.017	0.121	0.110	**0.131**
40	0.080	0.097 ± 0.017	0.128	0.112	**0.131**
50	0.079	0.101 ± 0.016	**0.130**	0.112	**0.129**
60	0.077	0.100 ± 0.016	0.130	0.108	**0.134**
70	0.080	0.102 ± 0.014	0.128	0.102	**0.134**
80	0.078	0.102 ± 0.014	0.124	0.104	**0.135**
90	0.079	0.102 ± 0.013	0.123	0.104	**0.137**
100	0.087	0.105 ± 0.011	0.118	0.104	**0.139**

Table C.22: Clustering purity on the animal with attributes (AWA) dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.101	0.108 ± 0.013	0.117	**0.139**	0.127
3	0.100	0.110 ± 0.017	0.127	**0.144**	0.135
4	0.101	0.109 ± 0.018	0.126	**0.137**	0.135
5	0.100	0.110 ± 0.018	0.132	**0.142**	0.140
6	0.099	0.108 ± 0.017	0.131	**0.146**	0.141
7	0.098	0.107 ± 0.019	0.132	**0.143**	**0.142**
8	0.098	0.106 ± 0.018	0.124	**0.145**	0.140
9	0.098	0.107 ± 0.021	0.132	**0.145**	0.143
10	0.098	0.108 ± 0.022	0.134	**0.147**	0.144
15	0.099	0.108 ± 0.023	0.142	**0.155**	0.151
20	0.099	0.119 ± 0.029	0.143	**0.156**	0.154
25	0.100	0.119 ± 0.028	0.141	0.154	**0.158**
30	0.095	0.117 ± 0.031	0.141	0.149	**0.153**
40	0.096	0.125 ± 0.029	0.142	0.150	**0.152**
50	0.097	0.132 ± 0.024	0.142	0.147	**0.154**
60	0.094	0.134 ± 0.024	0.144	0.143	**0.156**
70	0.093	0.136 ± 0.022	0.140	0.139	**0.159**
80	0.092	0.137 ± 0.023	0.143	0.142	**0.163**
90	0.090	0.136 ± 0.024	0.143	0.140	**0.165**
100	0.115	0.140 ± 0.016	0.140	0.146	**0.164**

Table C.23: Clustering normalized mutual information on the animal with attributes (AWA) dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	2.001 ± 0.014	1.985 ± 0.045	**1.957** ± 0.006	1.995 ± 0.003	**1.967** ± 0.005
3	1.993 ± 0.019	1.968 ± 0.061	**1.935** ± 0.010	1.992 ± 0.004	1.988 ± 0.007
4	1.992 ± 0.014	1.965 ± 0.062	**1.957** ± 0.013	1.990 ± 0.005	**1.937** ± 0.016
5	1.989 ± 0.019	**1.953** ± 0.078	**1.936** ± 0.018	1.990 ± 0.006	1.991 ± 0.005
6	1.988 ± 0.024	1.936 ± 0.105	**1.916** ± 0.019	1.988 ± 0.010	1.961 ± 0.009
7	1.989 ± 0.029	1.921 ± 0.127	**1.893** ± 0.018	1.988 ± 0.011	1.986 ± 0.008
8	1.959 ± 0.028	**1.907** ± 0.132	**1.894** ± 0.026	1.982 ± 0.012	1.985 ± 0.007
9	1.976 ± 0.026	**1.903** ± 0.156	**1.896** ± 0.030	1.981 ± 0.013	1.986 ± 0.007
10	1.974 ± 0.031	**1.907** ± 0.161	**1.907** ± 0.031	1.980 ± 0.015	1.985 ± 0.007
15	1.969 ± 0.017	**1.868** ± 0.250	1.916 ± 0.037	1.977 ± 0.015	1.983 ± 0.006
20	1.973 ± 0.024	**1.912** ± 0.216	**1.930** ± 0.036	1.975 ± 0.019	1.980 ± 0.008
25	1.976 ± 0.025	**1.932** ± 0.162	**1.913** ± 0.039	1.971 ± 0.024	1.981 ± 0.007
30	1.979 ± 0.035	1.884 ± 0.272	**1.858** ± 0.039	1.971 ± 0.030	1.980 ± 0.008
40	1.977 ± 0.040	1.925 ± 0.173	**1.854** ± 0.045	1.969 ± 0.033	1.979 ± 0.009
50	1.975 ± 0.037	1.955 ± 0.112	**1.887** ± 0.047	1.972 ± 0.036	1.977 ± 0.009
60	1.974 ± 0.027	1.962 ± 0.121	**1.881** ± 0.049	1.972 ± 0.040	1.974 ± 0.011
70	1.974 ± 0.044	1.964 ± 0.111	**1.909** ± 0.048	1.972 ± 0.042	1.973 ± 0.012
80	1.975 ± 0.053	1.960 ± 0.132	**1.877** ± 0.048	1.972 ± 0.043	1.970 ± 0.012
90	1.975 ± 0.059	1.969 ± 0.107	**1.938** ± 0.047	1.972 ± 0.042	1.970 ± 0.011
100	**1.976** ± 0.041	**1.974** ± 0.094	**1.977** ± 0.046	**1.973** ± 0.043	**1.966** ± 0.013

Table C.24: Davies-Boulding index on the animal with attributes (AWA) dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
1	0.786 ± 0.012	1.196 ± 0.986	**2.000** ± 0.008	0.708 ± 0.013	0.772 ± 0.011
2	0.784 ± 0.016	1.528 ± 1.804	**3.000** ± 0.015	0.708 ± 0.007	0.771 ± 0.014
3	0.787 ± 0.012	1.864 ± 2.616	**4.000** ± 0.012	0.708 ± 0.006	0.773 ± 0.020
4	0.782 ± 0.015	2.200 ± 3.430	**5.000** ± 0.010	0.708 ± 0.011	0.764 ± 0.013
5	0.799 ± 0.014	2.539 ± 4.239	**6.000** ± 0.010	0.708 ± 0.013	0.763 ± 0.012
6	0.798 ± 0.019	2.878 ± 5.048	**7.000** ± 0.012	0.708 ± 0.012	0.769 ± 0.016
7	0.798 ± 0.017	3.211 ± 5.865	**8.000** ± 0.012	0.708 ± 0.011	0.766 ± 0.012
8	0.795 ± 0.010	3.541 ± 6.686	**9.000** ± 0.010	0.708 ± 0.013	0.779 ± 0.014
9	0.800 ± 0.014	3.877 ± 7.499	**10.000** ± 0.011	0.708 ± 0.012	0.777 ± 0.011
10	0.814 ± 0.011	5.554 ± 11.569	**15.000** ± 0.011	0.708 ± 0.011	0.784 ± 0.016
11	0.821 ± 0.011	7.227 ± 15.644	**20.000** ± 0.005	0.708 ± 0.006	0.782 ± 0.012
12	0.824 ± 0.014	8.894 ± 19.726	**25.000** ± 0.004	0.708 ± 0.005	0.786 ± 0.011
13	0.825 ± 0.012	10.553 ± 23.818	**30.000** ± 0.007	0.709 ± 0.009	0.764 ± 0.012
14	0.793 ± 0.018	13.854 ± 32.022	**40.000** ± 0.022	0.707 ± 0.014	0.769 ± 0.019
15	0.786 ± 0.015	17.176 ± 40.201	**50.000** ± 0.026	0.698 ± 0.014	0.749 ± 0.018
16	0.740 ± 0.016	20.491 ± 48.389	**60.000** ± 0.013	0.722 ± 0.019	0.750 ± 0.018
17	0.723 ± 0.016	23.816 ± 56.563	**70.000** ± 0.015	0.737 ± 0.017	0.750 ± 0.019
18	0.718 ± 0.015	27.145 ± 64.734	**80.000** ± 0.016	0.745 ± 0.017	0.748 ± 0.021
19	0.712 ± 0.015	30.475 ± 72.903	**90.000** ± 0.017	0.749 ± 0.016	0.745 ± 0.019
20	0.708 ± 0.015	33.807 ± 81.070	**100.000** ± 0.018	0.763 ± 0.015	0.750 ± 0.018

Table C.25: One-vs-one SVM classification accuracy on the Berkeley protein dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.282 ± 0.349	0.257 ± 0.555	0.207 ± 0.276	0.313 ± 0.310	**0.374** ± 0.257
3	0.250 ± 0.315	0.253 ± 0.556	0.203 ± 0.275	0.260 ± 0.286	**0.375** ± 0.189
4	0.213 ± 0.270	0.224 ± 0.489	0.193 ± 0.244	0.201 ± 0.268	**0.366** ± 0.150
5	0.189 ± 0.243	0.212 ± 0.464	0.191 ± 0.241	0.195 ± 0.270	**0.346** ± 0.115
6	0.172 ± 0.216	0.208 ± 0.456	0.198 ± 0.254	0.224 ± 0.244	**0.285** ± 0.076
7	0.161 ± 0.200	0.203 ± 0.452	0.216 ± 0.285	0.178 ± 0.211	**0.274** ± 0.065
8	0.157 ± 0.188	0.205 ± 0.453	0.223 ± 0.297	0.178 ± 0.210	**0.245** ± 0.053
9	0.143 ± 0.181	0.196 ± 0.451	0.229 ± 0.310	0.172 ± 0.187	**0.236** ± 0.048
10	0.139 ± 0.174	0.196 ± 0.458	**0.240** ± 0.326	0.162 ± 0.168	0.215 ± 0.043
15	0.099 ± 0.160	0.161 ± 0.397	**0.220** ± 0.287	0.119 ± 0.146	0.166 ± 0.041
20	0.076 ± 0.157	0.152 ± 0.412	**0.236** ± 0.310	0.072 ± 0.127	0.141 ± 0.044
25	0.067 ± 0.158	0.146 ± 0.411	**0.239** ± 0.323	0.042 ± 0.089	0.136 ± 0.053
30	0.059 ± 0.153	0.141 ± 0.398	**0.235** ± 0.325	0.036 ± 0.090	0.128 ± 0.061
40	0.063 ± 0.144	0.142 ± 0.397	**0.227** ± 0.325	-0.002 ± 0.078	0.115 ± 0.064
50	0.078 ± 0.131	0.142 ± 0.381	**0.228** ± 0.324	-0.004 ± 0.089	0.092 ± 0.064
60	0.084 ± 0.120	0.175 ± 0.348	**0.343** ± 0.237	-0.011 ± 0.090	0.091 ± 0.064
70	0.104 ± 0.115	0.177 ± 0.333	**0.331** ± 0.228	0.004 ± 0.103	0.088 ± 0.068
80	0.117 ± 0.113	0.179 ± 0.303	**0.304** ± 0.210	0.021 ± 0.100	0.097 ± 0.078
90	0.119 ± 0.116	0.176 ± 0.301	**0.304** ± 0.208	0.074 ± 0.093	0.106 ± 0.081
100	0.093 ± 0.127	0.168 ± 0.308	**0.308** ± 0.203	0.136 ± 0.084	0.108 ± 0.091

Table C.26: Cophenetic correlation on the Berkeley protein dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.078 ± 0.058	0.099 ± 0.164	**0.143** ± 0.112	0.097 ± 0.038	0.118 ± 0.037
3	0.078 ± 0.058	0.116 ± 0.206	**0.166** ± 0.139	0.101 ± 0.036	0.149 ± 0.038
4	0.078 ± 0.057	0.120 ± 0.222	**0.173** ± 0.150	0.099 ± 0.038	0.153 ± 0.034
5	0.078 ± 0.057	0.129 ± 0.247	**0.172** ± 0.152	0.102 ± 0.039	0.170 ± 0.039
6	0.078 ± 0.058	0.130 ± 0.252	**0.172** ± 0.154	0.127 ± 0.038	0.163 ± 0.035
7	0.078 ± 0.058	0.129 ± 0.250	**0.170** ± 0.154	0.122 ± 0.038	0.162 ± 0.034
8	0.079 ± 0.059	0.129 ± 0.252	**0.169** ± 0.154	0.125 ± 0.037	0.157 ± 0.034
9	0.079 ± 0.059	0.136 ± 0.271	**0.188** ± 0.178	0.132 ± 0.035	0.161 ± 0.036
10	0.079 ± 0.059	0.137 ± 0.274	**0.187** ± 0.179	0.131 ± 0.033	0.158 ± 0.034
15	0.079 ± 0.058	0.142 ± 0.291	**0.192** ± 0.191	0.124 ± 0.036	0.149 ± 0.036
20	0.079 ± 0.059	0.145 ± 0.303	**0.191** ± 0.193	0.128 ± 0.042	0.145 ± 0.037
25	0.080 ± 0.060	0.146 ± 0.305	**0.187** ± 0.190	0.122 ± 0.040	0.142 ± 0.037
30	0.081 ± 0.061	0.151 ± 0.315	**0.187** ± 0.190	0.123 ± 0.041	0.137 ± 0.038
40	0.084 ± 0.065	0.155 ± 0.322	**0.189** ± 0.192	0.119 ± 0.045	0.129 ± 0.034
50	0.100 ± 0.078	0.162 ± 0.325	**0.196** ± 0.197	0.123 ± 0.046	0.123 ± 0.034
60	0.169 ± 0.108	0.186 ± 0.328	**0.200** ± 0.201	0.123 ± 0.046	0.120 ± 0.033
70	0.175 ± 0.114	0.190 ± 0.333	**0.207** ± 0.206	0.127 ± 0.045	0.117 ± 0.031
80	0.180 ± 0.124	0.195 ± 0.337	**0.213** ± 0.209	0.131 ± 0.045	0.115 ± 0.030
90	0.182 ± 0.128	0.196 ± 0.340	**0.217** ± 0.211	0.137 ± 0.041	0.117 ± 0.030
100	0.194 ± 0.140	0.199 ± 0.345	**0.215** ± 0.214	0.142 ± 0.036	0.115 ± 0.029

Table C.27: Area under the curve of the R_{NX} index on the Berkeley protein dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.700	0.729 ± 0.042	**0.788**	0.700	0.700
3	0.700	0.728 ± 0.040	0.785	0.796	**0.807**
4	0.700	0.719 ± 0.027	0.757	0.700	**0.774**
5	0.700	0.700 ± 0.000	0.700	0.700	**0.753**
6	0.700	0.706 ± 0.009	**0.718**	0.700	0.700
7	**0.700**	**0.700** ± 0.000	**0.700**	**0.700**	**0.700**
8	0.700	0.732 ± 0.045	**0.795**	0.700	0.700
9	**0.700**	**0.700** ± 0.000	**0.700**	**0.700**	**0.700**
10	**0.700**	**0.700** ± 0.000	**0.700**	**0.700**	**0.700**
15	0.700	0.732 ± 0.046	**0.797**	0.700	0.700
20	0.700	0.734 ± 0.049	**0.803**	0.700	0.700
25	0.700	0.735 ± 0.049	**0.804**	0.700	0.700
30	**0.700**	**0.700** ± 0.000	**0.700**	**0.700**	**0.700**
40	**0.700**	**0.700** ± 0.000	**0.700**	**0.700**	**0.700**
50	**0.700**	**0.700** ± 0.000	**0.700**	**0.700**	**0.700**
60	**0.700**	**0.700** ± 0.000	**0.700**	**0.700**	**0.700**
70	**0.700**	**0.700** ± 0.000	**0.700**	**0.700**	**0.700**
80	**0.700**	**0.700** ± 0.000	**0.700**	**0.700**	**0.700**
90	**0.700**	**0.700** ± 0.000	**0.700**	**0.700**	**0.700**
100	**0.700**	**0.700** ± 0.000	**0.700**	**0.700**	**0.700**

Table C.28: Clustering purity on the Berkeley protein dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.027	0.137 ± 0.116	**0.297**	0.152	0.245
3	0.027	0.144 ± 0.120	0.309	0.295	**0.346**
4	0.032	0.128 ± 0.083	0.234	0.110	**0.286**
5	0.032	0.089 ± 0.069	0.186	0.104	**0.255**
6	0.032	0.092 ± 0.076	**0.199**	0.101	0.165
7	0.032	0.079 ± 0.057	**0.159**	0.085	0.154
8	0.032	0.126 ± 0.120	**0.296**	0.077	0.162
9	0.032	0.083 ± 0.059	**0.166**	0.077	0.161
10	0.032	0.084 ± 0.062	**0.171**	0.075	0.160
15	0.033	0.128 ± 0.129	**0.311**	0.082	0.160
20	0.033	0.146 ± 0.135	**0.335**	0.079	0.159
25	0.033	0.145 ± 0.139	**0.341**	0.078	0.154
30	0.033	0.057 ± 0.019	0.077	0.084	**0.154**
40	0.033	0.055 ± 0.016	0.070	0.078	**0.159**
50	0.033	0.055 ± 0.018	0.075	0.061	**0.150**
60	0.120	0.083 ± 0.026	0.072	0.073	**0.137**
70	0.120	0.080 ± 0.028	0.066	0.069	**0.166**
80	0.107	0.078 ± 0.022	0.071	0.064	**0.148**
90	0.107	0.081 ± 0.018	0.067	0.061	**0.155**
100	0.109	0.078 ± 0.023	0.055	0.071	**0.137**

Table C.29: Clustering normalized mutual information on the Berkeley protein dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	1.729 ± 0.225	1.657 ± 0.528	**1.608** ± 0.251	1.718 ± 0.252	1.627 ± 0.365
3	1.670 ± 0.225	**1.615** ± 0.635	**1.608** ± 0.356	1.632 ± 0.243	1.716 ± 0.154
4	1.744 ± 0.228	1.648 ± 0.606	**1.603** ± 0.288	1.671 ± 0.284	1.742 ± 0.139
5	1.759 ± 0.229	1.650 ± 0.600	**1.603** ± 0.300	1.687 ± 0.266	1.750 ± 0.143
6	1.761 ± 0.229	1.650 ± 0.590	**1.603** ± 0.275	1.683 ± 0.266	1.705 ± 0.292
7	1.773 ± 0.229	1.653 ± 0.587	**1.603** ± 0.256	1.679 ± 0.282	1.688 ± 0.344
8	1.667 ± 0.288	1.593 ± 0.623	**1.533** ± 0.302	1.657 ± 0.292	1.754 ± 0.204
9	1.785 ± 0.288	1.633 ± 0.623	**1.533** ± 0.254	1.650 ± 0.297	1.756 ± 0.203
10	1.781 ± 0.228	1.658 ± 0.586	**1.604** ± 0.254	1.663 ± 0.264	1.765 ± 0.188
15	1.701 ± 0.288	1.611 ± 0.594	**1.532** ± 0.237	1.697 ± 0.242	1.769 ± 0.183
20	1.717 ± 0.288	1.640 ± 0.563	**1.532** ± 0.213	1.655 ± 0.258	1.767 ± 0.181
25	1.691 ± 0.288	1.637 ± 0.554	**1.532** ± 0.229	1.708 ± 0.210	1.743 ± 0.220
30	1.824 ± 0.288	1.681 ± 0.564	**1.532** ± 0.193	1.798 ± 0.123	1.771 ± 0.187
40	1.821 ± 0.288	1.681 ± 0.564	**1.532** ± 0.200	1.695 ± 0.298	1.716 ± 0.248
50	1.824 ± 0.288	1.680 ± 0.574	**1.532** ± 0.216	1.653 ± 0.279	1.693 ± 0.316
60	1.842 ± 0.354	1.648 ± 0.643	**1.418** ± 0.193	1.793 ± 0.113	1.714 ± 0.279
70	1.850 ± 0.354	1.651 ± 0.643	**1.418** ± 0.190	1.724 ± 0.230	1.680 ± 0.341
80	1.844 ± 0.148	1.763 ± 0.475	1.757 ± 0.189	**1.703** ± 0.223	1.767 ± 0.184
90	1.864 ± 0.147	1.804 ± 0.332	1.757 ± 0.159	1.739 ± 0.174	**1.699** ± 0.311
100	1.867 ± 0.148	1.805 ± 0.328	1.756 ± 0.149	1.790 ± 0.153	**1.735** ± 0.181

Table C.30: Davies-Boulding index on the Berkeley protein dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.333 ± 0.016	0.331 ± 0.021	0.328 ± 0.013	0.347 ± 0.005	**0.392** ± 0.010
3	0.420 ± 0.011	0.373 ± 0.068	0.327 ± 0.010	0.413 ± 0.028	**0.437** ± 0.018
4	0.436 ± 0.012	0.382 ± 0.078	0.328 ± 0.013	**0.482** ± 0.031	0.466 ± 0.018
5	0.484 ± 0.012	0.407 ± 0.111	0.329 ± 0.011	**0.501** ± 0.017	0.491 ± 0.019
6	**0.518** ± 0.006	0.428 ± 0.130	0.338 ± 0.024	**0.523** ± 0.025	**0.521** ± 0.014
7	0.557 ± 0.010	0.453 ± 0.148	0.350 ± 0.021	0.546 ± 0.019	**0.569** ± 0.018
8	0.550 ± 0.018	0.451 ± 0.142	0.353 ± 0.020	**0.590** ± 0.011	**0.593** ± 0.018
9	0.553 ± 0.011	0.459 ± 0.136	0.365 ± 0.026	0.587 ± 0.015	**0.609** ± 0.017
10	0.578 ± 0.009	0.479 ± 0.140	0.381 ± 0.013	0.601 ± 0.016	**0.613** ± 0.010
15	0.653 ± 0.009	0.536 ± 0.166	0.419 ± 0.008	**0.672** ± 0.013	0.662 ± 0.006
20	0.655 ± 0.012	0.583 ± 0.105	0.510 ± 0.014	0.687 ± 0.012	**0.705** ± 0.016
25	0.641 ± 0.009	0.605 ± 0.054	0.569 ± 0.012	**0.676** ± 0.015	**0.680** ± 0.018
30	0.630 ± 0.022	0.612 ± 0.036	0.594 ± 0.014	0.653 ± 0.010	**0.682** ± 0.012
40	0.512 ± 0.020	0.576 ± 0.093	**0.639** ± 0.013	0.559 ± 0.024	0.612 ± 0.012
50	0.382 ± 0.018	0.518 ± 0.194	**0.655** ± 0.010	0.403 ± 0.017	0.481 ± 0.013
60	0.339 ± 0.010	0.501 ± 0.231	**0.664** ± 0.020	0.344 ± 0.015	0.393 ± 0.013
70	0.331 ± 0.014	0.499 ± 0.239	**0.667** ± 0.016	0.336 ± 0.013	0.374 ± 0.011
80	0.317 ± 0.013	0.493 ± 0.250	**0.669** ± 0.016	0.322 ± 0.009	0.365 ± 0.013
90	0.319 ± 0.015	0.487 ± 0.239	**0.656** ± 0.014	0.322 ± 0.011	0.363 ± 0.015
100	0.319 ± 0.011	0.480 ± 0.229	**0.642** ± 0.011	0.320 ± 0.011	0.361 ± 0.014

Table C.31: One-vs-one SVM classification accuracy on the Cora dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	**0.339**	**0.337** ± 0.002	0.335	0.302	0.302
3	**0.355**	0.345 ± 0.010	0.335	0.344	0.310
4	0.349	0.342 ± 0.007	0.335	**0.363**	0.308
5	**0.384**	0.359 ± 0.025	0.335	0.350	**0.384**
6	**0.404**	0.369 ± 0.035	0.335	**0.405**	0.362
7	0.397	0.366 ± 0.031	0.335	**0.417**	0.367
8	0.397	0.366 ± 0.031	0.335	**0.404**	0.351
9	**0.399**	0.367 ± 0.032	0.335	0.390	0.347
10	0.381	0.358 ± 0.023	0.335	**0.411**	0.378
15	0.390	0.352 ± 0.038	0.314	**0.402**	0.353
20	0.302	0.330 ± 0.028	**0.358**	0.338	**0.359**
25	0.302	0.331 ± 0.029	0.360	0.387	**0.423**
30	0.302	0.330 ± 0.028	0.358	0.302	**0.425**
40	0.303	0.326 ± 0.024	0.350	0.302	**0.421**
50	0.308	0.335 ± 0.027	0.362	0.302	**0.460**
60	0.302	0.332 ± 0.030	0.362	0.309	**0.391**
70	0.302	0.336 ± 0.034	0.370	0.302	**0.390**
80	0.302	0.336 ± 0.034	0.371	0.302	**0.390**
90	0.302	0.341 ± 0.038	0.379	0.305	**0.448**
100	0.309	0.343 ± 0.035	**0.378**	0.303	0.366

Table C.32: Clustering purity on the Cora dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

K	Single view			Stacked views	MVSC-CEV
	Worst	Average	Best		
2	0.027	0.051 ± 0.024	**0.075**	0.009	0.029
3	0.027	0.069 ± 0.042	**0.111**	0.086	0.038
4	0.027	0.071 ± 0.044	0.115	**0.118**	0.086
5	0.027	0.081 ± 0.054	**0.135**	0.116	0.112
6	0.027	0.091 ± 0.063	**0.154**	0.147	0.110
7	0.027	0.087 ± 0.060	0.147	**0.166**	0.128
8	0.027	0.087 ± 0.060	0.147	**0.160**	0.122
9	0.027	0.088 ± 0.060	**0.148**	**0.147**	0.109
10	0.027	0.081 ± 0.054	0.135	**0.186**	0.118
15	0.022	0.073 ± 0.051	0.124	**0.159**	0.113
20	0.068	0.060 ± 0.008	0.052	0.058	**0.115**
25	0.091	0.058 ± 0.033	0.025	0.125	**0.189**
30	0.090	0.053 ± 0.037	0.016	0.023	**0.190**
40	0.108	0.066 ± 0.042	0.024	0.020	**0.195**
50	0.122	0.074 ± 0.048	0.026	0.016	**0.230**
60	0.117	0.067 ± 0.051	0.016	0.023	**0.153**
70	0.111	0.064 ± 0.046	0.018	0.028	**0.137**
80	0.112	0.063 ± 0.049	0.014	0.025	**0.138**
90	0.094	0.060 ± 0.035	0.025	0.021	**0.222**
100	**0.125**	0.072 ± 0.053	0.020	0.015	0.078

Table C.33: Clustering normalized mutual information on the Cora dataset of MVSC-CEV compared with single view and stacked views SC. K is the dimensionality of the projection.

Bibliography

[1] Ali E Abbas. A Kullback-Leibler View of Linear and Log-Linear Pools. *Decision Analysis*, 6(1):25–37, 2009.

[2] S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Altschul et al.. 1990. Basic Local Alignment Search Tool.pdf, 1990.

[3] Massih Amini, Nicolas Usunier, and Cyril Goutte. Learning from Multiple Partially Observed Views – an Application to Multilingual Text Categorization. *Advances in Neural Information Processing Systems 22*, pages 28–36, 2009.

[4] Michael Bacharach. Normal Bayesian dialogues. *Journal Of The American Statistical Association*, 74(368):837–846, 1979.

[5] J. M. Bates and C. W. J. Granger. The Combination of Forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969.

[6] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[7] Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 15(6):1373–1396, 2003.

[8] Donald Berndt and James Clifford. Using dynamic time warping to find patterns in time series. *Workshop on Knowledge Knowledge Discovery in Databases*, 398:359–370, 1994.

[9] M Breukelen, R P W Duin, D M J Tax, and J E Hartog. Handwritten Digit Recognition by Combined Classifier (1998).pdf. *Kybernetika*, 34(4):[381]–386, 1998.

[10] S Lawrence C. Giles, K.D. Bollacker. CiteSeer: An automatic citation indexing system. *Digital Libraries 98: Third ACM Conference on Digital Libraries*, pages 89–98, 1998.

[11] Xiao Cai, Feiping Nie, and Heng Huang. Multi-View K -Means Clustering on Big Data. In *The 23rd International Joint Conference on Artificial Intelligence*, pages 2598–2604, 2013.

[12] Xiao Cai, Feiping Nie, Heng Huang, and Farhad Kamangar. Heterogeneous Image Features Integration via Multi-View Spectral Clustering. In *Cvpr*, pages 1977–1984, 2011.

[13] E. J. Candes, X. Li, Y. Ma, J. Wright, E. J. Candes, X. Li, and Y. Ma. Robust Principal Component Analysis? *Preprint*, page 41, 2009.

[14] Rich Caruana. Multitask Learning. *Machine Learning*, 28(1):41–75, 1997.

[15] Arthur Carvalho and Kate Larson. A Consensual Linear Opinion Pool. 2012.

[16] Kamalika Chaudhuri, Sham M. Kakade, Karen Livescu, and Karthik Sridharan. Multi-view clustering via canonical correlation analysis. In *Icml*, ICML '09, pages 1–8, New York, NY, USA, 2009. ACM.

[17] Xiangyu Chen, Yadong Mu, Shuicheng Yan, and Tat-Seng Chua. Efficient large-scale image annotation by probabilistic collaborative multi-label propagation. *Proceedings of the international conference on Multimedia*, (January 2010):35–44, 2010.

[18] Sabina Chiaretti, Xiaochun Li, Robert Gentleman, Antonella Vitale, Marco Vignetti, Franco Mandelli, Jerome Ritz, and Robin Foa. Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood*, 103(7):2771–2778, 2004.

[19] Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng. NUS-WIDE: A Real-World Web Image Database from National University of Singapore. *Acmmm*, page 1, 2009.

[20] Robert T. Clemen and Robert L. Winkler. Combining Probability Distributiond from experts in Risk Analysis. *Risk Analysis*, 19(2):155–156, 1999.

[21] Roger M Cooke. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. 1991.

[22] Jose Costa Pereira, Emanuele Coviello, Gabriel Doyle, Nikhil Rasiwasia, Gert R G Lanckriet, Roger Levy, and Nuno Vasconcelos. On the role of correlation and abstraction in cross-modal multimedia retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):521–535, 2014.

[23] Trevor F Cox and Michael A A Cox. *Multidimensional Scaling, Second Edition*, volume 88. 2000.

[24] M Craven, D DiPasquo, D Freitag, A McCallum, T Mitchell, K Nigam, and S Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. *World Wide Web Internet And Web Information Systems*, (1):509–516, 1998.

[25] N Dalal and B Triggs. Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1:886–893, 2005.

[26] D L Davies and D W Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, 1(2):224–227, 1979.

[27] Fernando De La Torre. A least-squares framework for component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(6):1041–1055, 2012.

[28] Virginia R. de Sa and Dana H. Ballard. Category Learning Through Multimodality Sensing. *Neural Computation*, 10(5):1097–1117, 1998.

[29] V.R. de Sa, Virginia R De Sa, San Diego, and La Jolla. Spectral Clustering with Two Views. *Development*, pages 20–27, 1998.

[30] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[31] Mark Van der Laan, Katherine Pollard, and Jennifer Bryan. A new partitioning around medoids algorithm. *Journal of Statistical Computation and Simulation*, 73(8):575–584, 2003.

[32] Jean Antoine Désidéri. Multiple-gradient descent algorithm (MGDA) for multiobjective optimization. *Comptes Rendus Mathematique*, 350(5-6):313–318, 2012.

[33] Zhengming Ding and Yun Fu. Low-Rank Common Subspace for Multi-view Learning. In *Proceedings - IEEE International Conference on Data Mining, ICDM*, volume 2015-Janua, pages 110–119, 2015.

[34] Pinar Duygulu, Kobus Barnard, J De Freitas, David Forsyth, and Nando de Freitas. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. *Proc. European Conference on Computer Vision (ECCV)*, 2353:97–112, 2002.

[35] Martin Ester, Hans P Kriegel, Jorg Sander, and Xiaowei Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pages 226–231, 1996.

[36] Theodoros Evgeniou and Massimiliano Pontil. Regularized multi-task learning. *International Conference on Knowledge Discovery and Data Mining*, page 109, 2004.

[37] Bernhard N. Flury. Common Principal Components in K Groups. *Journal of the American Statistical Association*, 79(388):892–898, 1984.

[38] Ellen H. Fukuda and Luis Mauricio Graña Drummond. a Survey on Multiobjective Descent Methods. *Pesquisa Operacional*, 34(3):585–620, 2014.

[39] Jean Gallier. The Schur Complement and Symmetric Positive Semidefinite (and Definite) Matrices. *Complement*, pages 1–12, 2010.

[40] Christian Genest and Kevin J. McConway. Allocating the weights in the linear opinion pool. *Journal of Forecasting*, 9(1):53–73, 1990.

[41] Bo Geng, Dacheng Tao, Chao Xu, Linjun Yang, and Xian Sheng Hua. Ensemble manifold regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(6):1227–1233, 2012.

[42] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[43] Derek Greene. A Matrix Factorization Approach for Integrating Multiple Data Views. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases: Part I*, ECML PKDD '09, pages 423–438, Berlin, Heidelberg, 2009. Springer-Verlag.

[44] Derek Greene and Padraig Cunningham. Practical solutions to the problem of diagonal dominance in kernel document clustering. *Proceedings of the 23rd International Conference on Machine Learning (ICML'06)*, pages 377–384, 2006.

[45] Sven Ove Hansson. Decision Theory. *Technology*, 19(1):1–94, 2005.

[46] David R Hardoon, Sandor Szedmak, and John Shawe-Taylor. Canonical correlation analysis: an overview with application to learning methods. *Neural computation*, 16(12):2639–64, 2004.

[47] David R Hardoon, Sandor Szedmak, John Shawe-Taylor, Jorge G Adrover, Stella M Donato, N Balakrishnan, A Capitanio, B Scarpa, Rossella Di Leonardo, Giada Adelfio, Adriana Bellanca, Marcello Chiodi, Salvatore Mazzola, Tetsuro Sakurai, Tomoya Yamada, Hyejin Shin, Seokho Lee, Y Akba, C Takma, Multivariate Data Analysis, E Rolph, Ronald L Tatham, Renato Leoni, Multivariate Data Analysis, E Rolph, Ronald L Tatham, Spring Quarter, Robin K Henson, B Khalil, T B M J Ouarda, A St-Hilaire, W K J Renema, Hermien E Kan, B?? Wieringa, Arend Heerschap, Jp Gonzalez, Ja Bagnell, Simon Cook, Thomas Oberthur, Steen a. Andersson, Jesse B Crawford, Rural Water Management, Applied Life Sciences, Jh Meyer, P Rein, P Turner, K Mathias, Phesheya Dlamini, The Harvard, O Neville, Laxman Hegde, A Vinokourov, John Shawe-Taylor, N Cristianini, Multivariate Data Analysis, E Rolph, Ronald L Tatham, Klara Verbyla, South Africa, Klara Verbyla, L Sun, S Ji, S Yu, J Ye, Phuti Sebatjane, Phuti Sebatjane, Fábio R Marin, James W Jones, Sylvie Combes, Ignacio González, Sébastien Déjean, Alain Baccini, Nathalie Jehl, Hervé Juin, Laurent Cauquil, Béatrice Gabinaud, François Lebas, Catherine Larzul, Christophe Coudun, Jean Claude Gégout, Christian Piedallu, Jean Claude Rameau, R C Gore, Scientific Writing, Health Sciences, Jussi Paananen, Ahmed Douaik, Marc Van Meirvenne, Tibor Tóth, Det Norske Veritas, Directorate Marketing, Keith L Bristow, Brian A Keating, and The Mathematica. Canonical Correlation Analysis. *Journal of Multivariate Analysis*, 25(1):1–6, 2014.

[48] J. A. Hartigan and M. A. Wong. A K-Means Clustering Algorithm. *Applied Statistics*, 28(1):100–108, 1979.

[49] Geoffrey E Hinton and Sam T Roweis. Stochastic neighbor embedding. *Advances in neural information processing systems*, pages 833–840, 2002.

[50] Chih Wei Hsu and Chih Jen Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415–425, 2002.

[51] Thorsten Joachims. A statistical learning learning model of text classification for support vector machines. *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 128–136, 2001.

[52] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

[53] I T Jolliffe. Principal Component Analysis. *Encyclopedia of Statistics in Behavioral Science*, 3:1580–1584, 2005.

[54] Samir Kanaan-Izquierdo and Alexandre Perera-Lluna. Multiview multidimensional scaling. Technical report, B2SLAB - CREB - UPC, 2017.

[55] Samir Kanaan-Izquierdo and Alexandre Perera-Lluna. Multiview t-distributed stochastic neighbour embedding. Technical report, B2SLAB - CREB - UPC, 2017.

[56] Samir Kanaan-Izquierdo, Andrey Ziyatdinov, Raimon Massanet, and Alexandre Perera. Multiview approach to spectral clustering. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 1254–1257, 2012.

[57] Samir Kanaan-Izquierdo, Andrey Ziyatdinov, and Alexandre Perera-Lluna. Multiview: an R package for multiview pattern recognition. Technical report, B2SLAB - CREB - UPC, 2017.

[58] Samir Kanaan-Izquierdo, Andrey Ziyatdinov, and Alexandre Perera-Lluna. Multiview and multifeature spectral clustering using common eigenvectors. *Pattern Recognition Letters*, 2017.

[59] L Kaufman and P J Rousseeuw. Clustering by means of medoids, 1987.

[60] J B Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[61] J. B. Kruskal. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2):115–129, 1964.

[62] Abhishek Kumar and Hal Daumé. A co-training approach for multi-view spectral clustering. *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 393–400, 2011.

[63] Abhishek Kumar, Piyush Rai, and Hal Daume. Co-regularized Multi-view Spectral Clustering. *Nips*, pages 1413–1421, 2011.

[64] Jack Kyte and Russell F. Doolittle. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1):105–132, 1982.

[65] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, pages 951–958, 2009.

[66] Gert R G Lanckriet, Tijl De Bie, Nello Cristianini, Michael I Jordan, and William Stafford Noble. A statistical framework for genomic data fusion. 20(16):2626–2635, 2004.

[67] D D Lee and H S Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[68] John A. Lee, Emilie Renard, Guillaume Bernard, Pierre Dupont, and Michel Verleysen. Type 1 and 2 mixtures of Kullback-Leibler divergences as cost functions in dimensionality reduction based on similarity preservation. *Neurocomputing*, 112:92–108, 2013.

[69] Jiayi Li, Hongyan Zhang, and Liangpei Zhang. Superpixel-Level Multitask Joint Sparse Representation for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, (June), 2015.

[70] Yeqing Li, Feiping Nie, Heng Huang, and Junzhou Huang. Large-Scale Multi-View Spectral Clustering with Bipartite Graph. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 2750–2756, 2011.

[71] S X Liao and M Pawlak. Image analysis with Zernike moment descriptors. In *Electrical and Computer Engineering, 1997. Engineering Innovation: Voyage of Discovery. IEEE 1997 Canadian Conference on*, volume 2, pages 700—-703 vol.2, 1997.

[72] David G. Lowe. Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 2([8):1150–1157, 1999.

[73] Qing Lu and Lise Getoor. Link-based Classification. *Proceedings of the 20th International Conference on Machine Learning*, page 8, 2003.

[74] Laurens Van Der Maaten. Learning a Parametric Embedding by Preserving Local Structure. *JMLR Proceedings vol. 5 (AISTATS)*, pages 384–391, 2009.

[75] Laurens Van Der Maaten. Fast Optimization for t-SNE. *Neural Information Processing Systems (NIPS) 2010 Workshop on Challenges in Data Visualization*, 1(1):1–5, 2010.

[76] Laurens Van Der Maaten. {<}Van Der Maaten - 2014 - Accelerating t-{{}SNE{}} using tree-based algorithms.pdf{>}. *Journal of Machine Learning Research*, 15:3221–3245, 2014.

[77] Mehrdad Mahdavi, Tianbao Yang, and Rong Jin. Stochastic Convex Optimization with Multiple Objectives. *Advances in Neural Information Processing Systems*, pages 1115–1123, 2013.

[78] Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. Introduction to Information Retrieval. *Journal of the American Society for Information Science and Technology*, 1:496, 2008.

[79] Andres McCallum and Kamal Nigam. A Comparison of Event Models for Naive Bayes Text Classification. *AAAI/ICML-98 Workshop on Learning for Text Categorization*, pages 41–48, 1998.

[80] Tom Mitchell and Avrim Blum. Combining labeled and unlabeled data with co-training. *Proceedings of the eleventh annual conference on Computational learning theory*, pages 92–100, 1998.

[81] A Moore. K-means and Hierarchical Clustering. *Statistical Data Mining Tutorials*, pages 1–24, 2001.

[82] Fionn Murtagh and Pedro Contreras. Methods of Hierarchical Clustering. *Computer*, 38(2):1–21, 2011.

[83] S Nene, S Nayar, and H Murase. Columbia Object Image Library (COIL-20). *Technical Report*, 95:223–303, 1996.

[84] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Nips*, 14(14):849–856, 2001.

[85] Feiping Nie, Heng Huang, Xiao Cai, and Chris Ding. Efficient and Robust Feature Selection via Joint. *Advances in Neural Information Processing Systems*, 23:1813–1821, 2010.

[86] Shibin Parameswaran and Kilian Q Weinberger. Large Margin Multi-Task Metric Learning. *Nips*, pages 1–9, 2010.

[87] Max Planck and Ulrike Von Luxburg. A Tutorial on Spectral Clustering A Tutorial on Spectral Clustering. *Statistics and Computing*, 17(March):395–416, 2006.

[88] Novi Quadrianto and Christoph H Lampert. Learning multi-view neighborhood preserving projections. *Icml*, pages 425–432, 2011.

[89] R Development Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing Vienna Austria*, 0:{ISBN} 3–900051–07–0, 2016.

[90] C E Shannon and W Weaver. The Mathematical Theory of Communication. *The mathematical theory of communication*, 27(4):117, 1949.

[91] Ling Shao, Li Liu, and Mengyang Yu. Kernelized Multiview Projection for Robust Action Recognition. *International Journal of Computer Vision*, 118(2):115–129, 2016.

[92] Jianbo Shi and Jitendra Malik. Normalized Cuts and Image Segmentation Normalized Cuts and Image Segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(March):888–905, 2005.

[93] T F Smith and M S Waterman. Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1):195–197, 1981.

[94] Robert R Sokal and F James Rohlf. The Comparison of Dendrograms by Objective Methods. *Taxon*, 11(2):33–40, 1962.

[95] Erik L L Sonnhammer, Sean R Eddy, Ewan Birney, Alex Bateman, and Richard Durbin. Pfam: Multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Research*, 26(1):320–322, 1998.

[96] G W Stewart and Harcourt Bruce Jovanovich. Matrix perturbation theory. *Mathematics and Computers in Simulation*, 33(1):74, 1991.

[97] Alexander Strehl and Joydeep Ghosh. Cluster Ensembles A Knowledge Reuse Framework for Combining Multiple Partitions. *Journal of Machine Learning Research*, 3:583–617, 2002.

[98] Tadelis, Drew Fudenberg, and Jean Tirole. *Game Theory*, volume 2008. 1984.

[99] Xiaoou Tang. Texture information in run-length matrices. *IEEE Transactions on Image Processing*, 7(11):1602–1609, 1998.

[100] Nickolay T. Trendafilov. Stepwise estimation of common principal components. *Computational Statistics and Data Analysis*, 54(12):3446–3457, 2010.

[101] L J P van der Maaten and G E Hinton. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, 2008.

[102] Laurens Van Der Maaten, Geoffrey Hinton, and Geoffrey Hinton van der Maaten. Visualizing Data using t-SNE, 2008.

[103] T van Erven and P Harremoes. Renyi Divergence and Kullback-Leibler Divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

[104] Sandro Vega-Pons and José Ruiz-Shulcloper. a Survey of Clustering Ensemble Algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(03):337–372, 2011.

[105] Luis von Ahn and Laura Dabbish. Labeling images with a computer game. *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, pages 319–326, 2004.

[106] Hua Wang, Feiping Nie, and Heng Huang. Multi-view clustering and feature learning via structured sparsity. *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 28:352–360, 2013.

[107] Martha White, Yaoliang Yu, Xinhua Zhang, and Dale Schuurmans. Convex Multi-view Subspace Learning. *Proc.NIPS*, pages 1673–1681, 2012.

[108] Tian Xia, Dacheng Tao, Tao Mei, and Yongdong Zhang. Multiview spectral embedding. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(6):1438–1446, 2010.

[109] Shiming Xiang, Feiping Nie, Gaofeng Meng, Chunhong Pan, and Changshui Zhang. Discriminative least squares regression for multiclass classification and feature selection. *IEEE Transactions on Neural Networks and Learning Systems*, 23(11):1738–1754, 2012.

[110] Bo Xie, Yang Mu, Dacheng Tao, and Kaiqi Huang. M-SNE: Multiview stochastic neighbor embedding. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 41(4):1088–1096, 2011.

[111] Zhe Xue, Guorong Li, and Qingming Huang. Joint Multi-View Representation Learning and Image Tagging. *Aaai*, pages 1366–1372, 2016.

[112] Qiyue Yin, Shu Wu, Ran He, and Liang Wang. Multi-view clustering via pairwise sparse subspace representation. *Neurocomputing*, 156:12–21, 2015.

[113] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 68(1):49–67, 2006.

[114] Lefei Zhang, Qian Zhang, Liangpei Zhang, Dacheng Tao, Xin Huang, and Bo Du. Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding. *Pattern Recognition*, 48(10):3102–3112, 2015.

[115] Qian Zhang, Lefei Zhang, Bo Du, Wei Zheng, Wei Bian, and Dacheng Tao. MMFE: Multitask multiview feature embedding. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2016-Janua:1105–1110, 2016.

[116] Tianhao Zhang, Dacheng Tao, Xuelong Li, and Jie Yang. Patch alignment for dimensionality reduction. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1299–1313, 2009.

[117] Xuran Zhao, Nicholas Evans, and Jean-Luc Dugelay. A subspace co-training framework for multi-view clustering. *Pattern Recognition Letters*, 41(0):73–82, 2014.