

Dimensionality reduction techniques for high dimensional data: State of the art

Suixin Jiang

Department of Math & Statistics, American University, 4400 Massachusetts Avenue,
NW, Washington, DC 20016, United States

sj6778a@alumni.american.edu

Abstract. In the past decades, tons of data have been generated every single day as people increasingly rely on electronic products and networks in their lives. The developments in techniques and storage capabilities provide a fundamental condition for analyzing those huge-volumed data. Each captured feature represents a dimension of the data. So far, high dimensional data analysis has become a challenging task in various study fields. Redundant and irrelevant features can be removed by using different dimensionality reduction techniques. Effective information extraction can be achieved by using the proper method. This paper reviews the most widely used dimensionality reduction techniques and their application fields. It can be found that though DRTs have been successfully applied to many areas (i.e., image, audio/video data, biomedical), DRTs still need to be improved and developed to achieve better classification and prediction accuracy. Inter-method combinations will remain the focus of research in the future. Computation time and cost may not be a limitation anymore as the computation power of the computer is still developing. So, the development of the algorithm is becoming particularly important. This study provides a brief introduction to widely used DRTs and their variants, it will be helpful for understanding HDD analysis, and more DRTs will be researched in future work.

Keywords: high dimensional data, dimensionality reduction technique.

1. Introduction

Tremendous data has been created and recorded continuously in different application fields during the past decades. Meanwhile, the data's dimensionality, complicity, and scale are all rising incredibly. High Dimensional Data (HDD) has been applied to many different areas including biomedical, genetics, education, the internet, etc. [1]. The new evolved formats of a huge amount of HDD have been included but not limited to text, speech signals, digital images, and videos.

When conducting Machine Learning (ML) methods to analyze HDD, the high dimensionality can present numerous challenges in accurate classification, pattern recognition, and visualization. Due to high computational complexity, a study in dozens or even hundreds of dimensions would be formidable. The curse of dimensionality represents increasing dimensionality to a fixed model can lead to overfitting. The only solution is to add tremendously new data for each extra dimension [2].

Feature extraction and feature selection are two ways of reducing high dimensionality. By combining the current features, feature extraction produces a subset of new features. The new feature set preserves

the original dataset's information at the maximum level. Feature selection is choosing the most instructive feature subset from the data. Applying a proper technique to reduce the dimensions can significantly save time and decrease the effort required to extract or select informative features for analysis.

Before applying ML models, Dimensionality Reduction Techniques (DRTs) provide an effective solution to cut down on the number of original variables. Numerous DRTs can be used to improve resource utilization and reduce computing time. These techniques are applicable to many kinds of data [3]. However, since each method was designed to maintain a certain characteristic of the original data, implementing DRTs is very challenging. A specific DRT could be problematic for some applications and appropriate for others.

Moreover, the core functionality of the DRT can be changed by modifying the internal parameters. This paper provides a basic summary of DRTs and their practical variants. Numerous potential problems of DRT have been emphasized. The following sections are organized in this order, Section 2 explores three widely used DRTs and their variants, Section 3 presents the common issues of both HDD and DRTs, and Section 4 presents a brief discussion.

2. Dimensionality reduction techniques

The number of redundant and unnecessary features can be reduced from the original HDD by using a proper DRT. It converts HDD into a dimension-reduced representation while maintaining the original features as much as possible. It is simple and effective to process, analyze, and visualize low dimensional data. DRT utilization can also speed up computation time and reduce storage requirements.

Formally, DRT transforms an HDD $Y = [y_1, y_2, y_3, \dots, y_m] \in \mathbb{R}^{m \times k}$ having m observations with k dimensions can be transformed into a low dimensional data $Z = [z_1, z_2, z_3, \dots, z_m] \in \mathbb{R}^{m \times j}$ having m observations with j dimensions, where $j \ll k$ in ideal condition [4]. Table 1 shows some main discriminative features of DRTs.

Table 1. Discriminative features of DRTs.

Feature	Feature Description
Linear technique	Low dimension results from a linear combination. Simple geometric interpretation and computational efficiency.
Non-Linear technique	Low dimension result from a nonlinear transformation. Dealing with complex non-linear data. Parameters need to be optimized.
Supervised learning	Using labeled input data, therefore knowing output data.
Unsupervised learning	Using unlabeled input data for discovering patterns.

2.1. Principal component analysis

Pearson first introduced Principal component analysis (PCA). He proposed this method as a discovery of the closest-fitting lines and planes to a set of spatial points. Hotelling then developed it independently. His derivation model is regarded as a sign of the maturity of PCA. PCA is the oldest and most widely used DRT in multivariate analysis. It analyzes data where a few inter-correlated quantitative dependent variables are often used to record the observations. The data dimensions can be reduced by various PCA strategies while retaining as much data variance as possible [5]. Different algorithms, such as eigenvalues, factor analysis, and linear regression can be used to compute PCA. As a fundamental mathematical analysis method, PCA is widely used in practical applications such as demography, geography, image, speech, visualization, and other disciplines. It aims to pick the most informative features out of the original HDD and then represent them as a new collection of independent features named principal components (PCs). Normally, the first picked-out principal component occupies the

biggest variance of the HDD, and the second picked-out principal component occupies the biggest variance of the rest HDD, etc.

Mathematically, let Y be a data matrix for dimension $m \times k$ (rows \times columns), which has a column-wise zero empirical mean. Observation and variables are represented by each row and column. It is possible to calculate the PCs $z_i \in \mathbb{R}^m$ as a linear weighted combination of features.

$$Z = YA \quad (1)$$

With $Z = [z_1, z_2, z_3, \dots, z_k] \in \mathbb{R}^{m \times k}$ and $A = [a_1, a_2, a_3, \dots, a_k] \in \mathbb{R}^{k \times k}$. Generally, only the top j dominating PCs are concerned since they account for most of the data variability. So, by easily using the Singular Value Decomposition, PCA enables one to effectively reduce the dimensions from k to j . The dominating j PCs can be utilized for showing the original data in dimension-reduced space and for classification and regression tasks [6].

Figure. 1 illustrates the PCA of a dataset symbolized within an elliptical subspace and the separation of two major principal components from a set of variables. In the past few decades, PCA has been developed into many different variants and applied to various study fields. Some PCA variants will be briefly introduced in the following paragraphs. Hubert et al. [7] pointed out the first extracted PCs might not fully represent the variance of the original data since classical PCA is particularly sensitive to outlying observations. By combining robust scatter matrix estimation with projection pursuit ideas, they proposed Robust PCA (ROBPCA), which shows more accuracy in both noncontaminated and contaminated data. Serneels and Verdonck [8] developed a new expectation robust (ER) algorithm to apply the robust PCA to data that contains missing values. Based on a simulation study, they concluded that ER approach performs well on different sizes of datasets.

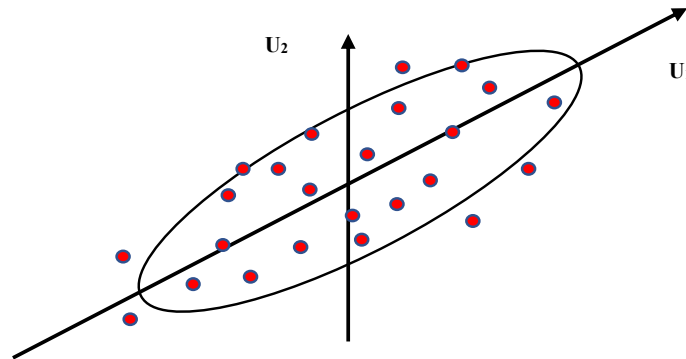


Figure 1. Analysis of principal component.

The Generalized PCA (GPCA) problem, the identification of mixtures of subspaces problem, was addressed by Vidal et al. [9] in their innovative algebraic geometric proposal. They demonstrated that GPCA is comparable to factoring a homogeneous polynomial in the absence of noise, and they also came up with a formula for the number of subspaces. Additionally, GPCA was evaluated on fictitious data. Several 2D and 3D segmentation applications were shown. Wang et al. [10] proposed a method of tracking moving objects from videos. It involves a combination of an incremental 2DPCA object characterization and a Maximum Likelihood blob object region tracking. The updated algorithm for the projected covariance matrices significantly reduces the computation time.

PCA methods are not intended for multilevel analysis. For extracting core inter and intra geometric components from multilevel data, Di et al. [11] proposed Multilevel Functional PCA, which combines standard multilevel mixed models and classical PCA. Multilevel functional PCA enables an effective breakdown of the observed functional variability. Years later, Multivariate Functional PCA (MFPCA), another novel statistical approach for data on diverse fields that varies in feature, was proposed by Happ and Greven [12]. A straightforward theoretical connection between single- and multiple-variable

functional PCA was developed. And based on the univariate counterparts, MFPCA provides a simple estimate method for multivariate functional principal components. Therefore, it can be used for sparse or measurement-erroneous data.

Metsalu and Vilo [13] presented a web tool ClustVis, which is established on PCA and Heatmap. It can show how multivariate data clusters. The potential application fields include but are not limited to gene expression study and high-resolution image analysis for cancer prognosis. The outputs from PCA and Heatmap can be downloaded in different formats with multiple detail options. Journée et al. [14] developed a new approach to Sparse Principal Component Analysis (SPCA). Four novel algorithms are included for computing a matrix's sparse principal components. The search space' dimension can be reduced drastically when the data matrix has much fewer rows than columns. The analysis of pitprops data and gene expression data demonstrated Generalized Power approach is performing better than the existing method in both the quality of the results and computational efficiency. A method that combined feature selection and subspace learning to remove redundant and irrelevant features was introduced by Yi et al. [15] named Joint Sparse PCA (JSPCA). Particularly, by applying $L_{2,1}$ -norms to a loss term and a regularization term. This method can eliminate pointless variables and lessen the impact of outliers. The experimental results showed that JSPCA has better feasibility and effectiveness than other PCA variants.

2.2. Linear discriminant analysis

In general, the linear discriminant function presented by Fisher was thought to be the earliest version of Linear Discriminant Analysis (LDA). It involves the methods that are used in Statistics. LDA seeks to define or differentiate two categories of objects or occurrences by finding a linear combination of their features. To reduce the dimension for further classification. The resultant combination can be employed as a linear classifier. Concretely, it ensures the greatest amount of class separability [16]. Some of the applications of LDA include face recognition, voice recognition, credit card fraud detection, tumors distinguishing between known classes, and so on.

The original data matrix is intended to be mapped onto a dimension-reduced space via LDA. It requires calculating the maximum degree of separation between different classes, the minimum difference between samples and the mean of each class, represented by S_B and S_W , respectively. The dimension-reduced space is generated by the transformation matrix (W_{max}) [17].

Mathematically, $W \in \mathbb{R}^{m \times j}$ ($j \ll k$) is a linear transformation matrix. It will be used for mapping HDD $Y \in \mathbb{R}^{m \times k}$ onto a dimension-reduced data $Z \in \mathbb{R}^{m \times j}$.

$$Z = YW^T \quad (2)$$

The transformation matrix (W) can be computed as in Equation. 3 [17].

$$W_{max} = \frac{W^T S_B W}{W^T S_W W} \quad (3)$$

Where S_B and S_W can be computed as in Equation. 1 and 2 [17].

$$S_B = \frac{1}{m} \sum_{j=1}^C m_j (\bar{x}_j - \bar{x}) (\bar{x}_j - \bar{x})^T \quad (4)$$

$$S_W = \frac{1}{m} \sum_{j=1}^C \sum_{k=1}^{m_j} (\bar{x}_j - \bar{x}) (\bar{x}_j - \bar{x})^T \quad (5)$$

Figure 2. shows two possible lower dimensional projected spaces of a group of data. Projection on LD1space satisfies the requirements of maximizing S_B and minimizing S_W .

During the past decades, numerous discriminant analysis techniques have been put out, each aimed at a particular data distribution type. However, in most cases, the distribution type is usually unclear, and researchers cannot guarantee the best algorithm is selected to analyze data. A method that aims to apply to a broad range of sample distributions was introduced by Zhu et al. [18] named Subclass Discriminant Analysis (SDA). Two criteria were derived to determine the practical partition of each class into a collection of subclasses. Additionally, this made it possible for them to identify the best subspace representation. Subsequently, a new scatter matrix of between-subclass can be defined.

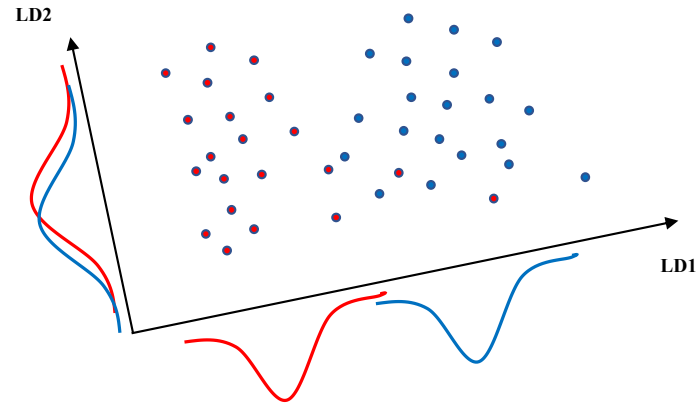


Figure 2. Analysis of Linear Discriminant.

Gkalelis et al. [19] proposed Mixture Subclass Discriminant Analysis (MSDA) years later. In their study, two shortcomings were pointed out when using SDA to analyze Gaussian homoscedastic subclass structured data. By modifying the objective function of SDA, MSDA showed higher classification performance. The within-class variance in LDA is typically singular since the number of dimensions is often significantly bigger than the number of observations. 2-Dimensional Linear Discriminant Analysis (2DLDA), a new LDA approach was proposed by Ye et al. [20]. The data represented type is the primary distinction between classical LDA and 2DLDA. The classical LDA uses a vectorized representation, while 2DLDA uses a matrix representation. Smaller-sized matrices are produced because of the eigen decomposition. Therefore, the computational time of 2DLDA is less required than classical LDA. Experimental results showed that 2DLDA performs better in classification tasks. According to Li et al. [21], though 2DLDA alleviates the singularity problem under some satisfying conditions, it might still run into the same problem. The test data generalizability is not considered either by LDA or 2DLDA, and the lack of a control term on the confidence interval might lead to an over-fitting problem. To address these problems, a novel generalized Lp-norm ($p > 0$) 2-dimension LDA (G2DLDA) framework was introduced by Li et al. [21]. Moreover, to achieve robustness, a proper “p” can be selected. In addition, the introduced regularization term can avoid singularity and over-fitting problems. Experimental results showed that G2DLDA performs better in terms of generalization.

Furthermore, Ran et al. [22] presented an efficient Generalization of Exponential Discriminant Analysis (GEDA), which replaced the Euler matrix exponential function with a general exponential function. Due to the property difference between these two functions, samples from distinct classes are separated at a greater distance by using GEDA. Sharma et al. [23] proposed the Improved Regularized LDA (IRLDA). It is a feature selection technique for gene study. The improvement lies in IRLDA is that the regularization parameter α can be determined without using any heuristic approach. The proposed method is further used to analyze gene expression datasets. Experimental results showed that IRLDA performs better in gene classification studies with limited samples. Robust Sparse LDA (RSLDA) was introduced by Wen et al. [24]. It is an effective feature extraction method that uses the L2,1-norm to restrict the projected matrix. To increase the resistance to noise and ensure the extracted features maintain the important information, both sparse norm matrix and orthogonal matrix are included. RSLDA showed great performance on image classification tasks.

2.3. Locally linear embedding

An unsupervised non-linear DRT was introduced by Roweis and Saul [25] named Locally Linear Embedding (LLE). LLE researches the local symmetries of linear reconstructions for discovering non-linear structures in HDD. Moreover, the local data’s structure is maintained in the embedding space. To put it simply, if the points in HDD input are close to each other, they ought to be adjacent neighbors in embedding space too. When performing the non-linear dimension reduction procedure, LLE can

successfully discover manifold structures by employing the eigen vector technique. The efficiency of LLE can be significantly improved if more local weight vectors are involved in calculating the optimal projected weight W_{ij} .

In terms of the algorithm [26], let X_i be one of the M real-valued vectors of data with D dimensionality, and each data point's K nearest neighbors can be identified by Euclidean distance or other methods. The cost function then measures reconstruction errors:

$$\varepsilon(W) = \sum_{i=1}^N \|X_i - \sum_{j=1}^K W_{ij \in A_i} X_{j \in A_i}\|^2 \quad (6)$$

As in Equation. 1 [26].

The cost function is minimized while keeping two limitations in mind. Firstly, for any data point X_i , if X_j is not belonging to the neighbor group of point X_i , then $W_{ij} = 0$. Secondly, the sum of the weight matrix's rows should be equal to one, that is $\sum_j W_{ij} = 1$. Then as expected, once the reconstruction weights W_{ij} are computed, they will be used to project every single vector X_i to a dimension-reduced vector Y_i . This process can be done by choosing coordinates Y_i in d -dimension to minimize the embedding cost function:

$$\delta(Y) = \sum_{i=1}^N \|Y_i - \sum_{j=1}^K W_{ij \in A_i} Y_{j \in A_i}\|^2 \quad (7)$$

As in Equation. 2 [26].

Figure 3. shows steps in locally linear embedding: (1) identify point X_i 's K nearest neighbors. (2) reconstruct linear relationship and obtain reconstruction weights W_{ij} (3) Compute low dimensional embedding vector Y_i .

Kouropteva et al. [27] proposed Supervised LLE (SLLE), which uses class labels to perform dimensionality reduction. And whether a data point belongs to another's K nearest neighbors depends on its membership information. Two variants of SLLE are proposed that only modify the method used to identify each point's K nearest neighbors. Experiments on handwritten digit recognition demonstrated that both variants yield the best accuracy. Hettiarachchi and Peters [28] proposed Multiple Manifold LLE (MMLLE). It is designed for learning multiple manifolds in data with various classes. The nearness of manifolds is used as a metric to explore the most suitable dimension-reduced embedding space. In multiple manifold spaces, MMLLE is trying to maintain the remoteness and structure of each manifold. Experimental results showed that MMLLE is performing well in object recognition and classification. LLE and its variants have also played a significant role in gene study. A method for the gene classification task was introduced by Lang et al. [29]. It uses a Signal-to-Noise Ratio (SNR) to remove noisy genes, followed by using LLE to project gene data into low dimensional space. The gene subset is constructed next by using SNR again. Experimental results showed the method could be applied to identify the difference between differentially expressed genes. Min et al. [30] proposed the Relevant Component LLE (RLLE). The core idea of RLLE is that it changes the represented data's feature space by assigning different weights to relevant and irrelevant dimensions. The ReliefF algorithm is being used to achieve this goal. After removing irrelevant attributes, the distance between the samples is computed. Then K nearest neighbor points of each sample are obtained, and the locally linear embedding is calculated. Experimental results on analyzing various DNA microarray data showed that the proposed method is highly effective in classifying different expressed gene classes. In recent years, gene selection and classification methods have been developed. A technique that combines LLE and Neighborhood Rough Set (NRS) was proposed by Sun et al. [31]. The irrelevant genes were dropped by measuring the separability between each class. The redundant genes were dropped by comparing the correlation coefficient. The noise was also removed by using wavelet analysis. In the tumor gene categorization study the proposed LLE_NRS showed great feasibility and effectiveness. Xu et al. [32] proposed a supervised LLE and Spearman's rank correlation coefficient (SLLE-SC2), which can be used for feature selection. In particular, the co-expression genes were removed by using the SC2 approach. Experimental results showed that SLLE-SC2 has great potential applicability to specific tumor problems.

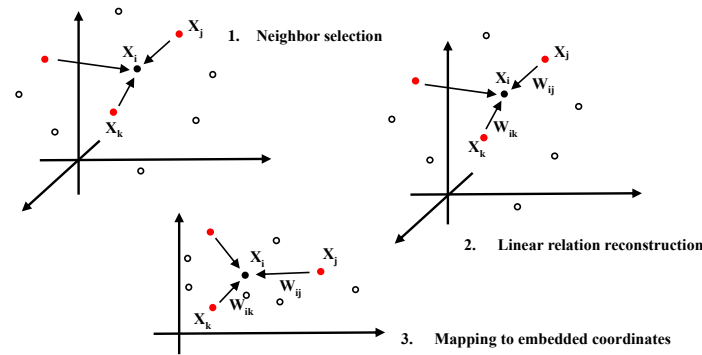


Figure 3. Analysis of Locally Linear Embedding.

3. Comparative study

3.1. Issues of HDD

Normally, outliers, noise, missing values, and small-sample-size are the most common issues of HDD. Respectively, outliers have much bigger or smaller values than all other values in the dataset. Outliers are outside of the normal distribution usually, so that is challenging to analyze a specific group of values when there are outliers. Similarly, it is hard to identify noise from actual data. Noise can either mask outliers or interfere with the detection of deviance. Data with missing values is hard to get accurate results after processing. Finally, a small-sample-size issue can cause an over-fitting problem. TABLE 2 shows a list of typical HDD issues with the DRTs that can be used to address each issue.

Table 2. Common HDD issues and corresponding available solution DRTs.

HDD Issues	Solution DRTs	References
Multi-variate	MFPCA	[12]
Outliers	ER-PCA, JSPCA	[8], [15]
Response Time	2D-PCA	[10]
Singularity	2DLDA, G2DLDA	[20], [21]
Small sample size	GEDA	[22]
Sparsity	SPCA, JSPCA	[14], [15]
Unknown subspace	GPCA	[9]

3.2. Issues of DRTs

Unfortunately, there is no guidance for selecting a certain appropriate DRT according to the type of HDD, and researchers may have to spend lots of time trying to find a suitable one. How to select a DRT correctly and effectively is a significant problem that requires more attention. Another issue is the identification of the redundancy level of redundant features and then removing some of them from the original HDD. Occasionally, some features may be essential for next-step analysis but are identified as redundant or irrelevant. Information is potentially lost when mapping low dimensional data. In addition,

many DRTs are sensitive to noise, and the existence of noise can have an immediate impact on DRTs' performance. Lastly, incorrect classification or predictions may result from extreme values of a certain variable.

Some common problems may arise while applying DRTs to real-life data. Identifying the nature of data and features for analysis always requires professional knowledge and relevant research experience. Hard to interpret the analysis results is one of the reasons hindering DRTs' application since it is not possible to keep all the original information after processing dimensionality reduction.

4. Discussion

DRTs have drawn greater attention in recent decades for use in gene expression research, face recognition, image classification, and other tasks. This study covers the most widely used methods, PCA, LDA, and LLE. The functionalities and available variants of each method are discussed. Due to its simplicity and effectiveness, PCA and its different variants are still widely used, but the existence of outliers might negatively impact PCA's performance. LDA and its different variants are widely used for pattern classification and feature extraction tasks on picture, text, and gene data. Moreover, G2DLDA resolves the classical LDA's singularity problem, while EDA and GEDA can solve the SSS problem. LLE and its different variants are widely used for image recognition and classification task. Gene expression studies also use LLE variants for selecting significant genes for a certain disease. Furthermore, LLE is only working for manifolds that have been well-sampled. It can retain the HDD's local structure in dimension-reduced space.

References

- [1] D. Amaratunga and J. Cabrera, "High-dimensional data, " *J. Natl. Sci. Found.* 44 (1), 3–9 (2016).
- [2] W. K. Vong, A. T. Hendrickson, D. J. Navarro and A. Perfors, "Do additional features help or hurt category learning? the curse of dimensionality in human learners, " *Cogn. Sci.* 43 (3), (2019) e12724.
- [3] I. K. Fodor, "A survey of dimension reduction techniques, " *Lawrence Livermore National Lab., CA (US)*, 2002 Tech. rep..
- [4] J. Shlens, A tutorial on principal component analysis, *Machine Learning*, arXiv:1404.1100v1 (2014).
- [5] S. Ahmadkhani and P. Adibi, "Face recognition using supervised probabilistic principal component analysis mixture model in dimensionality reduction without loss framework, " *IET Comput. Vision* 10 (3), 193–201 (2016).
- [6] C. O. S. Sorzano, J. Vargas and A. Pascual Montano, "A survey of dimensionality reduction techniques, " *Machine Learning*, arXiv:1403.2877v1 (2014).
- [7] M. Hubert, P. J. Rousseeuw and K. V. Branden, "Robpca: a new approach to robust principal component analysis, " *Technometrics.* 47 (1), 64–79 (2005).
- [8] S. Serneels and T. Verdonck, "Principal component analysis for data containing outliers and missing elements, *Comput. " Stat. Data Anal.* 52 (3), 1712–1727 (2008).
- [9] R. Vidal, Y. Ma and S. Sastry, "Generalized principal component analysis (gpca), " *IEEE Trans Pattern Anal Mach Intell.* 27 (12), 1945–1959 (2005).
- [10] T. Wang, I. Y. Gu and P. Shi, "Object Tracking Using Incremental 2D-pca Learning and MI Estimation, "in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE. 1, 1933–1936 (2007).
- [11] C. Z. Di, C. M. Crainiceanu, B. S. Caffo and N. M. Punjabi, "Multilevel functional principal component analysis, " *Ann. Appl. Stat.* 3 (1), 458–488 (2009).
- [12] C. Happ and S. Greven, "Multivariate functional principal component analysis for data observed on different (dimensional) domains, " *J. Am. Stat. Assoc.* 1–11 (2018).
- [13] T. Metsalu and J. Vilo, Clustvis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap, *Nucleic Acids Res.* 43 (W1), W566–W570 (2015).

- [14] M. Journée, Y. Nesterov, P. Richtárik and R. Sepulchre, "Generalized power method for sparse principal component analysis," *J. Mach. Learn. Res.* 11 (Feb), 517–553 (2010).
- [15] S. Yi, Z. Lai, Z. He, Y. Cheung and Y. Liu, "Joint sparse principal component analysis," *Pattern Recognit.* 61, 524–536 (2017).
- [16] A. Tharwat, T. Gaber, A. Ibrahim and A. E. Hassanien, "Linear discriminant analysis: a detailed tutorial," *AI Commun.* 30 (2), 169–190 (2017).
- [17] C. N. Li, Y. H. Shao, W. Yin and M. Z. Liu, "Robust and Sparse Linear Discriminant Analysis via an Alternating Direction Method of Multipliers," *IEEE Transactions on Neural Networks and Learning Systems.* 31 (3), 915–926 (2020).
- [18] M. Zhu and A. M. Martinez, "Subclass discriminant analysis, IEEE Trans," *Pattern Anal. Mach. Intell.* 28 (8), 1274–1286 (2006).
- [19] N. Gkalelis, V. Mezaris and I. Kompatsiaris, "Mixture subclass discriminant analysis", *IEEE Signal Process. Lett.* 18 (5), 319–322 (2011).
- [20] J. Ye, R. Janardan and Q. Li, "Two-dimensional Linear Discriminant Analysis," in *Advances in Neural Information Processing Systems 17*, edited by L. Saul, Y. Weiss and L. Bottou, (NIPS 2004), pp. 1569–1576.
- [21] C. N. Li, Y. H. Shao, W. J. Chen and N. Y. Deng, "Generalized two-dimensional linear discriminant analysis with regularization," *Neural Networks.* 142, 73–91 (2021).
- [22] R. Ran, B. Fang, X. Wu and S. Zhang, "A simple and effective generalization of exponential matrix discriminant analysis and its application to face recognition, IEICE Trans. *Inf. Syst.* 101 (1), 265–268 (2018).
- [23] A. Sharma, K. K. Paliwal, S. Imoto and S. Miyano, "A feature selection method using improved regularized linear discriminant analysis," *machine Vision and Applications.* 25, 775–786 (2014).
- [24] J. Wen, X. Z. fang, J. R. Cui, L. K. Fei, K. Yan, Y. Chen and Y. Xu, "Robust Sparse Linear Discriminant Analysis," *IEEE Transactions on Circuits and Systems for Video Technology,* 29 (2), 390–403 (2019).
- [25] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science, New Series.* 290 (5500), 2323–2326 (2000).
- [26] B. Ghogogh, A. Ghodsi, F. Karay and M. Crowley, "Locally Linear Embedding and its Variants: Tutorial and Survey," *Machine Learning*, arXiv:2011.10925v1 (2020).
- [27] O. Kouropteva, O. Okun and M. Pietikäinen, "Supervised Locally Linear Embedding Algorithm for Pattern Recognition." in *Iberian Conference on Pattern Recognition and Image Analysis*, Springer, 2003, pp. 386–394.
- [28] R. Hettiarachchi and J. F. Peters, "Multi-manifold lle learning in pattern recognition," *Pattern Recognition.* 48 (9), 2947–2960 (2015).
- [29] Y. X. Lang, Z. Qin and X. Li, "An Effective Gene Selection Method for Cancer Classification Based on Locally Linear Embedding," *Journal of Computational and Theoretical Nanoscience.* 8, 2108–2111 (2011).
- [30] X. P. Min, H. Wang, Z. W. Yang, S. X. Ge, J. Zhang and N. X. Shao, "Relevant Component Locally Linear Embedding Dimensionality Reduction for Gene Expression Data Analysis", *Metallurgical & Mining Industry.* 4, 186–194 (2015).
- [31] L. Sun, J. C. Xu, W. wang and Y. Yin, "Locally linear embedding and neighborhood rough set-based gene selection for gene expression data classification", *Genetics and Molecular Research.* 15 (3), (2016).
- [32] J. C. Xu, H. Y. Mu, Y. Wang and F. Z. Huang, "Feature Genes Selection Using Supervised Locally Linear Embedding and Correlation Coefficient for Microarray Classification," *Computational and Mathematical Methods in Medicine.* 2018.