

Performance Analysis of Dimensionality Reduction Techniques: A Comprehensive Review



Deepthi Mishra  and Saurabh Sharma 

Abstract Dimensionality reduction is comprised of techniques which are applied for lessening the dimensions of high-dimensional data. When there is increment in the data size and in its characteristics, then dimensionality reduction is applied to convert dataset into lesser dimensions with keeping the information short and precise. In other words, the process maintains the conciseness of the information without loss. The definition for dimension reduction can be stated as reducing n dimensional data for n dimensional space to i dimensional dimensions where $i < n$. To fulfill this objective, few techniques which are combination of mathematics and statistics are applied for dimensionality reduction are discussed in the paper. The paper focusses on the concept, techniques, and applications of dimensionality reduction. Additionally, the aim of the paper is to analyze and compare different techniques of dimensionality reduction with visualized results. Among different techniques of dimensionality reduction, results visualize LDA to be more informative and accurate.

Keywords Dimensionality reduction · Principle component analysis PCA · Linear discriminant analysis LDA · Independent component analysis ICA

1 Introduction

Usually, the enormous and high-throughput data is known as big data [1]. The biggest problem while dealing big data is to proficiently be processing it with elementary results. As statistical techniques are exceptionally good to handle low-dimensional data but many a times many statistical techniques are insufficient to deal with high-dimensional data. Due the drawbacks of increasing dimensionality in the dataset, reducing the number of features or dimensions has become a mandatory part which should be followed in such a way so that it can generate optimum solutions to the model.

D. Mishra (✉)

G. L. Bajaj Institute of Technology and Management, Greater Noida, India

S. Sharma

BITS PILANI, Pilani, India

So, to overcome the problems, some new statistical and computational techniques are required which can manage such huge data efficiently. By applying dimension reduction methods, the reduction of the original n dimensions to f dimensions takes place. In other words, it can be mentioned as the procedure of generating principle components by dropping few numbers of random variables fundamentally, it can be further split into two ways, one is feature selection and another is feature extraction.

Now to resolve the issues due to curse of dimensionality, innovative and new computational methodologies are to be required with statistical thinking. Usually, those statistical techniques which are applied for low-dimensional data necessarily not beneficial for high-dimensional data. So, this generates a requirement to develop some new effective techniques which are comprised of statistics and mathematics for data analysis and explorations of high-dimensional datasets. For dimensionality reduction, feature selection and feature extraction statistical techniques play a vital role to design accurate model for machine learning.

The benefit of dimension reduction methods is as all can deal with large amount of heterogenous data very efficiently such that n dimensions can be reduced to i dimensions. Generally, the techniques are applied after data gathering. Traditionally, dimensionality reduction is one of the methodologies of data reduction in data preprocessing.

Generally used methods for dimensionality reduction which are discussed in the paper are principal components analysis (PCA), linear discriminant analysis (LDA), and independent component analysis (ICA). Widely use of dimensionality reduction methods is achieved as it reduces the number of characteristics and noisy data by applying mathematical computations and transformations very efficiently [2, 3]. The key aspect of dimensionality reduction is to detect correlations between the features so as to improve accuracy in the result.

As for linear structure dataset, principal component analysis is computationally efficient [4].

The purpose of study presents different dimensionality reduction techniques in detail named as principal component analysis, linear discriminant analysis, and independent component analysis. Distinct techniques of dimensionality reduction are introduced thoroughly in the paper with algorithms and their functioning. The main objective of study is to furnish the knowledge of dimensionality reduction which can be further applied to different databases.

The contribution of the paper is to enlighten the researchers by exposing the study of dimensionality reduction with its various techniques. The goal of the paper is to clarify the requirement and applications of dimensionality reduction to researchers. The paper presents the visualization of different dimensionality reduction techniques in Python 3.7.

The rest of the paper is designed as, Sect. 2 includes the previous work done in dimensionality reduction. Section 3 consists of prerequisite knowledge content for dimensionality reduction. Need and applications of dimensionality reduction are discussed in Sect. 4. Different techniques of dimensionality reduction such as PCA, LDA, and ICA are analyzed in the next three sections. Section 8 shows implementation results. Paper ended with conclusion and references.

2 Literature Review

In current scenario, data is gathered very easily and in convenient way. Data can be collected for many purposes. It can be accumulated in an extraordinary rapid speed. It handles such data; data preprocessing plays an important role for effectively applying machine learning and data mining. To reduce and downsize the data, dimensionality reduction generates effective results.

Big data is the combination of various data streams within huge amount generated from heterogenous data sources. The distinguishing characteristic of big data is its gigantic volume which is incremented recurrently which causes curse of dimensionality [5]. So, it is necessary to apply dimensionality reduction for reducing the size of data while saving the beneficial data [6]. Big data is identified by its some characteristics such as [6]

Volume—It stands for vast amount of data that is big data size. There is no maximum limit of the data size for big data and no parameter and definitions in standings of volume for consideration of data as big data.

Veracity—It states about the noise, abnormalities, biases, and outliers in the huge data. When data arrives from heterogenous sources with wide variety of data measures, it can be prone to anomalies and deviations. So, there is requirement of efficacy of big data systems in respect with unfailling data sources.

Velocity—It refers the speed at which the big data generated. In other words, it is directly proportional to the frequency of collection of data entry. Computing methods then applied for processing of such data.

Variety big data is the collection and combination of heterogenous data from various sources. Data may be structured or unstructured.

Variability—Numerous sources generate various kind of data in abundance amount at different speeds which can generate the issue of inconsistency in data. Variability property of big data manages the related issues.

Value—It defines the worth of the data being gathered in big data repository from various sources. Its efficacy, effectiveness, and applicability are defined by the property.

The properties of big data in form of 6 V's make the big data system to manage the data processing and its calculations very efficiently and effectively in prospects of computations, business data processing, financial data processing, and many more. It helps in data compression and hence reduced storage space.

In statistics, the significance of dimensionality is the number of attributes in the dataset. High-dimensional data means the dataset having large number of attributes due to which computations become tremendously tough.

Feature selection and feature extraction are two methods which are applied for dimensionality reduction which improves the quality of data by removing irrelevant information. Few examples for high-dimensional data can be as digital images, spatial databases, speech signals, and real-world data which required dimensions to be reduced for data analyzing.

Some linear techniques for dimensionality reduction which are basic techniques such as principal components analysis (PCA), linear discriminant analysis (LDA), and independent component analysis (ICA) are discussed in the paper, though nonlinear data cannot be efficiently handled by these linear techniques.

There are some nonlinear techniques: Kernel PCA, Isomap, maximum variance unfolding, diffusion maps, locally linear embedding, Laplacian eigenmaps, Hessian LLE, local tangent space analysis, Sammon mapping, multilayer autoencoders, locally linear coordination, and manifold charting.

A symbolic regression technique is presented [7] for dimensionality reduction.

3 Prerequisite Knowledge for Dimensionality Reduction

3.1 Statistics

Statistics is a branch of mathematics dealing with data collection, organization, analysis, interpretation, and presentation [8]. It can be stated in steps as defining problem, analyzing problem, data collection and cleaning, analysis, and computation on data, and finally generating conclusion and interpretation of result [9].

Standard Deviation. It is the amount of distribution of a set of data from its mean [10]. It is the statistical method to assess the measure of dispersion, distribution, and variability of data as the value of standard deviation is high the higher the degree of deviation from its mean.

In statistics, the standard deviation (SD, also represented by the lower case Greek letter sigma σ or the Latin letter s).

Variance. In statistics, variance refers to the spread of a dataset. It is the method of measurement to calculate the square of the difference between the data points spread and the mean values of those data points which further divided by number of data points. Variance is a measure of the variability or spread in a set of data.

Covariance. It is applied to find correlation between variables such that it finds measurement of change between two variables. If the result is positive, then it is considered that two variables are positively correlated in same direction. If outcome is negative, it indicates that the variables are inversely related.

$$\sum (X_m - X)(Y_m - Y)/N \quad (1)$$

where

N total number of records.

X & Y mean values.

3.2 *Matrix*

The data can be represented in the form of matrix. The matrix is represented in the form of rows and columns. The size of a matrix is also called the matrix dimension or matrix order. If you want to add (or subtract) two matrices, their dimensions must be the same.

3.3 *Eigenvector and Eigenvalue*

Eigenvector is a nonzero vector that can be changed by a scalar factor.

$$K(v) = Tv \tag{2}$$

where t may be a scalar termed eigenvalue.

Certain vectors x is in the same direction as Ax . Those are the eigenvectors and A is considered as eigenvalue. It is basically applied in principal component analysis.

4 **Need and Application of Dimensionality Reduction**

Currently, data collection is tedious as the data accumulates in an unprecedented rate. This enormous data needs the preprocessing techniques of data mining. The goal of dimensionality reduction state is that low-dimensional data is to be generated while preserving information of high-dimensional data. As it is assumed that high-dimensional data is unintended dimensions of low-dimensional measured data, dimensionality reduction is an effective tactic to rationalizing huge data [11].

Data Compression: As the data is reduced dimensionally hence reducing storage space which in turn provides efficient retrieval of data from storage. Dimensionality reduction supports to analyze and retrieve data in faster for machine learning algorithms with removal of extraneous variables.

Noise Removal: The study says accuracy and efficiency of query processing is degraded swiftly as the dimension increases. So, it is an important part of data preprocessing, whereas data preprocessing is a significant role for machine learning and data mining. During the process, redundant features are removed which results in positive effect on query accuracy.

Visualization: It provides different projection of high-dimensional data onto 2D or 3D. When data is reduced to lower dimensions, visualization becomes easy.

Speed: Performance and efficiency can be increased by applying dimensionality reduction techniques in large set of dataset which further increases the speed of training data.

Analysis of parameters: Reduction of multicollinearity improves the analysis of parameters of models related to machine learning. If there is decrement in the size of the parameters directly, there is also reduction in sample size.

Evades curse of dimensionality: Reduction in noise and multicollinearity improves the conception process which in turn reduces curse of dimensionality.

There are various applications of dimensionality reduction such as document classification, image processing, biomedical, pattern recognition, and text categorization. There are various applications of dimensionality reduction such as document classification, image processing, biomedical, pattern recognition, text categorization [12–14].

5 Principal Component Analysis

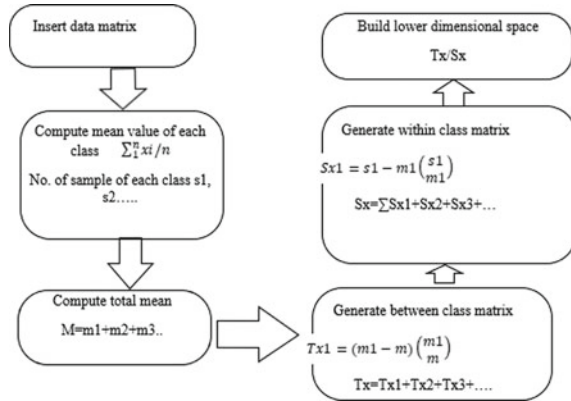
It is a methodology which can be applied to shrink the dimensions to lesser variables to ease the calculations to get accurate and precise results. The mathematics behind PCA can be stated as deriving principal components by finding largest variance with orthogonal values. First deriving a principal component by finding largest variance among the projection in the direction in feature space. Second principal component is orthogonal to first as maximizing the variance in all directions. Its concept can be reflected as rotating the data lies with original axes to new orthogonal axes.

It is applied in many ways, as it is one of the oldest techniques which is in turn a strong method for dimensionality reduction. PCA is nonparametric and unsupervised, so it can be applied various forms of analysis such as computer graphics, big data, and neurosciences. It can reduce high-dimensional data to lower-dimensional data with minimal effort.

PCA can be implemented by applying various stages of calculations. In short, it can be concluded that, initially dataset is identified, further mean values are calculated for each attribute values, then new dataset is constructed by subtracting mean from the data values. Covariance matrix is calculated from the dataset. Eigenvalues and eigenvectors are calculated from eigenmatrix. A feature vector is formed by choosing the components as eigenvector having the highest eigenvalues chosen as feature vector for PCA. As the feature vector is selected, take transpose of vector, and multiply it on the left of the original data, transposed. Previously, it can be performed on a square symmetric matrix.

While computing PCA, a new orthogonal axis is calculated so that axes of original coordinates of dimensions can be rotated and applied on the same. It is computed and applied in a way so that new axes can be accord with original one [15].

Fig. 1 Functioning of LDA



6 Linear Discriminant Analysis

This is another method for dimensionality reduction. In some characteristics, it is better than PCA. It has been discussed in later section in the paper. The objective of LDA is to preserve class separability information during dimensionality reduction. It is based on Fisher criterion [16]. LDA can do feature combination as it can transform all the original features to new ones.

LDA applies two different approaches, which can be named as class-dependent and class-independent transformation. It is based on supervised subspace learning further which is based on Fisher criterion.

Few steps required for implementing LDA are presented in (see Fig. 1) and can be stated as—Initially the dataset should be written in formulated form. In other words, it can be said that data can be represented in matrix form.

Mean values are computed for each class in the dataset. Further scalar matrix is computed additionally computing eigenvectors. The next step is to choose linear discriminants for new features from calculated eigenvectors by selecting top k eigenvectors, finally deriving new subspace from transformed samples.

7 Independent Component Analysis

ICA is statistical computation method to decompose multivariate data which can be in the form of signals, random variables, or may be some measurements [17]. Data applied and analyzed by ICA can be acquired by any application area such as digital images, document datasets or speech signals, waves, and many more. Let the data be combination of mixed variables of independent components.

$$x_1, x_2, x_3 \dots x_n.$$

ICA assumes all the components as separate and independent components in the mixture of variable [18]. It searches out those components which are independent and non-Gaussian. It is based on two basic principles, first is nonlinear decorrelations and second is maximum non-Gaussian condition. Those independent components are termed as factors or sources. It is also based on the facts that there should be independency in source signals. The values should be in non-Gaussian distribution in source signals. It discovers the independent signals by discovering statistical-independent components.

Generally, algorithms for ICA applies statistical techniques such as mean calculations, eigenvectors, and eigenvalues computation to detect independent components.

ICA can be divided into two cases: one is linear ICA and another nonlinear ICA. Further linear ICA can be defined in two different types such as noiseless and noisy.

Let the data can be defined in form of random vector $z = (z_1, \dots, z_n)K$ having hidden components $m = (m_1, \dots, m_n)K$. The observed data z will be transformed, by applying static transformation X such as $m = X_z$, turning vector of independent components m by applying function (m_1, \dots, m_n) .

8 Results

The results illustrate the implementation of few techniques of dimensionality reduction on iris dataset and wine dataset. Both datasets are retrieved from UCI machine learning repository [19]. Sample size of iris dataset is 150 records. Sample size of wine dataset is 178 records with 13 features. The dataset allows for several new combinations of attributes and attribute exclusions or the modification of the attribute type for the research and knowledge management process [20]. Seventy percent of dataset is applied as training dataset, and thirty percent of dataset is applied as test dataset. This is because more sample data is applied for classification to offer accurate results. The dataset should be applicable for classification.

The techniques are implemented in Python 3.7 on above-mentioned datasets. It is a prevalent open source nowadays which is a beneficial tool to deal with machine learning and statistical analysis. It provides perfect visualization and accurate outcomes.

Results of PCA, ICA, and LDA implemented in Python 3.7 on iris dataset are clearly visible in (see Figs. 2, 3, and 4).

9 Discussions

In this section, the discussion and comparison of different techniques of dimensionality reduction are presented as stated in (see Fig. 5). See Fig. 5 shows the accuracy results of PCA, LDA, and ICA. LDA arises as most prominent one by offering accuracy and better results than others. See Figs. 6 and 7 that represent the graphical

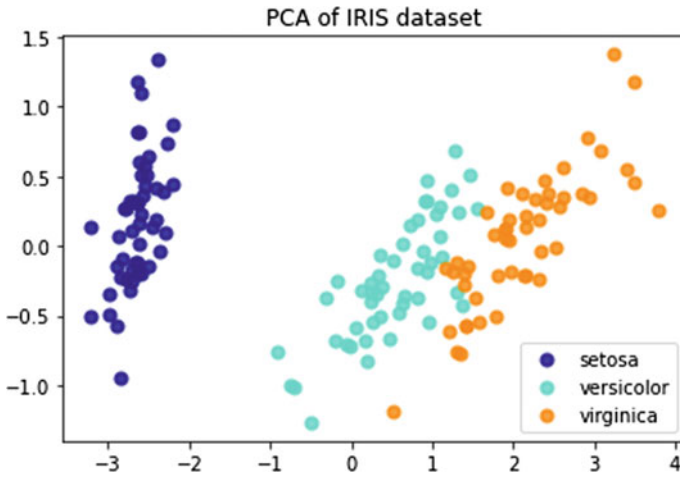


Fig. 2 The results of PCA with eigenvectors

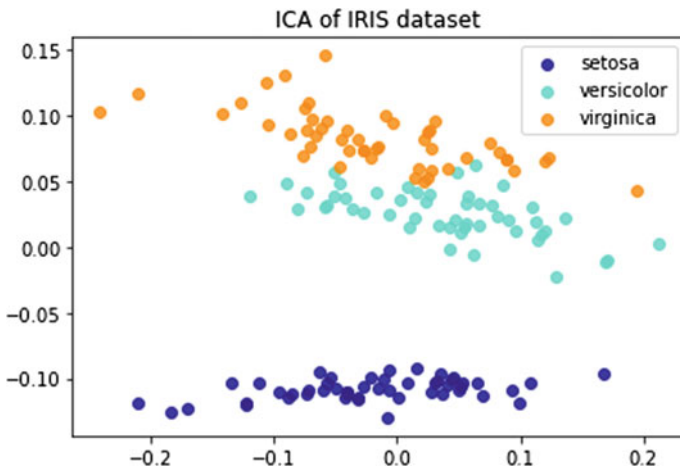


Fig. 3 The results of ICA with eigenvectors

representation of comparison of LDA, PCA, and ICA. It can be clearly interpreted from comparative results that LDA performs very well among the three techniques.

Implementation is done in Python 3.7. If class separability is considered then, it can be concluded that PCA achieves better results if samples per class are fewer in comparison to LDA which works better with dataset having large samples. Both are linear transformation techniques as PCA is considered and based on unsupervised techniques as it overlooks class specifications and LDA is considered as supervised. Further, PCA finds directions of maximal variance by applying the concept of eigenvectors usually called principal components while LDA finds feature subspace that

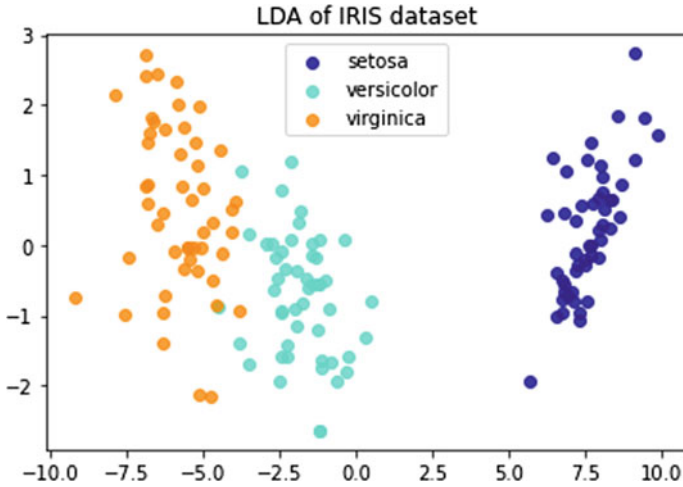


Fig. 4 Results for LDA with eigenvectors

Fig. 5 Comparative analysis of PCA, ICA, and LDA

	precision	recall	f1-score	support	
	0	1.00	1.00	1.00	15
	1	0.93	0.88	0.90	16
	2	0.87	0.93	0.90	14
accuracy				0.93	45
macro avg		0.93	0.93	0.93	45
weighted avg		0.93	0.93	0.93	45
The Accuracy for LDA is 0.9333333333333333					
[[15 0 0]					
[0 12 4]					
[0 1 13]]					
	precision	recall	f1-score	support	
	0	1.00	1.00	1.00	15
	1	0.92	0.75	0.83	16
	2	0.76	0.93	0.84	14
accuracy				0.89	45
macro avg		0.90	0.89	0.89	45
weighted avg		0.90	0.89	0.89	45
The Accuracy for PCA is 0.8888888888888888					
[[0 15 0]					
[0 16 0]					
[0 14 0]]					
	precision	recall	f1-score	support	
	0	0.00	0.00	0.00	15
	1	0.36	1.00	0.52	16
	2	0.00	0.00	0.00	14
accuracy				0.36	45
macro avg		0.12	0.33	0.17	45
weighted avg		0.13	0.36	0.19	45
The Accuracy for ICA is 0.3555555555555557]					

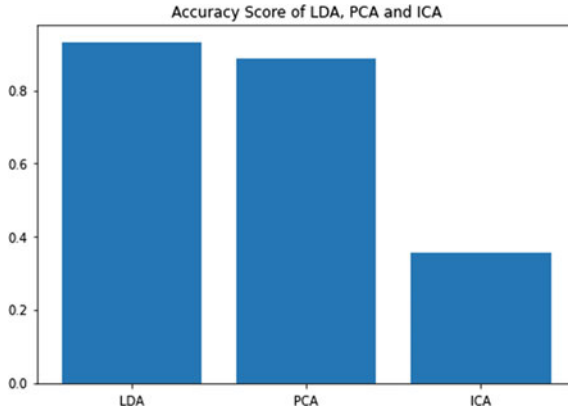


Fig. 6 Graphical representation of comparison of LDA, PCA, and ICA on IRIS

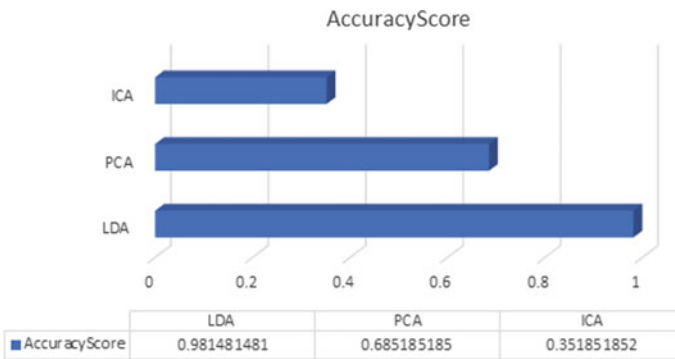


Fig. 7 Accuracy score on WINE database

maximizes the class separability in the dataset, which design LDA more generalized technique for dimensionality reduction. The study shows that PCA works on finding most variance in the dataset while LDA works on finding most variance for the individual class separability. Few drawbacks of PCA can be like as it where datasets are large enough to handle by only mean and covariance. Secondly, it is not necessary to generate linear correlation between variables.

Further, moving on the discussion, study finds that the PCA can remove correlations but not higher-order dependence but ICA can remove both correlations and high-order dependence. For ICA, all the components are equal unlike PCA which in turn considers only some components. PCA applies eigenvectors and its components are orthogonal, but components of ICA are not orthogonal. For selecting number of components, PCA applied variance criteria but ICA does not apply any such criteria for component selection. PCA searches directions that represents data in a $\Sigma \lambda_0 -$

x^2 sense, and ICA searches such components that are most independent from each other.

PCA and LDA are generally used but ICA is not usually applied in datasets. PCA and LDA are dimensionality reduction techniques meanwhile can be applied for classification of data.

10 Conclusion

Dimensionality reduction is the procedure to lessen m dimensions into d dimensions. The paper presents a study of dimensionality reduction techniques which can be further beneficial to researchers and academicians. This study provides worthy information about requirement of preliminary knowledge for dimensionality reduction and its techniques. The study reveals, LDA performs well on large samples in comparison with PCA. LDA provides more accurate results than PCA and ICA. LDA is more robust technique providing more precise outcomes. It is a beneficial process to reduce the redundancy. Those features which are correlated can be considered for reduction process. The study suggested which algorithm performs better according to data type as all the techniques have different way of calculations. User should ponder before applying the technique of dimensionality reduction according to the type of datasets.

References

1. Badaoui F, Amar A, Hassou LA, Zoglat A, Okou CG (2017) Dimensionality reduction and class prediction algorithm with application to microarray Big Data. *J Big Data* 32(4)
2. James AP, Dimitrijević S (2012) Ranked selection of nearest discriminating features. *Hum-Centr Comput Inform Sci* 2–12
3. Cunningham JP, Ghahramani Z (2015) Linear dimensionality reduction: survey, insights, and generalizations. *J Mach Learn Res* 16:2859–2900
4. Ye AQ, Ajilore OA, Conte G, Gad Elkarim J, Thomas-Ramos G, Zhan L, Shaolin Y, Kumar A, Magin RL, Forbes AG, Leow AD (2012) The intrinsic geometry of the human brain connectome. *Brain Inform* 197–210
5. Zhang Y, Zhou Z-H (2010) Multi-label dimensionality reduction via dependence maximization. *ACM*
6. Rehman MHU, Liew CS, Abbas A, Jayaraman PP, Wah TY, Khan SU (2016) Big data reduction methods: a survey. *Data Sci Eng* 265–284
7. Icke L, Rosenberg A (2010) Dimensionality reduction using symbolic regression. In: *GECCO'10*, ACM, Portland, Oregon, USA
8. James G, Witten D, Hastie T, Tibshirani R (2013) *An introduction to statistical learning*. Springer, New York
9. Field A, Miles J, Fields Z (2012) *Discovering statistics using R*. SAGE, Thousand Oaks
10. Trevor H, Robert T, Jerome F (2009) *Elements of statistical learning*. Springer, New York
11. Ramadevi GN, Usharani K (2013) Study on dimensionality reduction techniques and applications. *Publ Probl Appl Eng Res* 4(1):134–140

12. Sunita, Rana V (2018) An optimizing preprocessing algorithm for enhanced web content. In: Proceedings of SoCTA 2018, pp 63–71
13. Bhatia MK (2018) User authentication in big data. Proc SoCTA 2018:385–393
14. Goel R, Sharma A, Kapoor R (2018) State-of-the-art object recognition techniques: a comparative study. SoCTA Proc 2018:925–932
15. Delac K, Grgic M, Grgic S (2006) Independent comparative study of PCA, ICA, and LDA on the FERET data set. Int J Imaging Syst Technol 15:252–260
16. Gu Q, Li Zh, Han J (2011) Linear discriminant dimensionality reduction. In: Machine learning and knowledge discovery in databases. ECML PKDD 2011. Lecture Notes in Computer Science, vol 6911, Springer, Berlin, Heidelberg
17. Hyvarinen A, Karhunen J, Oja E (2001) Independent component analysis. Wiley, Hoboken
18. Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. Neural Netw 13(4–5):411–430
19. <https://archive.ics.uci.edu/ml/datasets/iris>
20. Fisher RA (1936) The use of multiple measurement in taxonomic problems. Ann Eugenics 7:179–188