

Recent Trends in Dimension Reduction Methods

Sehban Fazili¹, Jyotsna Grover², Samar Wazir³, Ila Mehta⁴
{sehbanfazili@gmail.com, Jyotsana.iitd@gmail.com, samar.wazir786@jamiahamdard.ac.in,
ilamehta88@gmail.com}

^{1,2,3}Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi

⁴ECE Department, NSIT, Delhi

Abstract: Dimensionality Reduction (DR) techniques helps us to focus on only that data which is essential or important to us. Basically, these techniques help us reduce a feature of a data element. DR is a process that reduces amount of variables into account, on obtaining the group of principle variables. DR is divided into two simpler methods that are feature elimination and feature extraction. The different methods used to reduce the dimensionality are: PCA (Principle component analysis), LDA (Linear discriminant analysis) and GDA (Generalized Discriminant analysis). Reduction of the dimensionality may be linear or nonlinear, depending on the method used.

Keywords: Principle component analysis, Linear discriminant analysis, Generalized Discriminant analysis.

1. Introduction:

Reduction of dimensionality, it's a technique that reduces a feature set's dimensions. A feature set can refer a database of hundreds of columns, or a collection of points that make up a large 3D sphere. DR takes down the quantity of columns to say thirty, or transforms the sphere into a 2D space circle. [7] The greater the number or quantity of features the more complex the model is.

Advantages of Dimensionality Reduction:

- Less disingenuous means improving model efficiency
- Less measurements means minimum computation and greater efficiency
- Fewer data indicates less storage space required
- Helps in reducing or removing redundant features and noise [3]

There are many reduction techniques, but the majority of methods fell under two specific classes: [5,8]

1. Feature Selection or Elimination
2. Feature Extraction

Elimination of feature is just what it indicates that is, by eliminating features, we minimize the feature space. [6]

In the extraction function, we build 10 “new” autonomous variables, in which every “new” experimental feature can be a blend of every of the 10 “old” autonomous features or variables. Nonetheless, we deliberately build these newly formed variables and arrange these unique variables according to whether they determine our dependent variables’. [5,8].

Linear Dimensionality Reduction methods:

- Principal Component Analysis (PCA)
- Factor Analysis
- Linear Discriminant Analysis (LDA)

Non-Linear Dimensionality Reduction methods:

- Multi-Dimensionality scaling (MD’s)
- Isometric feature mapping (isomap)
- Locally linear embedding (LLE)
- Hessian Eigen mapping (HLLE)
- Spectral embedding (Laplacian eigen maps)

The above description of the topic Dimensionality reduction has given us a basic idea of what dimensionality reduction is and what are its advantages and what methods it uses to reduce dimensions of a feature set.

2. Dimensionality Reduction (Literature Survey)

Dimensionality Reduction relates to the process of transforming a collection of large-scale data into smaller-scale data provided it clearly and accurately conveys specific information. [7] Usually these methods are used to solve machine learning problems in order to obtain better classification or regressive functionalities.

Dimensionality reduction methods are categorized in figure 1:

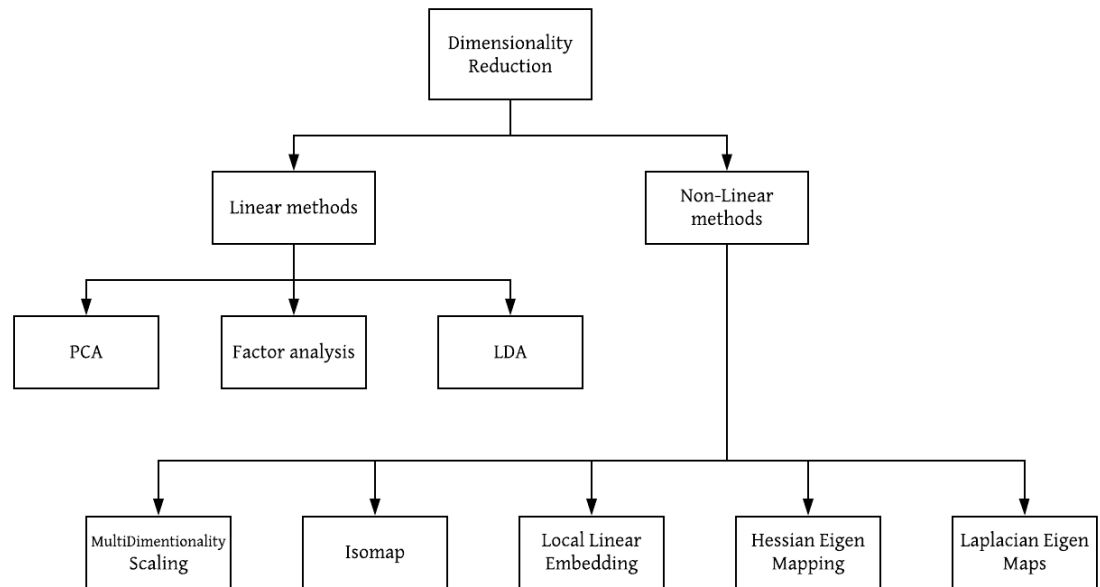


Figure 1 Dimensionality Reduction Methods

3. Literature Survey of linear methods for Dimensionality Reduction

Perhaps some of the most popular and well-known methods for reducing dimensionality are those that apply linear transformations. Within all methods of dimensional reduction, orthogonal projections are commonly used techniques. [1] Linear techniques or methods of transformation are used when the data are in linear subspace.

The most commonly used linear methods for dimensionality reduction are:

3.1 PCA (Principal Component Analysis)

This technique was originally formulated by Pearson in 1901. [1,2] It is a method which is used for reduction of dimensionality in continuous data. The principal components here are the features with maximum variance. This spins and projects the data along the path of growing or rising variance.

PCA may also be outlined as a mathematical method for transforming a series of observations possibly associated variables into a set of linearly unrelated variables values.

Every one of the main components is selected in a certain way that it could reflect the majority of the variance still accessible and all of these main components are orthogonal to one another. The first principal variable has the highest variance in all the main components.

Used in:

- Finding the interrelationship amongst variables throughout the data
- Translating and analysing data
- As number of variables reduces, further study is simplified
- It is also used to visualize genetic divergence and population-relatedness [3,10]

Principle Axis Method:

PCA simply searches for a linear variable blend and then extracts the maximum variance from the variables. Upon completion of this process, it completely eliminates it and searches for next linear combination that provides an explanation of the maximum percentage of the residual variance that essentially results to orthogonal factors. We evaluate total variance in this process.

- **Eigen vector:** A non-zero vector that remains parallel only after multiplication of matrixes. Let's assume that x is matrix M 's own r dimension vector with $r \times r$ dimension if both Mx and x are parallel. Then find the solution of $Mx = Ax$, where both x and A cannot find their own vector and their own values. [10]

Now we will state under Eigen-Vectors that the main components of the variable display both common and special variance. It is essentially a variance-focused method which seeks to produce complete variance and similarity with all elements. Primarily the linear combos of the initial variables are the main elements measured by their input to summarizing the variance in a particular orthogonal dimension.

- **Eigen Values: Generally** known as characteristic roots. Essentially, it analyses its variance throughout all variables that are taken into account for by the latter component. The ratio of values would be the ratio of the factor's informative significance to the variables. If the factor is small therefore the rationalization of variables is less contributory. In simpler terms, it determines the amount of variance that the factor accounts in the overall provided database. For all the variables, we can determine the value of the factor as the total of all its squared factor load.

Consequently, PCA is a tool that unites:

1. Covariance-matrix: A test of how every variable is equated.
2. Eigen-vectors: The ways in which our data were distributed.
3. Eigen-values: A relative value of those various directions.

PCA blends existing indicators and helps us to decrease relatively unimportant eigenvectors.[3]

3.2 Factor Analysis

This is a method used to decrease or minimize a huge number of variables into equivalent lesser ones. To determine the most significant, the values of the experimental data are represented as functions of a number of potential causes. The findings should be induced by a linear transformation of the latent factor in the lower dimensions and the inclusion of Gaussian noises. This method has become the one of the most commonly used statistical method, especially in psychology and behavioral sciences. [1,3]

It also can be defined as a technique for analyzing a complete matrix of all similarities between a range of different variables to uncover the latent sources of variance that might explain the similarities between many apparently different experiments.

It is one of the oldest structural models which was built in 1904 by Spearman. He attempted to explain the relationships between a set of test scores, and proposed that a model with a single common factor that he called "Intelligence" might produce these scores, with a specific factor for each and every test.

Analysis of factors is a method used to assess the similarities of variables or items within the pool of items for a growing measure.

3.3 Linear Discriminant Analysis (LDA)

This approach constructs data in such a way as to optimize the class separability. Definitions of the same type are closely linked by the projection. Definitions from different classes are not placed together by the projection. LDA is also known as normal discriminant analysis (NDA). LDA's aim is to project the data in such a way as to maximize the separation between classes. [1,2]

Additionally, LDA is closely linked to PCA and factor analysis in that both seek linear combinations of variables that help to understand the data. LDA seeks to mold the distinction between the classes of information. PCA, on the other hand, takes no class difference into account, and factor analysis constructs the function combinations expected to be based on disparities instead of similarities.

LDA operates where continuous quantities are the measurements made on independent variables for each test. The comparable approach is discriminant correlation analysis when dealing with categorical independent variables.

4. Literature Survey of Non-linear methods for Dimensionality Reduction

Non-linear methods of transformation are used when a linear subspace does not contain the data. It is founded on the manifold theory that, in a high dimensional system, the most relevant information is contained in a limited number of low dimensional manifolds. Suppose if a linear subspace is a sheet of aluminum, then a rolled-up foil of aluminum is a simple example of a nonlinear manifold. Informally, this is considered a Swiss roll, a traditional question in the area of the reduction of non-linear dimensionality. These methods are useful to analyze data with high signal to noise ratio, example images of physical objects.

Some non-linear methods are as following:

4.1 Multi-dimensionality Scaling

It is a way for projecting data onto a low-dimensional space, while approximately holding the space among every pair of samples within the original dataset. [5,7,9]

It is a class of methods and a large literature in its own right. [1] A method used to analyse data correlation or differences to geometric space distances. Projects of lower dimension details such that data points that are near to each other (in terms of Euclidean distance) [2] in the higher dimension are also similar to the lower dimension.

4.2 Isometric feature mapping (Isomap)

Projects information to such a lower dimension while retaining the geodesic distance [2,4] as opposed to Euclidean distance as in MD's. [9] The nearest range among two points on a curve is the geodesic distance.

4.3 Locally Linear Embedding (LLE)

Recovers non-linear global structure out of linear fits. [2] Given sufficient data, every local path of the manifold may be penned as a linear, measured sum of its closest neighbour.

4.4 Hessian Eigen mapping (HLE)

Projects lower-dimensional data while retaining the local neighbourhood like LLE, however utilizes the Hessian operator to obtain this result better and thus the name. [9] High dimensional data sets arise in many real-world problems. [2]

4.5 Laplacian Eigenmaps

This method uses spectral methodologies to reduce dimensionality by mapping adjacent inputs to adjacent outputs. This retains localization instead of linearity to the state.

This method is also known as Spectral Embedding. [4] Local properties are based on the pairly distance between neighbours in Laplacian eigen maps.

5. Conclusion

The paper presents a comprehensive survey of different dimensionality reduction methods. From the survey it is clear that Principal component analysis (PCA) that is Linear method is more preferred over other methods. PCA performs better where number of samples per class is less while as non-linear dimensionality reduction methods are useful to analyse data with high signal to noise ratio. For example, images of physical objects.

References

- [1]. Cunningham JP, Ghahramani Z. Linear dimensionality reduction: Survey, insights, and generalizations. *The Journal of Machine Learning Research*. 2015 Jan 1;16(1):2859-900.
- [2]. Sarveniazi A. An actual survey of dimensionality reduction. *American Journal of Computational Mathematics*. 2014 Mar 20;4(02):55.
- [3]. <https://www.youtube.com/watch?v=xebPVQ1f7nM>
- [4]. Van Der Maaten L, Postma E, Van den Herik J. Dimensionality reduction: a comparative. *J Mach Learn Res*. 2009 Oct 26;10(66-71):13.
- [5]. Padmaja DL, Vishnuvardhan B. Survey of dimensionality reduction and mining techniques on scientific data. *International Journal of Computer Science & Engineering Technology*. 2014;1(5):1062-6.
- [6]. Padmaja DL, Vishnuvardhan B. Comparative study of feature subset selection methods for dimensionality reduction on scientific data. In 2016 IEEE 6th International Conference on Advanced Computing (IACC) 2016 Feb 27 (pp. 31-34). IEEE.

- [7]. Nishana SS, Surendran S. Graph embedding and dimensionality reduction-a survey. International Journal of Computer Science & Engineering Technology (IJCSET). 2013;4(1):29-34.
- [8]. Matt Berms "A One-Stop Shop for Principal Component Analysis"
- [9]. Judy T Raj "A beginner's guide to dimensionality reduction in Machine Learning"
- [10]. <https://www.geeksforgeeks.org/ml-principal-component-analysispca/>