ISSN No: 2456-2165

Implementation of Dimensionality Reduction Techniques in Hospital Management

[1]S.Gnana Sophia, Research Scholar,
S.T. Hindu College, Nagercoil,
Affilated by:Manonmaniam Sundaranar University,
Abishekapatti,Tirunelveli-627012,TN,India

[2]Dr. K. K Thanammal, Associate Professor,
Department of Computer Science and Applications
S.T.Hindu College, Nagercoil
Affilated by: Manonmaniam Sundaranar
University,Abishekapatti,Tirunelveli-627012,TN,India

[3]Dr. S S Sujatha Associate Professor,
Department of Computer Science and Applications
S.T.Hindu College, Nagercoil
Affilated by: Manonmaniam Sundaranar
University,Abishekapatti,Tirunelveli-627012,TN,India

Abstract:- The major approach of machine Learning is Dimensionality Reduction. Overriding of the learning model will be produced if the higher number of features applicable in the dataset. The medical records in various hospitals are high dimensional in nature, which produce poor performance. The techniques of Dimensionality Reduction are applicable to resolve the features and the dimensionality of data sets are reduced. For the design of various dimensional hospital data. The Linear and non-linear methods of dimensionality Reduction techniques will be adapted and their ability is compared. Here the K-means Clustering algorithm is used.

Keywords:- *Machine Learning; Dimensionality Reduction (DR);* Principal Component Analysis(PCA *);*

I. INTRODUCTION

Many machine learning technique may not be adequate for higher dimensional data. Dimensionality reduction is needed for visualization, ie it projects the high dimensional data onto 2D or 3[1]. The valuable storage and regeneration are done in Data Compression and in Noise removal. Here, there will be the positive effect on the accuracy of query. Dimensionality Reduction is applicable in Customer Relationship Management, Text Mining, Image Retrieval , Face Recognition, Intrusion Detection etc.,. Medical data usually contains high amount of data which is hidden in the mode of images or signals [7]. There are three Dimensionality Reduction techniques: Linear Discriminant Analysis, Principal Component Analysis and Kernel Principal Component Analysis [3]. Many applications are using dimensionality reduction to diminish the dimensionality of data for the purpose of high processing speed. The data is diminished is called is known as features. Increasing processing speed is the advantage of dimensionality reduction.

The features can devotedly keep main information from original data. In feature extraction, it can explore the low-dimensional features to result in the high dimensional image. So it can lower the perception time and storage space [8]. An important application of dimensionality reduction is to build up the processing speed. There types of Linear Dimensionality Reduction are PCA, LDA, and Kernel PCA, ISOMAP LLE MDS and MVU are the types of Non-linear dimensionality reduction.

1. **Linear Discriminant Analysis (LDA)** It brings out the purpose to find the linear combination of Separating two or more classes of objects It is mainly used in pre-processing step for machine learning applications. This is an example tool for dimensionality reduction for linear method. It can target on maximizing the separability among known categories [7],

2. **Principal Component Analysis (PCA**) In dimensionality reduction of linear method, PCA is an important tool. Independent variables are numeric in nature. PCA is the commonly used unsupervised machine learning algorithm for the collection of applications 5]. They can be used in Dimensionality Reduction, compression etc.,The main aim of PCA is to obtain the principal components and it can characterize the data points with a set of principal components. These cannot chosen at random because the principal components are vectors [6]. The Principal Components can abolish the blast by reducing the large number of features to a particular group of Principal components and a feature vector and forming the principal components

➢ *Implementation of PCA on 2-D dataset.*

Normalize the data, Compute the covariance matrix, compute eigen values and eigenvectors, When 2D image is converted into i-D image, it can produce large vector space. Let us consider P as a N dimensional vector, F, the reconstructed image of size m x n will be resolved as Z = FP ----------------(1)

Using the above equation, we will get ND projected vector Z. The equation (1) will produce
J (P) = Trace (M)……… (2)

Where M is the covariance matrix.
M =meanvalue {(Z-meanvalue (Z)) (Z – meanvalue (Z)) T}
= meanvalue {[FP – meanvalue (FP)] [FP – meanvalue(FP)]T }
= meanvalue {[(F – meanvalue (F) P] [(F – meanvalue (F) P] T} (3)
Therefore, Trace (M) = P T {[meanvalue(F – meanvalue(F)] T (F – meanvalue(F)]}P……………. (4)

The excellent projection axes can be determined by maximizing the principle which exihibits to the highest eigen values. The projection axes F1, …,Fd are called eigen vectors of M of first 1 to 20 largest eigen values.

- *Advantages of PCA*

Because of it is established on linear algebra, the computation becomes easier and can be solved by computers. It speeds up the machine learning algorithms, its coverage becomes faster instead of original dataset when it can trained on Principal components. It aims to distinguish the correlation between variables [6]

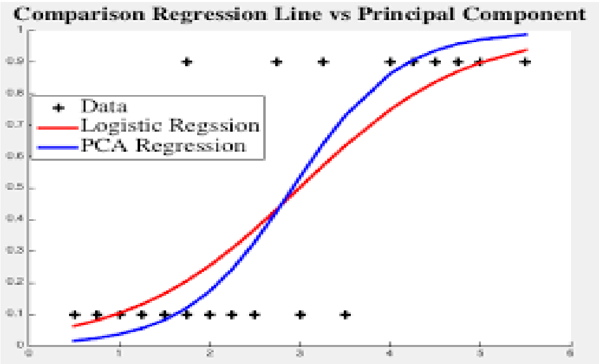

Figure 1 : Comparison of Regression Line Vs Principal Component Analysis

When we want to establish that variables in data are independent to each other PCA is needed. PCA is used.in a situation if we wish to diminish the number of variables in a data set with many variables in it and to clarify data and variables in a data set.

The standard PCA can find linear principal components to perform data in lower dimensions. Suppose if we want nonlinear principal component, and if we administer standard PCA for the data, it cannot produce good representative direction.. To avoid this problem Kernel Principal Component Analysis (KPCA) can be used.

3. Kernel Principal Component Analysis (KPCA): The Kernel Principal Component Analysis (KPCA) because it achieve PCA in a new space. It can develop the PCA to a high dimensional feature space using the “kernel Trick”

Linear: Something having to do with a line**.**
Non-linear: It is a function where the graph is a straight line

Because of it is a linear method, PCA cannot distinguish linear data adequately.KPCA can drop nonlinear organization in the data. This is the most universally used algorithm for dimension reduction. The PCA can calculate the co-variance matrix of m X m matrix X

$$C = \frac{1}{m} \sum_{i=1}^{m} \mathrm{Xi} X i^{T}$$

It will venture the data into the first k eigenvectors of the particular matrix. After being transposed into the high dimensional space, The KPCA can starts calculating the covariance matrix of the data

It is a nonlinear PCA and can be advanced by using kernel method. To find PC in various space it uses Kernel trick. It achieves PCA in a new space, and result in vast variance than PCA.

The mapping function of Non-linear method is denoted by Ø so X is the mapping of a sample and can be written as X -> Ø(X) this is known as “kernel function “. The function defines the name kernel which calculates the dot product of the images of the samples X under Ø

K (X_i, Y_j) = Ø (X_i) Ø (X_j) T)

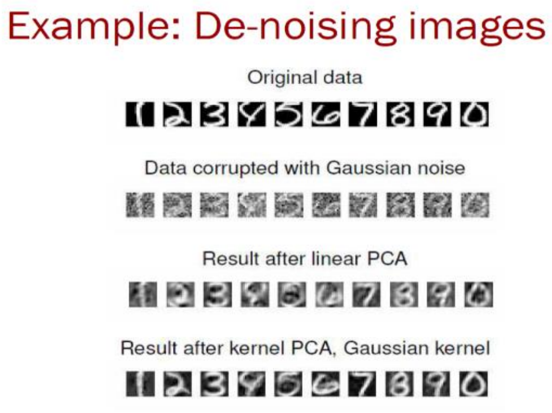
Example: De-noising images
Original data
Data corrupted with Gaussian noise
Result after linear PCA
Result after kernel PCA, Gaussian kernel

Figure 2: De-noising images

II. RELATED WORK

In the paper Research on KPCA and NS-LDA combined face recognition by Lei Zhao, Jiwen Dong and Xiuli Li, The authors proposed the method about KPCA plus NS-LDA for feature extraction and it can be applied for face recognition study [5]. It can be appreciated the performance of face recognition by the character of joining the applications of KPCA by the use of high order characteristic data and good feasibility of NS-LDA projection matrix. This method can forcefully reform the recognition rate [6]

III. PROBLEM FORMULATION

The linear PCA assumes that the conjunction between the variables are linear. If all the variables are assumed to be scaled in the digital level, then only their perception is logical. If we don’t select the number of principal components with care eventhough the principal components try to canvas maximum variance within the features in a dataset, it may loss some of the information when compared to the original list of features.

IV. PROPOSED SYSTEM

It uses kernel trick to discover principal components in high dimensional space. It directs new path which is based on covariance matrix of original variables. The computational complexity for KPCA to excerpt principal components and it takes more time. KPCA is a nonlinear version of Principal component Analysis which follows kernel tricks. The major concept is to activate the input vector x into the higher dimensional feature space through the kind of nonlinear mapping which can be selected by advance. It can give a good re-encoding of the data when it lies along a nonlinear manifold.

The way k-means algorithm works is as follows:

When we apply the original k-means clustering algorithm, we have to achieve the fumigate clusters. We have a set of data items with some properties. The aim is to classify that items into groups. This is an unsupervised algorithm. We can use the Euclidean distance for the computation of similarity. 11]. The steps for the algorithm is given as follows

Input X= {x1, x2,…..,xn}// set of n data items K // Number of clusters

Output: A set of k clusters

1. Define the total number of clusters *K.*
2. Initialize centroids by first dragging the dataset and then randomly choose *K* data points for the centroids without any restoration.
3. Carry iterating upto there is no difference in the centroids. i.e the framework of data points to clusters isn't changeable.

- Calculate the total of the squared distance between data points and all centroids. Select each data point to the nearest cluster (centroid).
- Calculate the average of all data points that belong to each cluster. It will compute the centroids for the clusters.

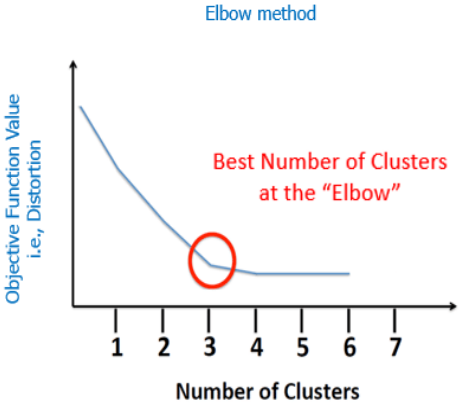

Figure 3: Centroids for Clusters

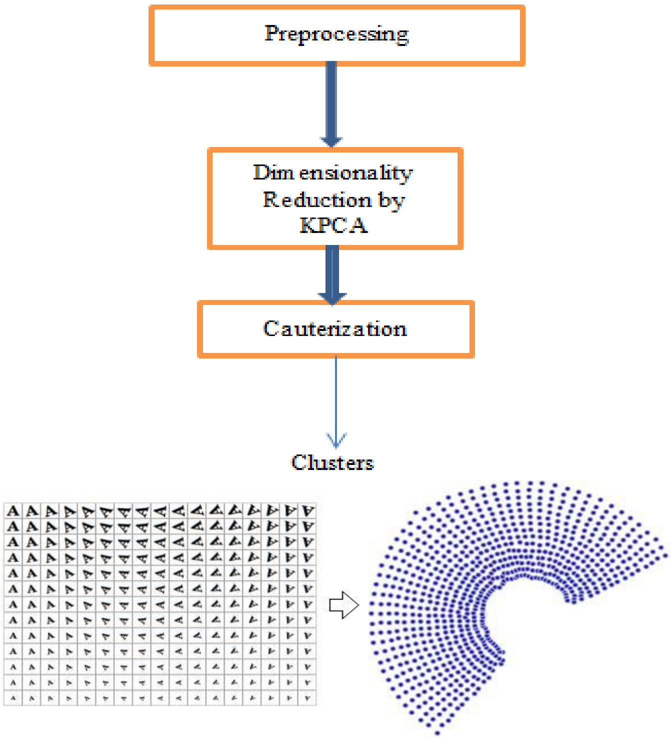

Figure 4: Proposed work for the problem

V. CONCLUSION

From the results determined we observe a course in the rates of algorithms. Since the datasets used have a linear data, we observe a perpendicular decline in the rates of nonlinear algorithms when compared to linear algorithms. It can specially pursuit to showcase the difference between the classes of data. It functions when the measurements made on independent variables for each observation are continuous quantities. The proposed algorithm works on unlabeled numerical data, Iterative technique, fast and efficient.

REFERENCES

[1]. Elkhadir, Zyad & Chougdali, Khalid & Benattou, Mohammed. (2016). Intrusion detection system using PCA and kernel PCA methods. 10.1007/978-3-319-30298-0_50.

[2]. S. Yin, Chen Jing, Jian Hou, O. Kaynak and H. Gao, "PCA and KPCA integrated Support Vector Machine for multi-fault classification," IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, 2016, pp. 7215-7220, doi: 10.1109/IECON.2016.7793188.

[3]. M. Fauvel, J. Chanussot and J. A. Benediktsson, "Kernel Principal Component Analysis for Feature Reduction in Hyperspectrale Images Analysis," Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006, Reykjavik, Iceland, 2006, pp. 238-241, doi: 10.1109/NORSIG.2006.275232.

[4]. Li Juan Cao, Kok Seng Chua and Lim Kian Guan, "Combining KPCA with support vector machine for time series forecasting," 2003 IEEE International Conference on Computational Intelligence for Financial

Engineering, 2003. Proceedings., Hong Kong, China, 2003, pp. 325-329, doi: 10.1109/CIFER.2003.1196278.

[5]. Xing Xiaoxue, Liu Fu, Shang Weiwei, Li Wenwen and Zhang Yu, "Research of PCA and KPCA in the characteristics simplicity of the gene data," Proceedings of 2013 2nd International Conference on Measurement, Information and Control, Harbin, 2013, pp. 669-672, doi: 10.1109/MIC.2013.6758051.

[6]. Jinghua Wang, Binglei Xie, Jiajie Xu and Haifen Chen, "A fast KPCA-based nonlinear feature extraction method," 2009 Asia-Pacific Conference on Computational Intelligence and Industrial Applications (PACIIA), Wuhan, 2009, pp. 232-235, doi: 10.1109/PACIIA.2009.5406645.

[7]. Comparative Study of Dimensionality Reduction Techniques for Spectral–Temporal Data Shingchern D. You 1,* and Ming-Jen Hung

[8]. Krzysztof SIWEK1 , Stanisław OSOWSKI1,2, Tomasz MARKIEWICZ1,3, Jacek KORYTKOWSKI1 Warsaw University of Technology (1), Military University of Technology (2), Military Institute of Medicine, Warsaw (3) Analysis of medical data using dimensionality reduction techniques

[9]. M. Karg, R. Jenke, W. Seiberl, K. Kühnlenz, A. Schwirtz and M. Buss, "A comparison of PCA, KPCA and LDA for feature extraction to recognize affect in gait kinematics," 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, Amsterdam, 2009, pp. 1-6, doi: 10.1109/ACII.2009.5349438.

[10]. L. Zhao, J. Dong and X. Li, "Research on KPCA and NS-LDA Combined Face Recognition," 2012 Fifth International Symposium on Computational Intelligence and Design, Hangzhou, 2012, pp. 140-143, doi: 10.1109/ISCID.2012.43.

[11]. Dr. B. Padmaja Rani2 1 Research scholar, Dept. of CSE, JNTUH, Hyderabad, Telangana 2 Professor of CSE, JNTUH, Hyderabad, “Kernel PCA Based Dimensionality Reduction Techniques for preprocessing of Telugu text documents for Cluster Analysis “