iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40572

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Dimensional Reduction Techniques for Huge Volume of Data

Soudagar Londhe[1], Manasi Patil[2]
[1, 2]*Dept. of Computer Engineering, Pune Institute of Computer Technology, Pune, India*

Abstract: ***Huge volume of data and information is needed with the expanding advancement in the current collection of tools, cloud storage, strategic techniques and increasing development of science technology. With the appearance of complete genome successions, the biomedical area has encountered an exceptional progression. This genomics has prompted the advancement of new high-produced strategies techniques that are huge amounts in measures of information and data, which inferred the exponential development of numerous organic and biological data sets. This paper represents different linear and non-linear dimensionality reduction techniques and their validity for different kinds of data information datasets and application regions.***
Keywords: High dimensional data, Dimensionality reduction, Linear techniques, Non-linear techniques, feature extraction, feature selection, Machine Learning

I. INTRODUCTION

In recent years, a tremendous large volume of data has been generated and used in various application areas. Also, the complexity, size, heterogeneity, and dimensionality of data information are growing rapidly. A huge amount of data is continuously and consistently generated in different formats like text, digital images, videos, and speech signals. Applications of High Data can be found in domains like social media, technology, medicine, web, and business.

High dimensionality data can result in accuracy, visualization, recognition, classification, and patterns and can cause overfitting. This issue can be avoided by adding subsequent data dimensions to each data point in exponential. With the Selection of features and extraction of features, i.e. feature transformation various dimensionality reduction can be implemented. By eliminating repetitive and unconnected features, extraction of feature transforms and changes initial datasets to the decreased dataset by conserving required information from the initial dataset. The selection of features collects the subsets from the data set that is the most relevant information data to the problem. Selecting the proper technique for dimensionality reduction can reduce the effort for feature analysis.

Reduction techniques offer a way to reduce input variables before applying them to machine learning models. It can be applied to pre-processing stage of data analysis and building models. Many reduction techniques are available with different data types but a particular technique may not be suitable for a particular application.

Paper is organized in a flow of sequence, where it describes different Dimensionality Reduction Techniques based on Linear and Non-Linear types of data set. Further datasets like Bioinformatics, Cancer Diagnosis and Prognosis and Character classification are tested for dimensionality reduction. Observation for before reduction and after reduction is computed.

II. DIMENSIONALITY REDUCTION TECHNIQUES

Dimensionality Reduction is a process of transforming and generating the high dimensional representation of datasets into low dimension representations. It transforms the original data set having higher dimensions and converts it into new data representing lower dimensionality while it preserves the original meanings of the data as much as possible. This data can be easily processed, analyzed, and visualized.

Formally, Dimension Reduction Techniques transforms the high dimensional data X = [x1, x2, …, xm] $\epsilon R^{m\times r}$ having the r dimensions and m observations into low dimensional data Z = [z1, z2, …, zm] $\epsilon R^{m\times k}$ where $k << r$ in an ideal case. Reduction techniques implicit, inverse and explicit map to construct it again from lower dimensionality data representation.

A. Linear Dimension Reduction Technique

Linear Dimension Reduction Techniques (LDRTs) use linear functions to transform higher dimensions into lower dimensions. The well-known technique is PCA. We have briefly discussed PCA below.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

1) *Principal Component Analysis:* PCA originally introduced by Karl Pearson, is an unsupervised linear mapping based on eigen vector. PCA preserves maximum amount of variance in original data by reduction using different strategies. It minimizes the dimensionality of feature space and conserves the maximum amount of variance. PCA is computed using different algorithms including factor analysis, eigen values, Linear Regression (LR) latent variable analysis.

Let X $\epsilon R^{m\times r}$ be an $m \times r$ matrix having m observations and r features. The PCs zi ϵR^{m}can be computed as a linear weighted combination of features.

Z = XW

Here, Z = [z1, z2, …, zm] ϵR^{m*r} and W = [w1, w2, …, wm] ϵR^{r*r} . This transformed data have the smallest reconstruction error with maximum variance among all projections.

PCA variants are designed for different data types, types and structures. The test results showed better LPCA performance compared to NN and PCA for speech and image data. upgraded for Robust PCA (RPCA) performance data to improve the robustness of the traditional algorithm. Hubert proposed Robust PCA (ROBPCA) based on reduction technique Projection Pursuit (PP) using robust scatter matrix data estimation. ROBPCA provided much more accurate and faster results compared to the previous PCA. To work with outsiders and non-data objects, Serneels and Verdonck introduced the Expectation Robust PCA and the test results show its suitability for the various sizes of data sets.

The algorithm is based on the search of orthogonal data directions which computes as much variance of the data as possible. The representation error of dimensionality reduction can be calculated by finding m orthonormal directions w_i for minimizing.

$J_{PCA} = E\{||x - \sum_{i=1}^{m} < w\_i, x> w\_i||^2\}$. From this objective function, projected of reduced vectors. $\chi = (<w\_i , x> ,\ldots, <w\_m , x>)^t$. This can be much more compactly written in $\chi = W^t x$, where W is a $M \times m$ matrix whose columns are the orthonormal directions w_i (or equivalently $W^tW = I$). The approximation to the original vectors is given by $x\hat{} = \sum\_(i = 1)\hat{}m < w\_i , x > w\_i$, or what is the same, $x\hat{} = W\chi$. Two dimensions graphical representation of a PCA transformation shown as ($x \epsilon R^2$). The rotated space shown by vectors $\chi = W^t x$ calculates the variance of the data in the original data space.

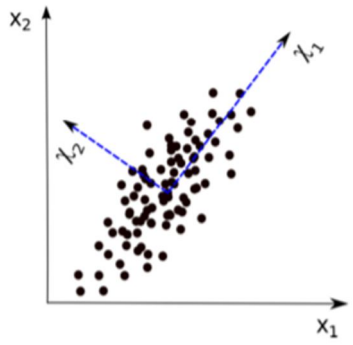


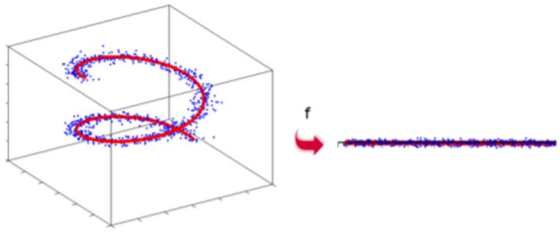
Fig 1. Graphical Representation of PCA transformation in two dimensions

We can rewrite the objective function as -

$J_{PCA} = E\{||x - W\chi||^2\} = E\{||x - WW\hat{}t\ x||^2\}\ \alpha\ ||X - WW^tX||\ \_F\hat{}2$.

Matrix projection of the input vectors to a lower-dimensional space ($\chi = W^t x$) is a technique in dimensionality reduction. Factors involved in this projection have a thrilling interpretation as explained below. Assume that we are analyzing scientific articles to a specific domain. Each article of word frequencies will be represented by a vector x, i.e., we choose a set of M words representative of our scientific area, and we determine how many times each word appears in each article. Each vector x is then orthogonally projected to the new subspace defined by vectors $\_i$. By considering each vector w_i as a dimension M, it can be directed as "topic" (topic is characterized M different words by their relative frequencies, two different topics can differ in relative frequencies of different M words). The projection of x onto each w_i relates an idea of how relatively important is topic w_i to represents the article x. More essential topics have massive projection values and, consequently, big values in the corresponding aspect of χ.

a) PCA for Curves and Surfaces: PCA helps to reduce data that are in *M* -dimensional space lie in some linear manifold. However, there are certain instances of data where it follows some curved type structure (slightly bent line). In this scenario, a straight line rounding the curve could not provide a good approximation to the actual data.



Original data space Transformed space

Fig 2. Representing Curved structure and transformed dataset

Given a bunch of predictions of the input vectors *x* with value zero average (assuming the original data is not zero- average, we can basically subtract the average from all elements), we can search for the line passing through the origin i.e. (0,0) and with direction w_1 (whose equation is $f(\chi) = w\_1 \chi$) that better fits this dataset, i.e., that minimizes $J\_(line) = E\{\lim_{0\to\infty} ||x - f(\chi)||\text{^}2 \}$. The infimum of the previous objective function implies that for every perception x_n we need to determine the point in the line (characterized by parameter χ_n) that is nearest to same point. The point $f(\chi\_n)$ is the orthogonal projection of the perception to the line. It can be demonstrated that the solution to this problem is the direction with the huge data variance, as similar arrangement as in PCA. A second new dataset can be constructed by a new dataset from the initial principal line by subtracting previously computed value $(x\_n\text{^}, = x\_n - f(\chi\_n))$.

2) SVD (Singular value Decomposition): SVD is a dimensionality reduction approach for matrix decomposition that is commonly employed when data is sparse. Rows of data are described as sparse when many of the values are 0. Some domains, such as recommender systems, have a user's rating for a small number of movies or music in the database and zero ratings for the rest. The bag of words model for text documents is another typical example, in which the document has a count or frequency for certain words but most words have a 0 value.

Examples of sparse data that can be used to reduce dimensionality using SVD:

a) Text Classification based on characters
b) Recommender Systems of platforms
c) Bag of Words Counts in passages
d) User-Song Listen Counts in music
e) User-Movie Ratings for reviews
f) One Hot Encoding
g) Customer-Product purchases for shopping

SVD factorizes the matrix $Y \epsilon R\text{^}(n \times k)$ into $UDV\text{^}T$. Matrix *U* and *V* are two orthogonal matrices whose dimensions are $n \times k$ and $k \times k$, respectively.

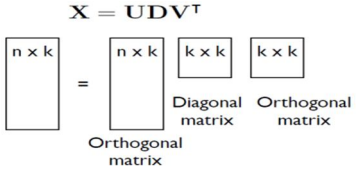


Fig 3. Matrix dimensions visualize

Matrix *D* is a $k \times k$ diagonal matrix containing singular values of the matrix *Y*. n is the number of singular values received. The ability of SVD to recreate the original matrix Y using the matrices U, V, and D is a key characteristic. SVD is computationally challenging, but with random needed sampling, sensitivity to outliers, and nonlinearities in the data, it can be made more efficient. It is difficult for visualization of the data through interpretation by the results as deciding factor.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

3) *Latent Semantic Analysis:* Latent Semantic Analysis (LSA), an unsupervised dimensionality-reduction technique that modifies text data in terms of r latent (hidden) features, where r is less than m (quantity of terms in data). With the growth of text documents, several challenges have arisen as a result of word phrase similarities in many disciplines and their linkage. A term can cause a conflict in both identification of data and the classification of data. For instance 'Fly' and 'fly', which has the same meanings but are differently used in place of noun and verb. LSA is a vector-based technique for comparing and representing text in low-dimensional data. LSA learns a semantic representation of text as well as associated word associations. LSA takes $n \times k$ words from the co-occurrence matrix $Y \epsilon R\hat{}(n \times k)$ as input. The local frequency of a given word I for a given document j is represented by each matrix element yij. The co-occurrence counts are then converted to definite weights in order to determine and predict information about a document's meaning. Previous dimension factors exist in the transformed matrix of weight values, which can be minimized using comparative SVD computation. For instance, LSA defines [$UDV\hat{}T$]=SVD(Y,m) where U and V are orthogonal data matrices.
4) *Locality Preserving Projections:* Locality Preserving Projections (LPP) is an unsupervised dimensional technique that structures and builds on neighborhood information of the data set. It signifies the idea and method of Laplacian of graph and calculates transformation matrix which maps the same data points from data sets to subspace. This representation map formed by can be thought of as a linear discrete approximation to a continuous forming map that emerges naturally from the scientific manifold's geometry.

LPP finds and computes $w1$, $w2$, … , wm vectors to map high dimensional dataset Y into low dimensional data by $w\hat{}T\ YDY\hat{}T\ and\ w$ should be equal to one.

$$argmin\_w\ =\ \Sigma\_ij \| w\hat{}T\ y\_i\ -\ w\hat{}T\ y\_j\ ||\_2\hat{}2\ A\_ij$$

A is connected and can be calculated using

$$A\_(ij)\ =\ exp(-(||\ y\_i\ -\ y\_j\ ||\hat{}2)/\beta)$$

Here β is originated to find eigen value decomposition data where all the average squared distance between all pairs of data can be reduced.

$$YLY\hat{}T\ w\ =\ \lambda YDY\hat{}T\ w$$

where $L\ =\ D\ -\ A$ and λ_i are eigen vectors.

Limitations of LPP is that it cannot be used for 2D image vectors due to the high singularity matrix issue. 2D saves local information and detects intrinsic manifold of image. It computed and resulted that 2D-LPP achieved a higher identification and recognition rate with better performance than LPP, 2D-LDA, and 2D-PCA when the identical data is applied for the same number of dimensions.

5) *Independent Component Analysis:* Independent Component Analysis (ICA) transforms a group of vectors into an independent set with maximally data points. Non-Gaussianity is used in ICA to identify independent components. The term "non-Gaussianity" refers to how far a random variable's distribution is from being Gaussian. Kurtosis and negentropy, for example. e.g. kurtosis and negentropy. ICA extracts independent components and objectives from the linear transformations of the original data.

Assume that the data observed $X\ =\ (x\_(1),\ x\_2\ ,\cdots,x\_m)\hat{}T$ is composed of using linear transformation of data with dimension $m \times p$ matrix T and non-Gaussian component vector $s\_i\ =\ (s\_(1),\ s\_2\ ,\cdots,s\_n)\hat{}T$.

$$X\_i\ =\ Ts\_i$$

ICA searches a linear mapping data W of the source vector s_i where each component data of an estimate v is as independent variance as possible.

$$v\_i\ =\ WX\_i\ =\ WTs\_i$$

6) *Projection Pursuit:* Projection Pursuit (PP) is an unsupervised dimension technique mainly used for the analysis of exploratory data. It is a non-parametric technique that finds lower-dimensional projections and explores and analyzes interesting patterns for data analysis. By identifying k dimensional data projection $X = [x\_(1), x\_2, \cdots, x\_k] \epsilon R\^(p \times k)\ (k < p)$ so that the projecting semantic data that maximize the predefined objective data function δ is called the projection pursuit index. It also measures the degree of riveting of projected data.

$argmax\_X\ \delta(YX)$ where $X\^T\ X = I$

The problem is divided by optimizations of relatively k problems. For each data optimization, one base of X is computed. The first base of data $x1$ is found by searching p units of dimensional data of unit length vector. Projected data then maximizes the one-dimension computed for the Projection index. To prevent the same projection data direction in successive following data iterations, projection aims to discard any information in the same direction from the source data. The process is repeated for all bases of data until all k subsequent bases are computed.

B. Non-Linear Dimensionality Reduction Techniques

A dimension reduction technique is associated and connected with a pair of data that consists of high-dimensional input data organized space to a low-dimensional corresponding output data space. A non-linear dimension reduction technique is used when a map's relationship is non-linear in nature. Following are some of the techniques.

1) *Multidimensional Scaling (MDS):* MDS computes and calculates the distance between data points from each pair in the original high-dimensional space data and maps it to a lower-dimensional space while it preserves the distances between points as much as possible. Multidimensional scaling works better when the value of the input distance matrix is combined into the elements of the d dimensional space region such that the relation of pairwise distances is preserved into embedded data space. By calculating sum of square errors between the non-similarities data and their corresponding data embedding inter-vector distances transformation of data can be achieved by calculating stress function.

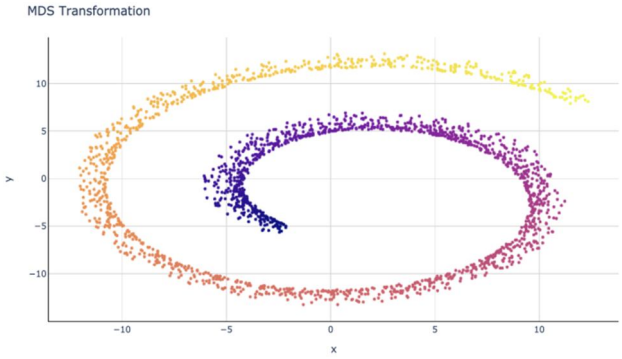


Fig 4. MDS Transformation

By computing and calculating distance between data points from the input data Y and store in distance matrix D. MDS is a technique for determining the low-dimensional coordinates of each data point in a dataset. The inter-point formation for relative distances $d\_ij\^$, of these data points should be always close to d_ij. MDS can be optimization problem and calculated as.

$$\delta(Z) = min(\sum_{i=1}^{m}\sum_{j=1}^{m}(d\_(ij) - d\_ij')\^2\ )$$

The stress function, which is formulated by sum-of-squares error, is used to determine adjacency. The most commonly used stress functions is calculated as:

$$Stress = \left(\frac{\sum_{i<j}(d\_(ij) - d\_ij')\^2)}{\left(\sum_{i<j}\left(d_{ij}^{i}\right)^2\right)}\right)$$

2) *Isomap:* Isomap is a technique that sets and combines different algorithms and uses a non-linear way to reduce dimensions of data points while preserving local default structures and geodesic distance. The distance between two elements in a data set that are in different manifolds is referred to as geodesic distance. Isomap computationally determines and calculates the geodesic distances between manipulated changed data points using a neighborhood graph theory. Each data point is connected with its k nearest neighbors similar points, the shortest distance between two points in data indicates a good estimate. It obtains a low-dimensional representation structure of data points by applying MDS on the resulting matrix. Isomap computes and identifies the nearby data points on the manifold M on the basis of the pairwise relation of Euclidean distance D_y. These neighbors data are stored in a weighted data matrix D_G that has the distances between neighbors. Isomap determines and computes the pairwise geodesic distance D_M between all pairs of data points on the same manifold M using all possible shortest path algorithms. The low dimensional data representation Z of high dimensional data X representation can be calculated using classical scaling formulated points on a pairwise geodesic matrix. The data vectors z_i are chosen accordingly so that the cost function of E is minimal.

$$E = || \sigma(D\_G - D\_z) ||\_F$$

where function σ transforms distances to inner products defined as

$$\sigma(D) = -HSH/2$$

Here, S and H are squared distances matrix and centering matrix, respectively.

3) *Locally Linear Embedding:* Locally Linear Embedding is an unsupervised dimensionality reduction method that tries to retain local data attributes. It computes and reduces n-Dimensions and parallelly it preserves the geometric data points features of the original previous non-linear feature structure. Suppose, if we have D-dimensions for data X_1, then we try to to reduce X_1 data to X_2 data in a feature space with d-dimensions.

$$X\_1 \in D \rightarrow X\_2 \in d$$

LLE predicts and identifies the neighbors of each data point set y_i using different algorithms. It calculates and identifies the weight data matrix W of data Y from neighbors' subsequent points. It computes to minimize the cost function.

$$\sum_{i=1}^{m} || y\_i - \sum\_(j = 1)\^m w\_ij y\_j ||\^2$$

Then low dimensional calculations are made for vector $〚 z 〛\_i$ data so that the cost function is minimized.

$$\sum_{i=1}^{m} || z\_i - \sum\_(j = 1)\^m w\_ij z\_j ||\^2$$

4) *Self-Organizing Map:* Self-Organizing Map (SOM) unsupervised Non-Linear dimensionality reduction techniques based on cognitive learning for analyzing nonlinear projections, complex data sets and multivariate. Data having mixed type features, the dissimilarity of two objects can be calculated on numerical and category-based features separately. The dissimilarity of data points can be calculated using the combination of squared Euclidean distance of data center points on numeric features and the calculated number of mismatches on categorical data features.

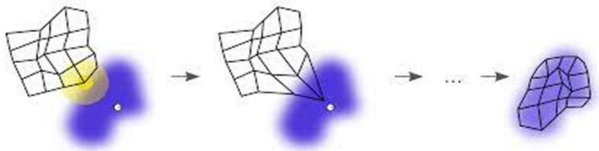


Fig 5. Self-Organizing map

A Self Organizing Map consists of a pairwise map of neurons $n\_1, n\_2, \cdots, n\_j$, each relating with a weight vector $w\_1, w\_2, \cdots, w\_j$, on a p-dimensional map. SOM analyses the distance to the weights of each neuron data corresponding for each y_i. The neuron having the lowest distance to its weight data is chosen for computing and plotted in direction analysis of corresponding y_i.

$$\delta\^, = min (|| y\_i - w\_k ||\^2) \text{ where } 1 \leq k \leq j$$

SOM can be used in real-life applications such as intrusion detection, recognition of protein folds, automatic organization of a massive collection of documents, market data analysis, development of risk-based prioritization for stagnation, the classification of fMRI, discard of the noise from 6D synthetic spectral image view data, and predication speed for weather climate and crop production data.

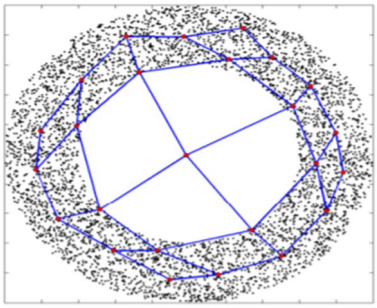


Fig 6.Representing data in form of a ring, Output map topology is represented by linking each vector to its neighboring blue edge

5) *Learning Vector Quantization (LVQ):* Learning Vector Quantization is a supervised algorithm used for different pattern recognition and statistical data classification. In LVQ, the competitive layer analyses and computes the input vectors later classifying them into correct classes using distance calculations. An input data layer, a single LVQ layer, and an output layer constitute LVQ. The output layer consists of nodes that are equal to distinct classes. LVQ specifies the classification using data of C classes prototypes. These are chosen as representative of the classification classes. With the help of space s_i and their respective class labels $c(s\_i) \epsilon 1, 2, \cdots, C$, they are distinguished by their location in dimension feature. For a given distance measure $d\hat{}, (s, y)$, the data point is assigned to a data class $c(s\_i)$ with $d\hat{}, (s\_i, y) \leq d\hat{}, (s\_j, y)$ for all $j \neq i$. Distance measured from computing can also be specified using a data matrix. LVQ cannot work well with the complex specific data sets having one or many pairwise relationships.
6) *T-Stochastic Neighbor Embedding:* T-Stochastic neighbor embedding (t-SNE) is an unsupervised algorithm for dimensionality reduction where it focuses on storing similar data points close with proper rearrangement and embedded the sets together in lower-dimensional space. It stores the local data structure using t-distribution to calculate and compute the similarities between two data set points. For data transformation of lower-dimensional, t-SNE mostly depends on conditional probability.

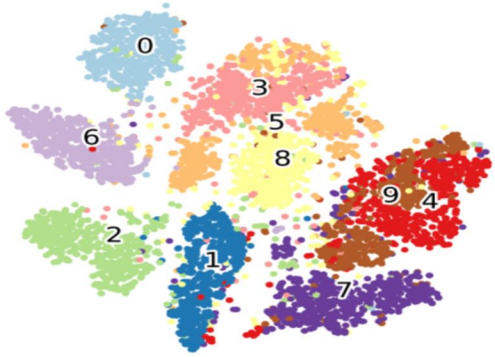


Fig 7. Large Features of data points represented in 2D space

For input data Y, SNE computes the conditional data probabilities p_ij. The goal and aim of t-SNE is to have k dimensionality map that reflects the similarities q_ij between two data points z_i and z_j.

$$P_{j|i} = \frac{(exp(-||y_i - y_j ||\hat{}2) / 2\sigma\_i\hat{}2)}{(\sum\_(k \neq i)\ exp(-||y\_i - y\_k ||\hat{}2) / 2\sigma\_i\hat{}2)}$$

$$q_{ij} = \frac{((1 + ||z\_i - z\_j ||\hat{}2)\hat{}(-1))}{(\sum\_(k \neq i)(1 + ||z\_i - z\_k ||\hat{}2)\hat{}(-1))}$$

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

t-SNE a distribution is the embedding space.

$$E\_(t-SNE) \sum_i \sum_j p\_ij \ log \frac{p\_ij}{q\_ij}$$

where,

$$P_{ij} = \frac{(p\_(j|i) + p\_(i|j))}{2m}$$

III. EXPERIMENTAL RESULTS

Generally, large data of real-world data are non-linear in nature. Spam data, three-dimensional data like Insurance Benchmark and cancer datasets are chosen for analysis. Here cancer dataset abstracts around 57 attributes and 26 instances, Spam contains around 4600 records for 57 attributes and Insurance dataset abstracts around 85 attributes with 750 total records for analysis. For all these high dimensional data computing and analyzing is very difficult and all these variables does not affect classification result. So, by removing irrelevant data we can reduce dimensionality data.

By removing attributes from the high dimensionality data takes less execution time, the accuracy may affect the loss of data for classification. For this project, dimensionality reduction is performed and performance is measured for Elapsed time and accuracy with the help of matlab.

A. Result of SVM Classification

High dimensional data has been used on SVM classification in this paper. Elapsed Time and Accuracy for value classification are noted and results are displayed using the table.

Dataset	Accuracy	Elapsed Time (in sec)
Bioinformatics	40.4667	1.6658
Cancer Diagnosis and Prognosis	77.9231	1.2544
Character classification	36.5217	0.3643

Table 1. Result of SVM Classification

To Show effectiveness Linear Discriminant Analysis is mainly processed for the dimensionality reduction. From analysis 86, 58 and 60 variable data are computed and then transformed into 32, 39 and 33 on bioinformatics, cancer diagnosis and character classification datasets respectively. In the next step, with the help of Principal Component Analysis thessse high dimensional data sets are processed which produces better data results than the LDA reduction method by giving 25, 8 and 14 respectively. Then ICA technique is performed on the high dimensional data which produces the results data 16, 8, and 5 bioinformatics, cancer diagnosis and character classification datasets respectively which is better compared to PCA and LDA dimensionality reduction techniques and the dimensions of low dimensional data are shown in the table.

Dataset	No of variables in HDD	Linear components	Principal Components	Independent Components
Bioinformatics	86	32	16	16
Cancer Diagnosis and Prognosis	58	39	16	8
Character classification	60	33	4	5

Table 2. Dimensions of low dimensional data

These lower-dimensional data from the above techniques are then again processed on the SVM classification and the corresponding Performance of the SVM classification is computed and calculated. From calculations, it proves and explains that for lower-dimensional data the performance of SVM is better and best computed than that of the higher-dimensional for SVM and the result is shown in the table.

Datasets	PCA		
	Dimension	Accuracy	Time
Bioinformatics (86)	16	49.76	1.22s
Cancer Diagnosis and Prognosis (58)	16	60.43	0.08s
Character classification (60)	4	71.09	1.15s

Datasets	LDA		
	Dimension	Accuracy	Time
Bioinformatics (86)	25	53	1.53s
Cancer Diagnosis and Prognosis (58)	8	39.46	0.01s
Character classification (60)	14	74.82	4.05s

Datasets	ICA		
	Dimension	Accuracy	Time
Bioinformatics (86)	16	59.23	1.64
Cancer Diagnosis and Prognosis (58)	8	70.23	0.008
Character classification (60)	5	73.13	1.22

Table 3. Performance Analysis

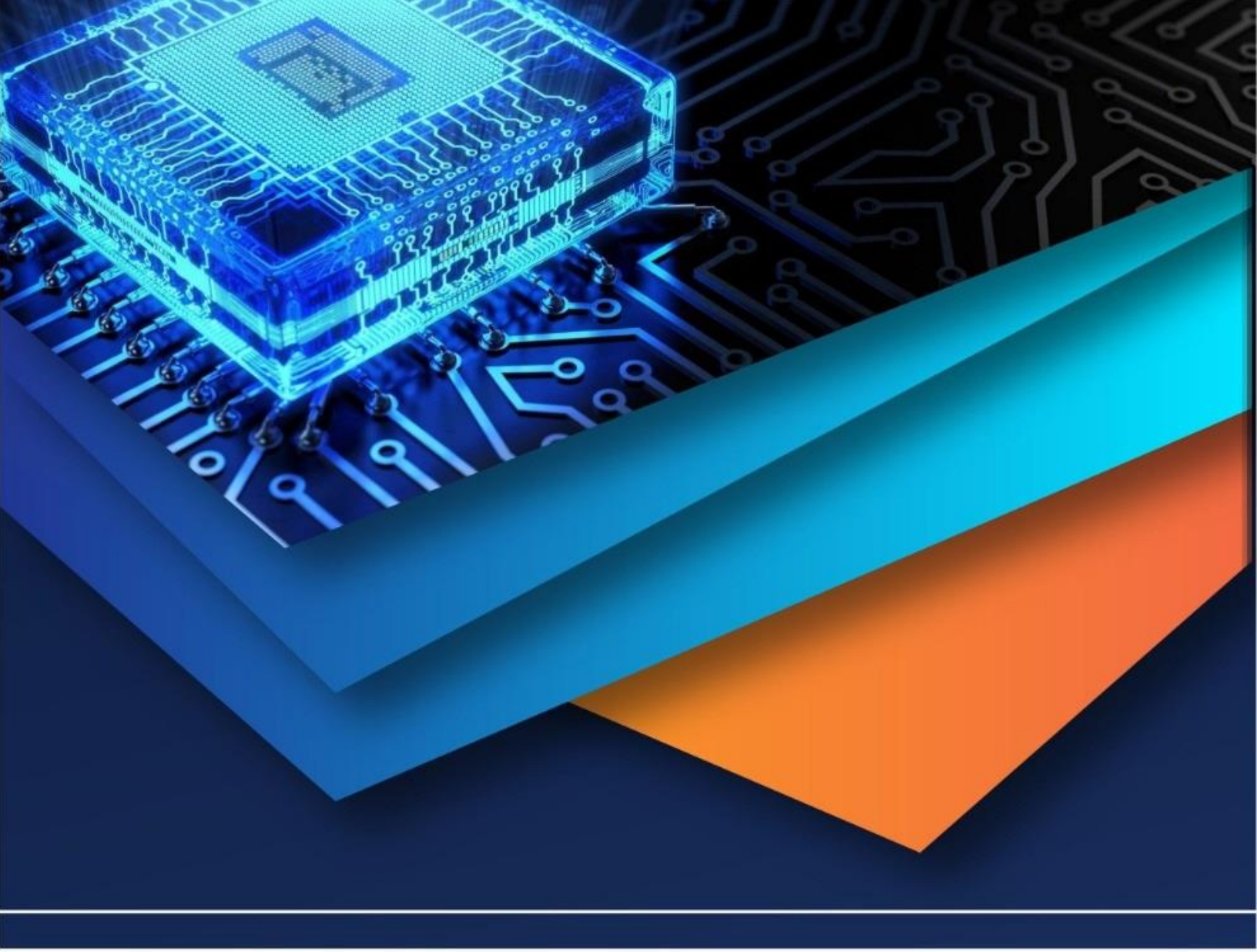
IV. CONCLUSION

In this paper, we present the different techniques to reduce the dimensionality of the original data points. As more and more data is generated, need for dimensionality reduction techniques also increased to reduce uncertainty in the decision-making of the data. For less computation power linear techniques are used which uses linear transformation. Time and cost are high for Non-linear techniques and have been correctly implemented in many different complex computations like audio, video and biomedical data.
From the survey of datasets from above results, it comes to know that, for handling data of linear dimensional Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are the best techniques. For handling non-linear dimensionality reduction Support Vector Machine (SVM), Independent Component Analysis (ICA) and Multi-Dimensional Scaling (MDS) are the best reduction techniques.

REFERENCES

[1] X.L. Zhang, Nonlinear Dimensionality Reduction of Data by Deep Distributed Random Samplings, in Asian Conference on Machine Learning, volume 2015, pp. 221–233.
[2] Zizhu Fan; Yong Xu; Zhang, D., "Local Linear Discriminant Analysis Framework Using Sample Neighbors", IEEE Transactions on Neural Networks, On page(s): 1119 - 1132 Volume: 22, July 2011
[3] Samarasena Buchala, Neil Davey, Tim M. Gale & Ray J Frank (2005) Analysis of linear and nonlinear dimensionality reduction methods for gender classification of face images, International Journal of Systems Science, 36:14, 931-942
[4] N. Kwak , "Principal component analysis based on L1-norm maximization", IEEE Transaction on Pattern Analysis and Machine Intelligence, volume. 30, no. 9, pages.1672 -1680 Year : 2008
[5] Chen Meng, Oana A. Zeleznik, Gerhard G. Thallinger, Bernhard Kuster, Amin M. Gholami, Aedín C. Culhane, Briefings in Bioinformatics, Volume 17, Issue 4, July 2016, Pages 628–641
[6] X. He, Incremental Semi-supervised Subspace Learning for Image Retrieval, in: Proceedings of the 12th Annual ACM International Conference on Multimedia, ACM, 2004, pp. 2–8.
[7] P. Huang, T. Li, Z. Shu, G. Gao, G. Yang, C. Qian, Locality-regularized linear regression discriminant analysis for feature extraction, Inf. Sci. (NY) 429 (2018) 164–176.
[8] S. Chen, H. Zhao, M. Kong, B. Luo, 2D-lpp: a two-dimensional extension of locality preserving projections, Neurocomputing 70 (4–6) (2007) 912–921.
[9] A. Hyvarinen, Fast and fixed-point algorithms for independent component analysis, IEEE Trans. Neural Netw. 10 (3) (1999) 626–634.
[10] T. Radüntz, J. Scouten, O. Hochmuth, B. Meffert, Automated eeg artifact elimination by applying machine learning algorithms to ica-based features, J. Neural Eng.14 (4) (2017) 46004.
[11] D. Zhou, X. Yang, Face Recognition Using Improved-lda, in: International Conference Image Analysis and Recognition, Springer, 2004, pp. 692–699.56
[12] S. Deegalla, H. Boström, K. Walgama, Choice of Dimensionality Reduction Methods for Feature and Classifier Fusion with Nearest Neighbor Classifiers, in 15th International Conference on Information Fusion (FUSION), IEEE, 2012, pp. 875–881.
[13] R. Vidal, Y. Ma, S. Sastry, Generalized principal component analysis (gpca), IEEE Trans Pattern Anal Mach Intell 27 (12) (2005) 1945–1959.
[14] G.H. Golub, C. Reinsch, Singular Value Decomposition and Least Squares Solutions, in: Linear Algebra, Springer, 1971, pp. 134–151.
[15] L. Cao, Singular Value Decomposition Applied to Digital Image Processing, in: Division of Computing Studies, Arizona State University Polytechnic Campus, Mesa, Arizona State University Polytechnic Campus, 2006, pp. 1–15.
[16] Y. Wang, L. Zhu, Research and Implementation of Svd in Machine Learning, in: IEEE/ACIS 16th International Conference on Computer and Information Science(ICIS), IEEE, 2017, pp. 471–475.
[17] F. Husson, J. Josse, B. Narasimhan, G. Robin, Imputation of mixed data with multilevel singular value decomposition, J. Comput. Graphical Stat. (2019) 1–26.
[18] P. Wiemer-Hastings, K. Wiemer-Hastings, A. Graesser, Latent Semantic Analysis, in: Proceedings of the 16th International Joint Conference on Artificial Intelligence, 2004, pp. 1–14.
[19] H.T. Tu, T.T. Phan, K.P. Nguyen, An Adaptive Latent Semantic Analysis for Text Mining, in: International Conference on System Science and Engineering (ICSSE), IEEE, 2017, pp. 588–593.
[20] M. Kedadouche, Z. Liu, M. Thomas, Bearing Fault Feature Extraction Using Autoregressive Coefficients, Linear Discriminant Analysis and Support Vector Machine under Variable Operating Conditions, in: Advances in Condition Monitoring of Machinery in Non-Stationary Operations, Springer, 2018, pp. 339–352.
[21] G. Baudat, F. Anouar, Generalized discriminant analysis using a kernel approach, Neural Comput. 12 (10) (2000) 2385–2404.
[22] J.A. Lee, A. Lendasse, M. Verleysen, Nonlinear projection with curvilinear distances: isomap versus curvilinear distance analysis, Neurocomputing 57 (2004) 49–76.
[23] J.H. Friedman, J.W. Tukey, A projection pursuit algorithm for exploratory data analysis, IEEE Trans. Comput. 100 (9) (1974) 881–890.
[24] J.A. Lee, M. Verleysen, Nonlinear Dimensionality Reduction, Springer Science & Business Media, 2007.
[25] B. Schölkopf, A. Smola, K.R. Müller, Kernel Principal Component Analysis, in: International Conference on Artificial Neural Networks, Springer, 1997, pp. 583–588.
[26] A.S. Rao, J. Gubbi, M. Palaniswami, Anomalous Crowd Event Analysis Using Isometric Mapping, in: Advances in Signal Processing and Intelligent Recognition Systems, Springer, 2016, pp. 407–418.
[27] X. Zheng, Y. Yuan, X. Lu, Dimensionality reduction by spatial-spectral preservation in selected bands, IEEE Trans. Geosci. Remote Sens. 55 (9) (2017) 5185–5197.
[28] M.M. Nezhad, E. Gironacci, M. Rezania, N. Khalili, Stochastic modeling of crack propagation in materials with random properties using isometric mapping for dimensionality reduction of nonlinear data sets, Int. J. Numer. Methods Eng. 113 (4) (2018) 656–680.

[29] D. Fitria, M.A. Ma'sum, E.M. Imah, A.A. Gunawan, Automatic arrhythmias detection using various types of artificial neural network-based learning vector quantization (lvq), Jurnal Ilmu Komputer dan Informasi 7 (2) (2014) 90–100.
[30] S. Hu, Y. Gu, H. Jiang, Study of Classification Model for College Students' M-learning Strategies Based on Pca-lvq Neural Network, in: 8th International Conference on Biomedical Engineering and Informatics (BMEI), IEEE, 2015, pp. 742–746.
[31] J. Lin, D. Gunopulos, Dimensionality Reduction by Random Projection and Latent Semantic Indexing, in: Proceedings of the Text Mining Workshop at the 3rd SIAM International Conference on Data Mining, 2003.
[32] F.M. Schleif, Small Sets of Random Fourier Features by Kernelized Matrix Lvq, in: Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM), 2017 12th International Workshop on, IEEE, 2017, pp. 1–5.
[33] L.v.d. Maaten, G. Hinton, Visualizing data using t-sne, Journal of machine learning research 9 (2008) 2579–2605. (Nov)
[34] A. Konstorum, E. Vidal, N. Jekel, R. Laubenbacher, Comparative analysis of linear and nonlinear dimension reduction techniques on mass cytometry data, bioRxiv (2018) 273862.
[35] A. Gisbrecht, A. Schulz, B. Hammer, Parametric nonlinear dimensionality reduction using kernel t-sne, Neurocomputing 147 (2015) 71–82.
[36] F.S. Tsai, K.L. Chan, Dimensionality Reduction Techniques for Data Exploration, in: 2007 6th International Conference on Information, Communications & Signal Processing, IEEE, 2007, pp. 1–5.
[37] U.D. Dixit, M. Shirdhonkar, Logo Based Document Image Retrieval Using Singular Value Decomposition Features, in: International Conference on Signal and Information Processing (IConSIP), IEEE, 2016, pp. 1–4.
[38] Rukshan Pramoditha, 11 Dimensionality reduction techniques you should know in 2021- Reduce the size of your dataset while keeping as much of the variation as possible, TowardsDataScience,
[39] S. Hao, Y. Xu, H. Peng, K. Su, D. Ke, Automated Chinese Essay Scoring from Topic Perspective Using Regularized Latent Semantic Indexing, in: 22nd International Conference on Pattern Recognition (ICPR), IEEE, 2014, pp. 3092–3097.
[40] Pulkit Sharma, The Ultimate Guide to 12 Dimensionality Reduction Techniques (with Python codes), Analytics Vidya, August 27, 2018
[41] F. Cai and V. Cherkassky, "Generalized SMO algorithm for SVM-based multitask learning," IEEE Trans. Neural Netw. Learn. Syst., vol. 23, no. 6, pp. 997–1003, Jun. 2012.
[42] S. Dhir and S.-Y. Lee, " Discriminant independent component analysis ", IEEE Transaction on Neural Networks, volume. 22, pages : 845 -857 2011



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY