

# Dimensionality Reduction and Classification through PCA and LDA

Telgaonkar Archana H.  
PG Student  
Department of CS and IT  
Dr. BAMU, Aurangabad

Deshmukh Sachin  
Associate professor  
Department of CS and IT  
Dr. BAMU, Aurangabad

## ABSTRACT

Information explosion has occurred in most of the sciences and researches due to advances in data collection and storage capacity in last few decades. Advance datasets with large number of observations present new challenges in data, mining, analysis and classification. Traditional statistical method breaks down partly because of the increase in the number of variables associated with each observation which is known as high dimensional data. Much of the data is highly redundant which can be ignored to extract features of dataset. The process of mapping of high dimensional data to lower dimensional space in such a way to discard uninformative variance from the dataset or finding subspace in which data can be easily detected is known as Dimensionality Reduction.

In this paper, well known techniques of Dimensionality Reduction namely Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are studied. Performance analysis is carried out on high dimensional data set UMIST, COIL and YALE which consists of images of objects and human faces. Classify the objects using knn classifier and naive bayes classifier to compare performance of these techniques. Difference between supervised and unsupervised learning is also inferred using these results.

## Keywords

Classification, Dimensionality reduction, KNN, LDA, PCA, naive bayes.

## 1. INTRODUCTION

Dimensionality reduction is the transformation of high-dimensional data into a meaningful representation of reduced dimensionality. In many real application such as face recognition [1], cross-media retrieval [2], text categorization, gene expression data classification [3], and image segmentation [4], pattern recognition, the dimensionality of data is very high, to deals with such high dimensional data need to reduce it without loss of information. Machine learning is used and served as the fundamental technique, in the topic of face detection, facial age estimation and face recognition. In face recognition use the most well-known eigenfaces which is based on principal component analysis (PCA)[5] and fisherfaces, which is based on linear discriminant analysis (LDA)[6] for dimensionality reduction and extracting personal features. machine learning method in pattern recognition, provide some theoretical analysis and practical guidelines to refine and improve the recognition performance.

For supervised method training set is labeled while in unsupervised method it is unlabeled. Labeled data are Limited, and to labeled large data would require expensive

human labor in practice. While unlabeled data is relatively easier to obtain. To effectively utilize both labeled and unlabeled data simultaneously semi-supervised learning was proposed and used in many real applications [7]. Many dimensionality reduction approaches have been proposed among that Principal Component Analysis (PCA) [5] and linear discriminant analysis (LDA) [6] are two popular methods which have been widely used in many classification applications. PCA is unsupervised while LDA is supervised and can achieve better classification results due to the utilization of label information. In this paper used PCA and LDA algorithm on the three image dataset that is coil-20, UMIST, YALE-B. PCA preserving as much of the variance in the high dimensional space as possible while LDA preserving as much of the class discriminatory information as possible.

## 2. DIMENSIONALITY REDUCTION

Increase the dimensionality refers to the problems associated with multivariate data analysis. While working with such large data needs to reduce the dimensionality, the goal of dimensionality reduction is to represent the data in a lower-dimensional space, by keeping some of the original properties of the data. Equation.1 shows the reduction of high-dimensional data space data [K] in to lower-dimensional space [N].

$$x = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \xrightarrow{\text{reduce dimensionality}} y = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} \quad (K \ll N) \quad (1)$$

At the intersection of several disciplines, data mining, databases, including statistics, pattern recognition, text mining, visualization, artificial intelligence and optimization Dimensionality reduction is important. There are several supervised (LDA, SVM, HNN), unsupervised (PCA, SVD, ICA) and semi supervised algorithm for dimensionality reduction. For dimensionality reduction, unsupervised PCA and supervised LDA use in this paper, on image dataset.

### 2.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is an unsupervised dimensionality reduction method, it is also named the discrete Karhunen–Loève transform (KLT), and Hotelling transform singular value decomposition (SVD) and empirical orthogonal function (EOF). PCA seeks to reduce the dimension of the data by finding a few orthogonal linear combinations (the principal components PCs) of the original variables with the largest variance. As per the number of the original variables there are as many PCs. The first several PCs explain most of the variance, so that disregarded the rest can be with minimal loss of information, for many datasets [1]. To reduce the

dimensionality of the huge data along with retaining as much information as possible in the original dataset, PCA is used.

To perform a Principal Components Analysis on data steps are [8]

1. Get image data: Suppose  $x_1, x_2, \dots, x_M$  are represented as  $N \times 1$  vectors

2. Compute the average of vector:  $\bar{x} = \frac{1}{M} \sum_{i=1}^M x_i$

3. Subtract the Mean:  $\Phi_i = x_i - \bar{x}$

4. Calculate the covariance matrix: Matrix  $A = [\Phi_1, \Phi_2, \dots, \Phi_M]$  ( $N \times M$  matrix) from this compute

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T$$

5. Compute eigenvalues and eigenvectors of the covariance matrix

$$C = \lambda_1 > \lambda_2 > \dots > \lambda_N \quad (\text{eigenvalues})$$

$$C = u_1, u_2, \dots, u_N \quad (\text{eigenvectors})$$

6. Forming a feature vector: eigenvectors are order by eigenvalue, highest to lowest. This gives the components in order of significance. The eigenvector with the highest eigenvalue is the principle component of the data set. Feature vector is formed by choosing the highest eigenvalue [8].

7. Deriving new dataset: we have chosen the principal components (eigenvectors) to keep in our data and formed a feature vector; we simply take the transpose of the vector and multiply it on the left of the original data set, transposed [8].

Final data = row feature vector \* row data adjust.

PCA is a technique applied in a large number of domains such as image compression, face recognition, patterns recognition, eigenfaces, and text categorization and computer vision.

## 2.2 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is supervised dimensionality reduction technique based on a linear projection from the high dimensional space to a low dimensional space by maximizing the between class scatter and minimizing the within-class scatter. LDA is also known as Fisher's linear discriminant. It is mainly used as a feature extraction step before classification and provides dimensionality reduction of feature vectors without loss of information. In LDA for all the samples of all classes, define two measures: 1) within-class scatter matrix

$$s_w = \sum_{j=1}^c \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T, \quad (2)$$

Where  $c$  is the number of classes,  $x_i^j$  is the  $i$  Th sample of class  $j$ ,  $\mu_j$  is the mean of class  $j$ , and  $N_j$  the number of samples in class  $j$ . next 2) between class scatter matrix

$$s_b = \sum_{j=1}^c (\mu_j - \mu)(\mu_j - \mu)^T, \quad (3)$$

Where  $\mu$  represents the mean of all classes. LDA methods applying the linear discriminant principle. This principle tries to maximize the ratio of determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples  $\frac{\det |s_b|}{\det |s_w|}$

Machine learning approach is used in Pattern recognition domain and it is difficult to work with the high-dimensional spaces; therefore, the dimensionality of the sample space must be reduced via feature extraction methods before applying the

classifier to data samples. To retain discriminatory information which the high-dimensional sample spaces provide, require good dimension reduction methods. Fig 1 shows the simple flow of feature extraction and classification.

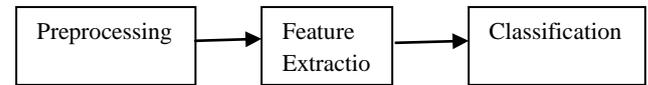


Fig 1: Feature extraction & classification

In preprocessing stage improve the representation of patterns, samples. Include registration, normalization, noise removal, resizing, segmentation depend upon nature of task. The goal of feature extraction is to reduce the number of features of patterns and retain as much as possible of their discriminatory information. After dimensionality reduction, the classification task, assign feature vector to a predefined class, based on their nature.

Feature extraction and dimension reduction can be combined in one step using principal component analysis (PCA), linear discriminant analysis (LDA) techniques, next apply the classifier on feature vectors in reduced-dimension space and perform classification. Fig 2 shows the testing phase of both PCA & LDA and final decision making based on classifier performance.

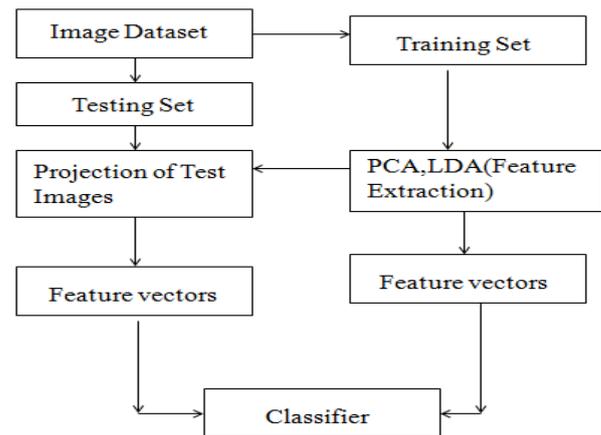


Fig 2: Step of testing phase PCA, LDA

## 3. CLASSIFICATION

Classification is the ordering or separation of samples into classes. There are different classifier such as support vector machines (SVM), decision tree (DT), neural network (NN), K-mean clustering, K-nearest Neighbor, naïve bayes etc. in this paper used two classifiers K-nearest neighbor and naïve bayes on image dataset after dimensionality reduction.

### 3.1 Naive Bayes classifier

A naïve bayes classifier is a simple probabilistic classifier with strong (naive) independence assumptions and based on applying Bayes theorem. The probability model would be "independent feature model"[9]. Naive Bayes classifiers can be trained very efficiently in a supervised learning setting, depending on the precise nature of the probability model. It is a simple classifier, requires a small amount of training data to estimate the parameters.

The naive Bayes probabilistic model:

The probability model is a conditional model for a classifier P (C|F<sub>1</sub>, F<sub>n</sub>) over a dependent class variable C with a small number of outcomes or classes, conditional on several feature variables F1 through F<sub>n</sub>.

Using Bayes' theorem reformulate the model  $p(C|F_1, F_n) = \frac{p(C)p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)}$ , also written as

Posterior = (prior×likelihood)/evidence.

Now assume that each feature F<sub>i</sub> is conditionally independent of every other feature F<sub>j</sub> for j ≠ i. This means that

$$p(F_i|C, F_j) = p(F_i|C)$$

$$= p(C) \prod_{i=1}^n p(F_i|C) \quad (4)$$

This means that under the above independence assumptions, the conditional distribution over the class variable C can be expressed like this:  $P(C|F_1, F_n) = \frac{1}{z} p(C) \prod_{i=1}^n p(F_i|C)$

Where z (the evidence) is a scaling factor, dependent only on F<sub>1</sub>, F<sub>n</sub>, i.e., a constant, if the values of the feature variables are known [9].

### 3.2 K Nearest Neighbor

The K-nearest neighbor algorithm is a non parametric, used for classifying based on closest training examples in the feature space. K-Nearest Neighbors (KNN) classification divides data into a test set and a training set. For test set row, the K nearest training set objects are found by Euclidean distance and the classification is determined by majority vote. If there is a condition, ties for the Kth nearest vector, all candidates are included in the vote. In KNN all the training data is needed during the testing phase and makes decision based on the entire training data set.

To choose the optimal value of neighborhood parameter k is key issue in k-nearest neighbor classification. The neighborhood parameter k, which controls the volume of the neighborhood and consequently the smoothness of the density estimates, plays an important role on the performance of a nearest neighbor classifier. Previous theoretical results (see e.g., Loftsgaarden and Quesenberry, 1965; Cover and Hart, 1968) suggest that if Euclidean distance is used for classification, one should vary k with n in such a way that  $k \rightarrow \infty$  and  $k/n \rightarrow 0$  as  $n \rightarrow \infty$  [10]. According to Shakhnarovich

et.al [11], larger values of k reduce the effect of noise on the classification, but make boundaries between classed issues distinct. Choosing an appropriate K is essential to make the classification more successful.

## 4. EXPERIMENTAL RESULT

### 4.1 Dataset

The In the experiments use four image databases, including Columbia Object Image Library (COIL)-20, YALE-B, and University of Manchester Institute of Science and Technology (UMIST) which is shown in Fig3. The UMIST database is a multiview face database, consisting of 575 images of 20 people, each covering a wide range of poses from profile to frontal views. The size of each cropped image is 32×32 with 256 gray levels per pixel [12]. The COIL-20 data set [13] consists of images of 20 objects viewed from varying angles at the interval of five degrees, total 72 images per object. Similarly, each image is down-sampled to the size of 32×32. The database YALE-B [14] consists of 38 subjects, with each

person having around 64 near frontal images under different illuminations condition; images are resized to 32×32 pixels.



Fig 3: samples of dataset COIL-20, UMIST, YALE-B

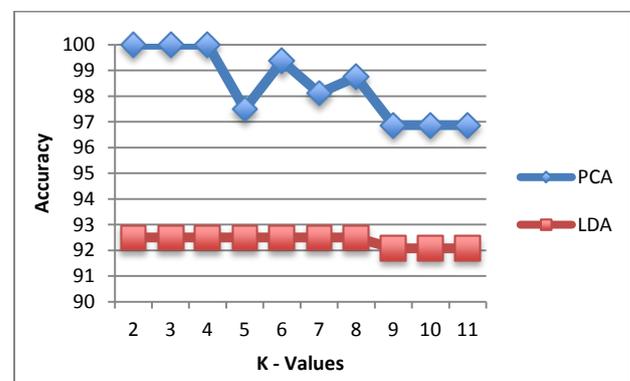
### 4.2 Performance analysis

Performance analysis take place on the basis of two dimensionality reduction technique PCA and LDA along with two classifier knn and naïve bayes on the object dataset COIL-20 and image dataset UMIST, YALE-B. In preprocessing stage down sample images in to 32×32 dimension, then reduce dimensionality using eigenvector by LDA and PCA method. For proper classification of samples in predefine classes use knn and naïve bayeson, on the basis of prediction value of classifier calculate mean accuracy and correlation on three dataset which is shown in table 1.

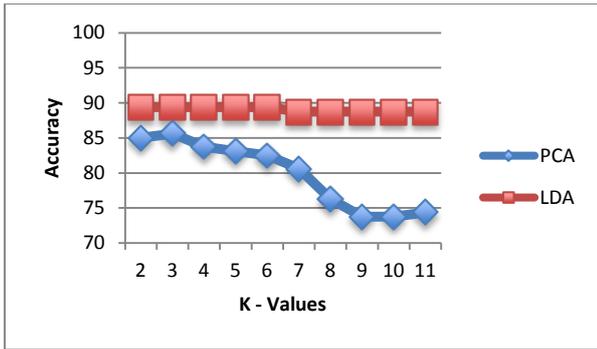
Table 1. Recognition Performance (Mean Accuracy, correlation)

Dataset	Method	KNN	Correlation	Naïve Bayes	Correlation
COIL-20	PCA	98.44	1	95.23	1
	LDA	92.37	1	86.74	1
UMIST	PCA	79.87	1	63.99	1
	LDA	89.06	1	83.87	1
YALE-B	PCA	58.75	1	94.35	1
	LDA	75.89	1	94.56	1

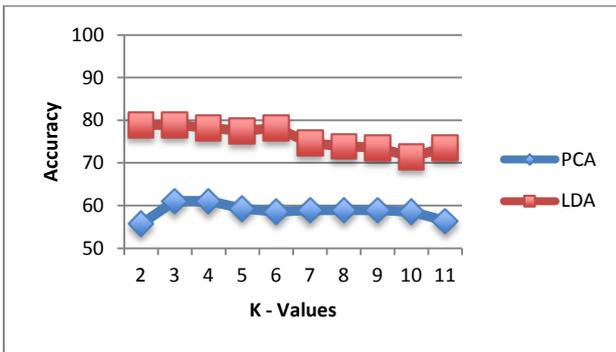
PCA on coil-20 dataset along with KNN and naïve bayes classifier gives better result, where LDA is supervised show accuracy in UMIST and YALE-B face dataset and perform better than PCA.



(A)

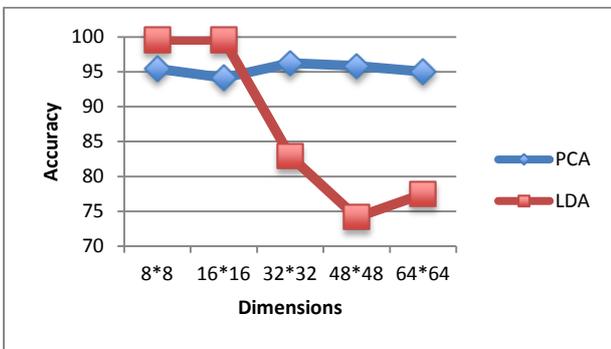


(B)

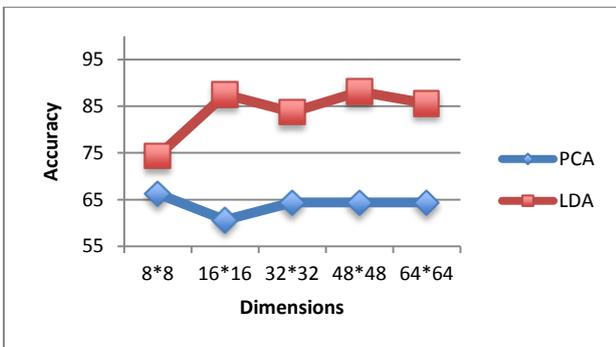


(C)

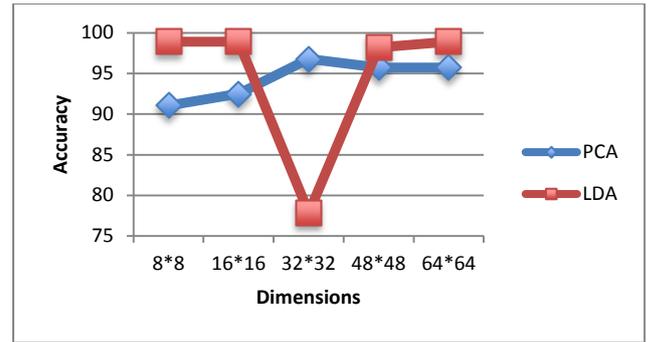
Fig 4: Accuracy versus the number of neighbors K in PCA,LDA on the three dataset (A) COIL-20,(B) UMIST,(C) YALE-B.



(A)



(B)



(C)

Fig 5: Accuracy versus the dimension in PCA,LDA on three dataset (A) COIL-20,(B) UMIST,(C) YALE-B.

Fig 4 shows the graph of accuracy versus k values for knn classifier and fig 5 shows graph of accuracy versus dimension for naïve bayes on three dataset. As per the k value increases the accuracy increase. knn is non parametric where naïve bayes is parametric classifier whose performance is calculated here based on dimension as a factor of sample. Naive bayes shows more variation in accuracy as compare to knn, accuracy of classifier depend upon feature extracted, as per dimension increase accuracy vary for specific dimension naïve bayes gives better result.

## 5. CONCLUSION

In this paper illustrate unsupervised PCA and supervised LDA dimensionality reduction method and make performance analysis of these techniques on high dimensional image dataset. Also analyze the accuracy of predicted values of classifier based on specific factor. As dimensionality is big challenge novel methods are expected to be developed depending upon the application. Labeled and unlabeled data utilize effectively, the semi supervised learning was proposed and has attracted much attention recently.

## 6. ACKNOWLEDGMENTS

We would like to thanks department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada Unuiversity, Aurangabad.

## 7. REFERENCES

- [1] Y. pang, Y. Yuan, and X. Li. 2008 Effective feature extraction in high dimensional space, IEEE Trans. Syst., Man, Cybern. B, Cybern.
- [2] Y. Yang, F. Wu, D. Xu, Y. Zhuang, and L.-T. Chia. 2010 Cross-media retrieval using query dependent search methods, Pattern Recognition.
- [3] R. Dudoit, J. Fridly, and T. P. Speed .2002 Comparison of discrimination methods for the classification of tumors using gene expression data, J. Amer. Stat. Assoc.
- [4] S. Xiang, F. Nie, C. Zhang, and C. Zhang. 2009 Interactive natural image segmentation via spline regression, IEEE Trans. Image Process.
- [5] M. Turk, A. Pentland. 1991 Face recognition using Eigenfaces, Computer Vision and Pattern Recognition, Proceedings CVPR'91, IEEE Computer Society Conference on (1991).
- [6] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman. 1997 Eigenfaces vs. fisherfaces: recognition using class

- specific linear projection, IEEE Trans. Pattern Anal. Mach.Intell.
- [7] Feiping Nie, Dong Xu, Xuelong Li.2011 Semi supervised Dimensionality Reduction and Classification through Virtual Label Regression.
- [8] Lindsay I Smith. 2002A tutorial on Principal Components Analysis.
- [9] Creative Commons Attribution-Share Alike 3.0 Unported <http://creativecommons.org/licenses/by-sa/3.0/>
- [10] Anil K. Ghosh. 2005 On optimum choice of k in nearest neighbor classification, Theoretical Statistics and Mathematics Unit, Indian Statistical Institute.
- [11] Shakhnarrovish, Darrell and Indyk. 2005 Nearest Neighbor Methods in learning and vision, MIT-press.
- [12] D. B. Graham and N. M. Allinson. 1998 Characterizing virtual eigensignatures for general purpose face recognition, in Face Recognition: From Theory to Applications
- [13] S. A. Nene, S. K. Nayar, and H. Murase. 1996 Columbia object image library (COIL-20), Columbia Univ., New York, Tech. Rep. CUCS-005-96
- [14] A. Georghiades, P. Belhumeur, and D. Kriegman. 2001 From few too many: Illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell.