A Hybrid Approach for Modeling High Dimensional Medical Data

Alok Sharma[1], Godfrey C. Onwubolu[1]

[1]*University of the South Pacific, Fiji*

sharma_al@usp.ac.fj, onwubolu_g@usp.ac.fj

Abstract. *This work presents the application of hybrid PCA and LDA to modeling high dimensional medical data, which is a real-life problem. For modeling and classifying medical data, we adopted this combination of two stage PCA and LDA procedure which is also known as Fisherface technique. During the training phase we applied this combination for extracting features from medical data. In the classification stage we introduced weighting ratio which is used with the conventional Euclidean distance measure to classify a given sample. For brevity we call this technique the weighted distance Fisherface technique. The presented technique shows promising results for medical data when compared with standard GMDH technique; in the two problems taken from the machining learning databases, the presented approach performed better than the standard GMDH.*

Keywords

Dimensionality reduction, inductive modeling, classification, PCA, LDA.

1 Introduction

Modeling and pattern classification plays crucial role in everyday life. The evolving computational demand makes this field very challenging and thus open for research. When high dimensional feature vectors are involved in the computation it makes the implementation of model and/or pattern classifier quite difficult and sometimes impossible. This limitation is usually referred as the *curse of dimensionality*. Efforts are undergoing to reduce the complexity in an efficient manner and at the same time achieve sufficient level of classification accuracy.

The two well known linear techniques for dimensionality reduction are principal component analysis (PCA) [6] and linear discriminant analysis (LDA) [4]. The goal of PCA is to find a parsimonious data space from the original data space such that the representation information is maximally preserved. The features in the reduced dimensional plane are transformed from higher dimensional space such that the mean squared error is minimum. On the other hand, LDA provides a reduced dimensional space such that the features of different classes or categories are maximally discriminated. In the case of solving high dimensional problem several authors applied PCA technique prior to the LDA technique [1][8][10][11].

In the work reported in this paper, we adopted this combination of two stage PCA and LDA procedure which is also known as Fisherface technique for *modeling* and *classifying* medical data. During the training phase we applied this combination for extracting features from medical data. In the classification stage, we introduced weighting ratio which is used with the conventional Euclidean distance measure to classify a given sample. For brevity we call this technique the weighted distance Fisherface technique. The presented technique shows promising results for medical data when compared with Group Method of Handling Data (GMDH) technique.

2 Model Description

Inductive modeling aims at constructing an efficient and effective model of high dimensional data. In a given set of inputs, system state, and outputs, the third component is always deducible with the other two at hand (see Fig. 1). A training dataset of inputs, **X**, and system states, **S**, can be used to estimate the ensuing outputs, **Y**, in a *prediction* or forecast model. It is a *modeling* or design problem to obtain a system, **S**, for given inputs and outputs **X** and **Y**. A *control* problem is to seek the optimal inputs, **X** for a given system states, **S**, that can be used to estimate the ensuing outputs, **Y**. These concepts are well described by Elder and Brown [5].

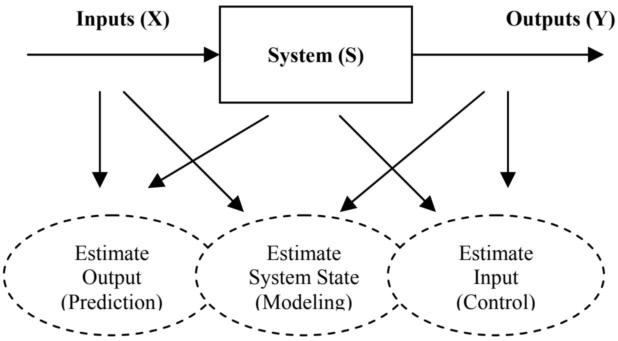

Fig. 1 Inductive modeling

The system presented here is the weighted distance Fisherface model which has been applied for medical data analysis. The basic block diagram of the model is shown in Fig. 2. The model has an output variable or target variable y which depends on d-dimensional input vector $\mathbf{x} = [x_1, x_2, ..., x_d]^T$ and the model parameters. The model parameters can be estimated by using training data where the output y and input $\mathbf{x}$ are known quantities. Once the model is estimated then it can be used to provide output y for any unknown $\mathbf{x}$.

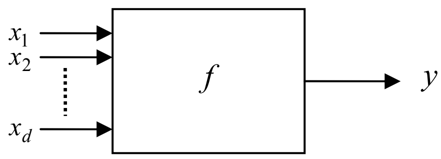

Fig. 2. Basic block diagram of the model

The target variable can be represented in the form of function f and the input as

$$y = f(\mathbf{x}) = f(x_1, x_2, \ldots, x_d) \quad (1)$$

The function f is a combination of two linear functions namely PCA and LDA. We conclude that our modeling approach in Fig. 2 follows the inductive modeling architecture in Fig. 1.

PCA finds a linear transformation $\mathbf{\Phi}$ which reduces d-dimensional data to h-dimensional feature vectors (where $h < d$) in such a way that the information is maximally preserved in minimum mean squared error sense. This linear transformation is known as PCA transform or Karhunen-Loéve transform (KLT) [6]. Since the transformation is from d-dimensional feature space to h-dimensional feature space the size of $\mathbf{\Phi}$ is $d \times h$. The h column vectors of the matrix $\mathbf{\Phi}$ are the basis vectors. The first basis vector is in the direction of maximum variance of the given feature vectors. The remaining basis vectors are mutually orthogonal and, in order, maximize the remaining variances subject to the orthogonal condition. Each basis vector represents a principal axis. These principal axes are those orthonormal axes onto which the remaining variances under projection are maximum. These orthonormal axes are given by the dominant/leading eigenvectors (i.e. those with the largest associated eigenvalues) of the measured covariance matrix. In PCA, original feature space is characterized by these basis vectors and the number of basis vectors used for characterization is usually less than the dimensionality d of the feature space [8][9].

In LDA, the dimensional embeddings are reduced in such a way that the orientation of the projected data of classes on an arbitrary line or space is well separated from each other. The transformation vectors are taken so that the criterion J is maximum, where J is the ratio of between-class scatter matrix (S_B) and within-class scatter matrix (S_W) [4]. In a c-class problem the LDA projects from d-dimensional space to $c-1$ or less dimensional space ($\mathbf{R}^{\mathrm{d}} \rightarrow \mathbf{R}^{\mathrm{c-1}}$). There are some limitations in applying LDA directly viz. matrix S_W can become singular due to high dimensionality of original feature vectors in comparison with low number of training vectors available. To overcome this limitation, a number of authors have proposed the use of PCA prior to the application of LDA [1][8][10][11] in the feature extraction stage.

In medical data analysis the PCA technique is applied for two main reasons

(i) the basis vectors that are of less importance can be discarded which would help in reducing the noise that could be present in medical data.
(ii) to overcome the singularity issue related with the direct application of LDA.

The LDA is applied after the application of PCA to give such feature space in which different class data are maximally separated under discriminant criterion. In addition, the weighted distance measure has been applied for classifying input vectors which was not applied previously in the literature. The following section briefly illustrates the mathematical details of the model.

2.1 Principal Component Analysis

The PCA transform can be found by minimizing mean squared error. To see this, let the feature vector be $\mathbf{x} \in \mathbf{R}^d$ (d-dimensional space), reduced dimensional feature vector be $\mathbf{z} \in \mathbf{R}^h$ and reconstructed feature vector be $\hat{\mathbf{x}} \in \mathbf{R}^d$. Then the mean squared error can be represented as

$$\text{MSE} = E[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$$

where $E[\bullet]$ is the expectation operation with respect to $\mathbf{x}$ and $\|\bullet\|^2$ is the norm squared value. We know that PCA transformation $\mathbf{\Phi}$ is of size $d \times h$ and it is used to do dimensionality reduction from d-dimensional space to h-dimensional feature space, i.e. $\mathbf{\Phi} : \mathbf{x} \rightarrow \mathbf{z}$ or

$$\mathbf{z} = \mathbf{\Phi}^{\mathrm{T}} \mathbf{x} \tag{2}$$

The PCA transformation $\mathbf{\Phi}$ can be obtained by minimizing mean squared error $E[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$ which turns out to be a generalized eigenvalue problem i.e.:

$$\Sigma_{\mathbf{x}} \boldsymbol{\varphi}_j = \lambda_j \boldsymbol{\varphi}_j \tag{3}$$

where $\mathbf{\Phi} = \{\boldsymbol{\varphi}_j : j = 1,2,\ldots,h\}$, $\boldsymbol{\varphi}_j \in \mathbf{R}^{\mathrm{d}}$, and $\Sigma_{\mathbf{x}}$ is covariance matrix of all input d-dimensional vectors. The expression λ_j denotes eigenvalues corresponding to $\boldsymbol{\varphi}_j$. The eigenvectors $(\boldsymbol{\varphi}_1,\ldots,\boldsymbol{\varphi}_h)$ of $\mathbf{\Phi}$ should be arranged such that their corresponding eigenvalues are in descending order $\lambda_1 > \lambda_2 > \ldots > \lambda_h$. This arrangement is, however, not a necessary step for PCA but it is mentioned here since the model conducts this arrangement process prior to the application of LDA.

2.2 Linear Discriminant Analysis

The LDA technique can be illustrated for two-class problem and multi-class problem. The LDA technique for multi-class problem is briefly described here. Let a c-class problem (assuming $c > 2$) is given with c unique class lables $\omega_1, \ldots, \omega_c$. In LDA the projection is from h-dimensional feature space to k-dimensional feature space where $k < h$ such that the samples or patterns of classes are well-separated. For a c-class problem the transformation can be given as:

$$\mathbf{s} = \mathbf{W}^{\mathrm{T}}\mathbf{z} \text{ where } \mathbf{s} \in \mathbf{R}^{\mathrm{k}} \text{ and } \mathbf{z} \text{ is from equation 2.} \tag{4}$$

The transformation matrix **W** is computed by maximizing Fisher's criterion $J(\mathbf{W}) = |\mathbf{W}^{\mathrm{T}} S_B \mathbf{W}| / |\mathbf{W}^{\mathrm{T}} S_W \mathbf{W}|$. The computation of between-class scatter matric S_B and within-class scatter matrix S_W can be computed from the vectors **z**. See Duda and Hart [4] for details of computing these matrices. The transformation matrix **W** is given by [4]:

$$S_B \mathbf{w}_i = \lambda_i S_W \mathbf{w}_i \tag{5}$$

where $\mathbf{W} = \{\mathbf{w}_i : i = 1, 2, \ldots, k\}$. The eigenvectors $\mathbf{w}_i$ (columns of **W**) correspond to the eigenvalues λ_i. Since the rank of between-class scatter matrix S_B is $c-1$ or less, $k \le c-1$. The next section elaborates the training phase of the model.

2.3 Training Phase of the Weighted Distance Fisherface Model

The training phase of the model is depicted in Fig. 3. The training data is processed through the PCA

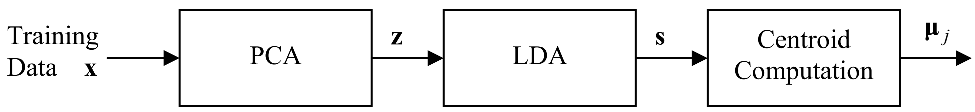

Fig. 3. Training phase of the model

block to give parsimonious data space. The feature vectors **z** are obtained after PCA process. These features are then sent to the LDA block which produces discriminant features **s**. The obtained features **s** of similar classes will be sent to the Centroid Computation block which will give respective centroid for each class. The training process can be summarized as follows:

1. Compute PCA transformation $\mathbf{\Phi}$ for input vectors **x** using equation 3.
2. Project the samples on lower dimensional space $\mathbf{R}^{\mathrm{h}}$ using equation 2. This will give **z** feature vectors.
3. Compute transformation **W** using eigenvalue decomposition (equation 5).
4. Project the **z** feature vectors on to k-dimensional space by using equation 4. This will give **s** feature vectors.[1]

[1] It can be observed from equations 2 and 4 that feature vector **s** is the linear combination of the elements of input vector **x** and can be represented as

$$s_m = \sum_{n=1}^{d} x_n v_{nm} \quad \text{for } m = 1, \ldots, k$$

where $v_{nm} = \sum_{j=1}^{h} w_{jm}\phi_{nj}$, x_n is the element of **x**, s_m is the element of **s**, w_{jm} (jth row and mth column) is the element of **W** and ϕ_{nj} (nth row and jth column) is the element of $\mathbf{\Phi}$.

5. Compute the centroid of each class. This will give centroid vector $\boldsymbol{\mu}_j$ for $j = 1,\ldots,c$.

The parameters that are required to store for testing phase are $\boldsymbol{\Phi}$, $\mathbf{W}$ and $\boldsymbol{\mu}_j$. The classification phase or testing phase is illustrated in the following section.

2.4 Classification Phase of the Weighted Distance Fisherface Model

The classification phase of the model is depicted in Fig. 4. The input vector **x** with unknown class label is entered in PCA block where feature vector **z** is computed using PCA transformation $\boldsymbol{\Phi}$. The transformed vector **z** is then processed through the LDA block which produced feature vector **s** using LDA transformation **W** at its output. The feature vector **s** then sent via the weighted distance measure block where weighted distance δ_j (for $j = 1,\ldots,c$) is computed. All the distance are evaluated in the comparison block and the class label is associated to the input vector for which the distance is minimum. The classification phase is summarized as follows:

1. Compute **z** for input vector **x** using equation 2 where $\boldsymbol{\Phi}$ is a known quantity.
2. Compute **s** from the transformation **W** and input **z** using equation 4.
3. Compute weighted distance δ_j between a test vector **s** and the centroid $\boldsymbol{\mu}_j$

$$\delta_j = (1/wt_j) \| \mathbf{s} - \boldsymbol{\mu}_j \|^2 \quad \text{for } j = 1,\ldots,c$$

The term wt_j is the weighting ratio which reflects *a priori* probability and can be given by $wt_j = \text{number of samples in class } j \,/\, \text{total number of samples in the training data}$.

4. Find the argument for which the weighted distance δ_j is minimum.

$$k = \mathbf{arg}\min_{j=1}^{c} \delta_j$$

Assign the class label ω_k to the test pattern **x**. The class label is the output of the model or the target variable i.e. $y = \omega_k$.

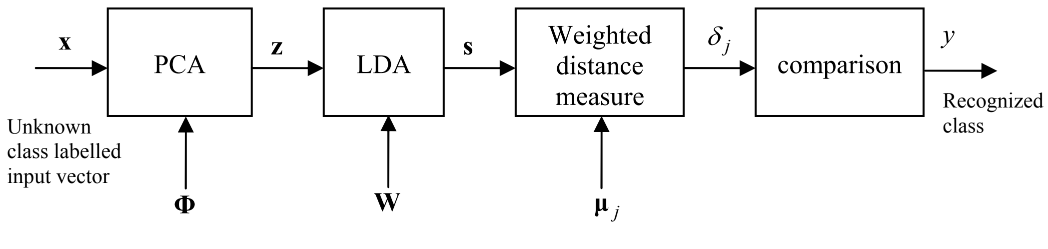

Fig. 4. Classification phase of the model

3 Heart Disease Medical Database

The Long Beach [2] heart disease medical database has been used in the experimentation. The database is provided by Robert Detrano of V.A. Medical Center, Long Beach and Cleveland Clinic Foundation. The dataset contains 200 cases and 75 attributes, but all published experiments [3][7] refer to using a subset of 13 of them. There are five classes in this dataset. The target variable discriminates between five levels of heart disease where label '0' indicates no presence of heart disease and '1', '2', '3', '4' represent the presence of heart disease at a gradually increased level. Therefore there are two basic types of classification problem (i) a binary classification problem that will detect the existence of heart disease (level 1-4) or no heart disease (level 0) and (ii) a five class classification problem that identifies all the five levels accurately. The system is modeled using the original 75 attribute Long Beach data and reduced Long Beach data of 13 attributes. The data is randomly split into 180 cases for learning the model and 20 cases for appraising the model. The experimentation has been described in the following section.

4 Experimentation on Medical Database

This section demonstrates the performance of the proposed classifier in comparison with GMDH model. The GMDH model was applied on heart disease medical dataset by Lemke and Mueller [7]. For two-class problem the target variable would be either 'has' or 'has not' that signify the presence of heart disease and absence of heart disease respectively. For five-class problem the target variable will be 0,1,2,3 and 4 that signify no heart disease and four levels of heart disease. The classification accuracy in percentage and the number of false classified cases are depicted in Tab. 1 for the testing data on Long Beach 75.

Tab. 1. Classification results for Long Beach 75 dataset

Test	GMDH [7]		Weighted distance Fisherface	
	Has/has not	Class 0-4	Has/has not	Class 0-4
False classified	2	5	0	4
Accuracy [%]	90.00	75.00	100.00	80.00

It can be observed from Tab. 1. that the weighted distance Fisherface model produces 100% accuracy for two-class problem whereas GMDH produces only 90% in this case. On the other hand, the proposed model is giving 80% accuracy for five-class problem whereas GMDH is giving only 75% accuracy. The improvement is significant in both the cases. In the case of two-class problem the PCA dimension (*h*) was 3 and LDA dimension (*k*) was 1 for the proposed model and for five-class problem $h = 7$ and $k = 4$.

Table 2 depicts the classification accuracy in percentage and the number of false classified cases for the testing data on Long Beach 13.

Tab. 2. Classification results for Long Beach 13 dataset

Test	GMDH [7]		Weighted distance Fisherface	
	Has/has not	Class 0-4	Has/has not	Class 0-4
False classified	7	16	5	10
Accuracy [%]	65.00	20.00	75.00	50.00

It is evident from Tab. 2. that the weighted distance Fisherface model produces 75% and 50% accuracy for two-class problem and five-class problem respectively. On the other hand, GMDH produces only 65% and 20% accuracy for two-class problem and five-class problem respectively. The PCA dimension was $h = 3$ and LDA dimension was $k = 1$ for two class problem and $h = 11$ and $k = 1$ for five-class problem. The values of *h* and *k* are selected for which the model is producing the best results. Thus the results obtained depicted are the best results that can be achieved by the model.

5 Conclusion

This paper has presented an inductive modeling approach based on combination of two stage PCA and LDA procedure which is also known as Fisherface technique for *modeling* and *classifying* medical data. For extracting features from medical data we have applied this combination of techniques and in the classification stage we added weighting ratio which was used with the conventional Euclidean distance measure to classify a given sample. We referred this technique as the weighted distance Fisherface technique. We conclude that our modeling approach in Fig. 2 follows the inductive modeling architecture in Fig. 1. For the two problems taken from the machining learning databases, the presented approach performed better than the standard GMDH. It can be said from the obtained results that the weigthed distance Fisherface model is producing better results than the GMDH model while experimented on heart disease medical database. The possible future research work would be to hybridize the techniques with GMDH-like networks.

References

[1] Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7), p. 711-720, 1997.

[2] Blake, C.L., Merz, C.J.: UCI repository of machine learning databases. http://www.ics.uci.edu/~mlearn, Irvine, CA, University of Calif., Dept. of Information and Comp. Sci., 1998.

[3] Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., Guppy, K., Lee, S., Froelicher, V.: International application of a new probability algorithm for the diagnosis of coronary artery disease. *American Journal of Cardiology*, 64, p. 304-310, 1989.

[4] Duda, R.O., Hart, P.E.: Pattern classification and scene analysis. John Wiley and Sons, New York, 1973.

[5] Elder IV, J.F., Brown, D.E.: Inductive and polynomial networks. Technical Report IPC-TR-92-009, Institute for parallel computation and department of systems engineering, University of Virginia, Charlottesville, VA, USA, 1992.

[6] Fukunaga, K.: Introduction to statistical pattern recognition. Academic Press Inc., Hartcourt Brace Jovanovich, Publishers, 1990

[7] Lekme, F., Mueller, J.-A.: Medical data analysis using self-organizing data mining technologies. *Systems Analysis Modelling Simulation*, 43(10), p. 1399-1408, 2003.

[8] Sharma, A., Paliwal, K.K., Onwubolu, G.C.: Class-dependent PCA, LDA and MDC: a combined classifier for pattern classification. *Pattern Recognition*, 39(7), p. 1215-1229, 2006.

[9] Sharma, A., Paliwal, K.K.: Fast Principal Component Analysis using Fixed-Point Algorithm", *Pattern Recognition Letters*, 28, p. 1151-1155, 2007.

[10] Swets, D.L., Weng, J.: Using discriminative eigenfeatures for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8), p. 831-836, 1996.

[11] Zhao, W., Chellappa, R., Phillips, P.J.: Subspace linear discriminant analysis for face recognition. Tech. Rep. CAR-TR-914, Center for Automation Research, University of Maryland, College Park, 1999.