MULTIPLE DISCRIMINANT ANALYSIS

Ramasubramanian Sundaram
Department of Electrical and Computer Engineering
Mississippi State University
Mississippi State, MS 39762 USA
email: sundaram@isip.mstate.edu

ABSTRACT

Fisher's linear discriminant is a well known discriminative technique for solving two-class problems. But when the number of classes is more than two, then several discriminative and representative techniques are used. Multiple Discriminant Analysis (MDA) is one such discriminative technique used for classification. Modified Multiple Discriminant Analysis (MMDA) is used when computations become unstable when using conventional MDA. MMDA extracts discriminant features from the observation patterns by using linear transformation and redefining the optimization index. Experiments were performed to compare MDA with MMDA on both static and temporal data. MMDA gives a classification error of 47.49% and 34.86% respectively and also shows significant improvement by using MMDA.

1. INTRODUCTION

In practical multi category problems, it is desirable that the features belonging to a particular class are far apart from the features of other competing categories. The feature extraction process needs to be effective so that salient features can be extracted that can differentiate between various classes [1]. Several methods like Principal Component Analysis [2,3], transform the input data so that the features are well separated and classification becomes easier. Linear transformations are extensively used because they are easy to compute and analytically tractable. Typically, these transformations involve projecting features from a high dimensional space to a lower dimensional space where they are well separated. This transformation could be representative or discriminative depending upon the problem.

Linear Discriminant Analysis (LDA) is one such discriminative technique based on Fisher's linear discriminant [4]. When the number of categories is more than two, then Multiple Discriminant Analysis is used. This transforms the data in such a way that the features are best separated in the least-squares sense. Modified Multiple Discriminant Analysis (MMDA) is used when computations become unstable due to singular matrix problems. In MMDA, the optimization index is modified to overcome singular matrix problems. This improves the classification accuracy significantly as evident from the experiments performed.

This paper is organized as follows. In the next section, we discuss the Fisher's linear discriminant for two class problem which serves as a building block for further discussion. Section 3 and 4 reviews the MDA algorithm and the changes needed to overcome the deficiencies of the conventional MDA. The experimental setup, training and testing process, followed by an analysis of the results are discussed in Section 5.

2. FISHER LINEAR DISCRIMINANT

Representative techniques transforms the data that are best represented in the least squared sense. But these features may not be the best way to discriminate between data in different classes. In fact, the directions that are discarded by representative techniques like PCA may be the directions that actually distinguish between several classes and hence may be very important. Discriminant analysis seeks to project the data along those directions that help in a better discrimination.

To make the concept much more simple, it would be elegant if the high dimensional data can be projected on to a line. This would help in visualizing data in a better way and would involve simple linear transformations. Fisher's linear discriminant is used for discrimination in case of two class problems. Even when the samples are well-separated in the higher dimensional space, projecting them onto a line may lead to confused clusters. But by changing the direction of the projecting line we may get a better discrimination in one particular orientation.

Suppose that we have a set of n d-dimensional samples $x_1, x_2, \ldots, x_n$ belonging to two different classes namely ω_1 and ω_2. If the output can be expressed as a linear combination of the components, then the scalar product

$$y = w' x \tag{1}$$

represents the projection of the high dimensional data onto a line. The magnitude of w is not very important but the direction of w is of significance. This is because the projections of samples belonging to different class need to be well separated while the samples belonging to the same class need to be clustered together. In other words, an optimal transformation should maximize the ratio of between class scatter to the within class scatter.

If x denotes a feature vector, then for a two category classification, the scatter matrices are given as follows

$$S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)' \tag{2}$$

and

$$m_i = \frac{1}{n_i} \sum_{x \in D_i} x \tag{3}$$

The total within-class matrix is given by

$$S_W = S_1 + S_2 \tag{4}$$

The relationship between the scatter matrix in the original space and the transformed space is given by

$$\tilde{s}_i^2 = w' S_i w \tag{5}$$

Similarly, the between-class scatter matrix is given by

$$S_B = (m_1 - m_2)(m_1 - m_2)' \tag{6}$$

and the relationship between the between class scatter matrices in the two different spaces is given by

$$(\tilde{m}_1 - \tilde{m}_2)^2 = w' S_B w \tag{7}$$

Since the criterion is to maximize the ratio of the between-class scatter matrix to the within-class scatter matrix, the optimal w is given by

$$J(w) = \frac{w' S_B w}{w' S_W w} \tag{8}$$

If S_W is non-singular, then (8) can be solved as a conventional eigenvalue problem and the eigenvectors of which gives the required transformation. Since the samples are well separated in the one dimensional space, a classifier can be easily built to get more accurate results.

3. MULTICATEGORY ANALYSIS

If the number of classes are more than two, then a natural extension of Fisher linear discriminant exists using c-1 discriminant functions and is called Multiple Discriminant Analysis (MDA) [3]. As in two category case, the projection is from a high dimensional space to a low dimensional space. If the input feature vector has d dimensions, then the projection is on to a c-1 dimensional space under the assumption that $d \geq c$. The optimal transformation should still maximize the ratio of between-class scatter to the within-class scatter. But unlike the two category case, the maximization should be done among several competing classes.

The within-class scatter matrix are calculated similar to (2) and the total within-class is given by

$$S_W = \sum_{i=1}^{c} S_i \tag{9}$$

The between-class scatter matrix slightly differs in computation and is given by

$$S_B = \sum_{i=1}^{c} n_i(m_i - m)(m_i - m)' \qquad (10)$$

where n_i is the number of training samples for each class, m_i is the mean for each class and *m* is the total-mean vector and is given by

$$m = \frac{1}{n}\sum_{i=1}^{c} n_i m_i \qquad (11)$$

As in two class problem, the criterion function is to maximize the ratio of between-class scatter and within-class scatter given by

$$J(W) = \frac{W'S_BW}{W'S_WW} \qquad (12)$$

The columns of the optimal rectangular matrix W are the eigenvectors of the eigenvalues of the equation

$$S_Bw_i = \lambda_i S_W w_i \qquad (13)$$

where w_i is the i^{th} column of the optimal transformation matrix W. If S_W is non-singular then (13) can be converted to a conventional eigenvalue problem.

4. MODIFIED MDA

Multiple Discriminant Analysis provides an elegant way for classification using discriminant features. But for small category problems, the rank deficiency of the scatter matrices poses a severe problem for classification. Rank deficiency are problematic in the sense that the eigenvalue computations may not converge to the required accuracy. This may give sub-optimal values for W and can affect the classification accuracy drastically.

To overcome this deficiency in the rank values, Modified Multiple Discriminant Analysis is used [5]. The criterion remains the same as in MDA but the transformation matrix is computed to maximize the between-class scatter and the transformation feature space is normalized for the within-class scatter. Since the transformation space is normalized, the rank of the within-class scatter matrix is no longer a problem which causes computational instability [6]. The criterion function is defined as

$$J(W) = W'S_BW \qquad (14)$$

and the within-class scatter of each dimension of the feature vector is recomputed as follows

$$r_i = w_i'S_W w_i \qquad (15)$$

where w_i is the i^{th} column of the transformation matrix.

5. EXPERIMENTS

In order to classify the input data using MMDA, a classifier was designed as follows to perform training and testing.

1. The mean of the each dimension of the transformed feature vector is calculated for each class i*:*

$$m_l^i = E\left\{y_l^i\right\} \qquad (16)$$

y_l^i is the transformed feature obtained by the transformation equation given by

$$y = W'x \qquad (17)$$

2. For the input test features that need to be classified, the feature distance to class C_k is defined as:

$$D_k(x) = \sum_{l=1}^{K} (w_l'x - m_l^k)/(\sqrt{r_l}) \qquad (18)$$

where r_l is obtained as in (15)

3. The test data is said to belong to a particular class if the distance is the closest to that class.

Two sets of data was chosen for training and testing. The first set had 11 distinct classes and each class had 48 training samples to train from. Each feature vector had 10 elements. Test set included a set of 379 test samples that were fully representative of all the classes. The second set was a temporal set of data that had 925 training samples and 5 distinct classes. Each feature vector had 39 elements and the test set included 225 test samples. The experiments were

Table 1: Performance of MDA and MMDA on both static and temporal test data. MMDA performs significantly better than MDA in both the cases

Test Data	Classification Error (%)	
	MDA	MMDA
Static Data	74.19	47.49
Temporal	71.14	34.86

performed on MDA and also on MMDA to verify improvements if any and the results are shown in Table 1.

The results show that the MMDA performs significantly better than the MDA for both static and temporal test data. This is because the rank of the within-class scatter matrix is very low in both the cases. Hence, if the within-class scatter is not normalized, the eigenvalue computation does not converge to the required accuracy. For MMDA, the eigenvalue computations converges to zero after several iterations while for conventional MDA it does not converge to zero. This leads to sub-optimal transformation matrix that causes significant classification errors during testing. MMDA does not use any temporal information for classification. Hence, the performance is not better for temporal test data because of any temporal information present in the training set. The performance is better for temporal set than for static set because of the amount of training data. The amount of training data for the temporal set is more than that for the static set. In addition to this, the number of distinct classes for the static set is more than that in the temporal set. This leads to more confusability among classes.

Another reason for temporal data performing better than the static data is the dimension of the feature vector itself. Multiple Discriminant Analysis tactically assumes that the dimensionality of the original feature vector is greater than or equal to the number of distinct classes. In case of static data, the feature vectors had a dimensionality of 10 and the number of distinct classes were 11. Since we built 10 discriminant functions for the static training data, the dimensionality was not reduced in any sense and this does not augur well for the algorithm. For temporal data, the original feature vector had 39 dimensions and the transformations projected this onto a 4 dimensional feature space for better discrimination.

6. CONCLUSIONS

This paper presented the advantages in using the Modified Multiple Discriminant Analysis over the conventional MDA. The MMDA and MDA algorithms perform classification by projecting the samples onto a lower dimensional space along the directions that are efficient for discrimination. In addition to classification based on discrimination, the MMDA algorithm overcomes the rank deficiency of the conventional MDA by normalizing the within-class scatter matrix prior to the eigenvalue computations. This helps in obtaining an optimal transformation matrix as the eigenvalue computations converge to the required accuracy. The results show significant improvement in using Modified MDA over MDA for both static and temporal test data.

7. REFERENCES

[1] H. Watanabe et al, "Discriminative Metric Design for Robust Pattern Recognition", IEEE Transactions on Signal Processing, vol. 45, pp. 2655-2662, 1997.

[2] Ian T. Jollife, Principal Component Analysis, Springer Verilag, New York, 1986.

[3] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification and Scene Analysis, Wiley Interscience, 2000.

[4] Geoffrey J. McLachlan, Discriminant Analysis and Statistical Pattern Recognition, Wiley, New York, 1992.

[5] Jiang Gao, Xiaoqing Ding, Youshou Wu, "On Improvement of Multiple Discriminant Analysis Method for Discriminative Feature Extraction", IEEE *SMC Conference Proceedings*, vol. 2, pp. 915-920, 1999.

[6] M. Zoltowski, "Solving the Generalized Eigenvalue Problem with Singular Forms", Proceedings of IEEE, vol. 75, no. 11, pp. 1546-1548, 1987.