Linear Discriminant Clustering in Pattern Recognition

*Zhaojia Sun[1], Miseon Choi[2], Young-Kuk Kim[1]
[1]Department of Computer Science and Engineering
Chungnam National University
[2]Human Resource Development Consortium for Next Generation Software in Information Technology, Chungnam National University
e-mail : hellojiajia_1@hotmail.com,{mschoi27, ykim}@cnu.ac.kr

Abstract

Fisher Linear Discriminant(FLD) is a sample and intuitive linear feature extraction method in pattern recognition. But in some special cases, such as un-separable case, one class data dispersed into several clustering case, FLD doesn't work well. In this paper, a new discriminant named K-means Fisher Linear Discriminant, which combines FLD with K-means clustering is proposed. It could deal with this case efficiently, not only possess FLD's global-view merit, but also K-means' local-view property. Finally, the simulation results also demonstrate its advantage against K-means and FLD individually.

I. Introduction

1.1 Fisher Linear Discriminant

FLD aims at finding the directional vector onto which distance between the cluster centers of the projected patterns is made as long as possible and dispersion of each class of them as small as possible. In two-class classification, the Rayleigh coefficient is as follows: $J(w)=\frac{w^T S_B w}{w^T S_W w}$ (1)

where $S_B=(m_1-m_2)(m_1-m_2)^T$
$S_W=\sum_{i=1,2}\sum_{x\in I_k}(x-m_i)(x-m_i)^T$

This research was supported by the **Ministry of Knowledge Economy**, Korea, under the Information Technology Research Center support program supervised by the Institute of Information Technology Advancement. (grant number **IITA-2008-C1090-0801-0031)**

where S_B is called a between-class scatter matrix, S_W a within-class scatter matrix, and m_k, I_k denotes the sample mean and the index set for class k, respectively. Maximizing the Rayleigh coefficient with respect to w is reduced to find w that satisfies $S_W^{-1}S_B w=\lambda w$ for the only positive nonzero eigen value λ. Because the matrix S_B is rank-one and so is the matrix S_W-1S_B, the matrix S_W-1S_B has zero eigenvalues except one positive eigenvalue [2].

1.2 K-means Clustering

K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The main idea is to define k centroids, one for each cluster. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. Then a loop has been generated, the k centroids change their location step by step until no more changes are done. The objective function [3]

$$J=\sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|x_n-\mu_k\|^2 \quad (2)$$

Which represents the sum of the squares of the distance of each point to its assigned vector μ_k. Our goal is to find values for the $\{r_{nk}\}$ and the $\{\mu_k\}$ so as to minimize J.

Where $r_{nk}=\begin{cases}1 & if\ k=argmin_j\|x_n-\mu_j\|^2\\ 0 & otherwise\end{cases}$

Equation (2) can be minimized by setting its derivative with respect to μ_k to zero giving

$$2\sum_{n=1}^{N} r_{nk}(x_n-\mu_k)=0 \quad (3)$$

Which we can easily solve for μ_k to give

$$\mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}} \tag{4}$$

II. K-means Fisher Linear Discriminant

Let's specify the principle of this algorithm combining with a simple classification case.

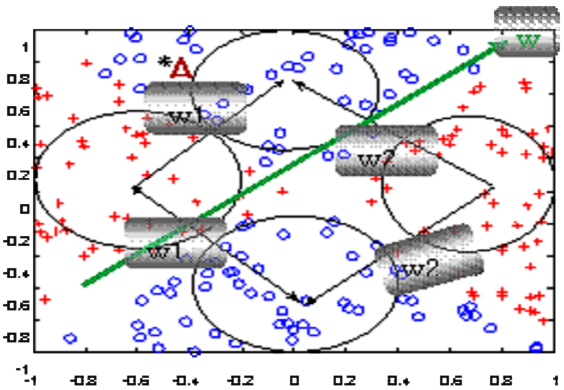

Figure1 Illustration of K-means FLD

Red "+" symbol represents class 1, meanwhile blue "o" represents class -1, when only FLD applied into this case, the green direction w is produced, which can't classify this case well. For this linear un-separable case, FLD can't deal with it efficiently, so we combine it with K-means clustering technique, and both them can cooperatively solve it well. Procedure as following:

Firstly, K-means clustering is applied into every class data, like Figure1, red class data is divided into two clusters, shown as left and right circle. Green class data: top and bottom clusters.

Secondly, fisher linear discriminant between any one red cluster and any one green cluster is obtained, shown as direction w11, w12, w21, w22.

Thirdly, given any one test data, described as point A, first of all, to determine which red cluster and green cluster it belongs to, like Figure1, point A is in left red cluster and top green cluster, then the FLD direction between these two clusters is adopted, namely direction w12.

Fourthly, only FLD direction w12 is used to determine the label for point A, ignoring other FLD directions.

III. Experimental Results and Analysis

Table1: Results comparison between more features and original KFD

Data (Misclassification rate %)	Common FLD	2-means FLD	3-means FLD	5-means FLD
Diabetis *Train: 468; Test: 300; Input:8*	23.22	22.27	21.45	20.89
Breast-cancer *Train:200; Test:77; Input:9*	27.14	25.12	26.45	27.14
Ringnorm *Train: 400; Test: 7000; Input:20*	24.95	22.34	22.16	21.19
Thyroid *Train: 140; Test: 75; Input:5*	14.67	5.8	6.73	7.47
Titanic *Train: 150; Test: 2051; Input:3*	22.50	20.73	19.38	18.86
Artificial data *Train: 200; Test: 2000; Input:4*	48.4	4.75	9.55	12.45
Waveform *Train: 400; Test: 4600; Input:21*	14.97	12.21	11.38	13.16

Table1 shows us the performance for classification between common FLD and K-means FLD, which can elevate the performance compared to common FLD algorithm

IV. Conclusion

FLD has some problem when dealing with linear un-separable classification problem, like Table1 the artificial data, only has 51.6 right classification rates. But K-means FLD combine K-means technique and FLD technique. It can deal with not only linear un-separable case, but also for separable one.

The simulation results shown as Table1 also demonstrate this. But the selection of parameter k is still a problem. Sometimes lower k is optimal, sometimes not. And our future work will focus on how to determine the optimal k.

Reference

[1] R.A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems,"Annals of Eugenics, vol. 7, part II, pp. 179-188, 1936.

[2] C. Xiang, X. A. Fan, Face Recognition Using Recursive Fisher Linear Discriminant, 2004

[3] Christopher M.Bishop, Pattern Recognition and Machine Learning, 2006 Springer Science + Business Media, LLC, 2006.

[4] S.Mika, Kernel Fisher Discriminant, 2002