



# A nested-loop Fisher discriminant analysis algorithm



Chunhui Zhao <sup>a,b,\*</sup>, Furong Gao <sup>a,c</sup>

<sup>a</sup> State Key Laboratory of Industrial Control Technology, Department of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China

<sup>b</sup> Key Laboratory of Advanced Control and Optimization for Chemical Processes, Shanghai, 200237, China

<sup>c</sup> Department of Chemical and Biomolecular Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong Special Administrative Region

## ARTICLE INFO

### Article history:

Received 27 January 2015

Received in revised form 8 June 2015

Accepted 15 June 2015

Available online 19 June 2015

### Keywords:

Fisher discriminant analysis

Nested loop

Iterative process

Orthogonal discriminant components

Data deflation

## ABSTRACT

Fisher discriminant analysis (FDA), as a very important method for feature extraction, has been widely used in different applications. However, some drawbacks of the conventional FDA algorithm have limited its success and applications. In order to improve the discriminant power, a new discriminant analysis algorithm is proposed based on Fisher's linear discriminant objective by developing a nested loop algebra, called nested-loop Fisher discriminant analysis (NeLFDA). The basic idea of the proposed NeLFDA is to overcome three important problems of the conventional Fisher discriminant analysis algorithm: (1) the within-class scatter matrix may be singular for eigenvalue decomposition, (2) the number of extracted discriminant components is limited by rank deficiency of the between-class scatter matrix, and (3) the discriminant components are correlated with each other for each class. The above problems are addressed in a nested-loop iterative process including inner-loop and outer-loop calculations. Using the proposed algorithm, its application to classification and fault diagnosis is evaluated on two examples. Illustration results show that the proposed algorithm can better separate different classes with improved discriminant power and provide more promising fault diagnosis performance.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Over the last few years, data-based statistical modeling and analysis methods have drawn much attention and found successful application in practice [1–17]. They are attractive options for handling various problems in many fields that are “data rich but information poor”. Some of the methods that have been developed recently for learning a low-dimensional feature space include principal component analysis, partial least squares, independent component analysis and canonical correlation analysis. Principal component analysis (PCA) [11,12] transforms a number of possibly correlated variables into a feature space spanned by a smaller number of uncorrelated variables known as principal components. Independent component analysis (ICA) [13] in contrast to correlation-based transformations such as PCA not only decorrelates the signals but also attempts to make the signals as independent as possible. The feature space is determined by both the second and higher order statistics of the original data. Canonical correlation analysis (CCA) [14] is a method of correlating linear relationships between two multidimensional variables. Partial least squares (PLS) [15,16] is a method similar to canonical correlation analysis but the object is usually to understand the influence of one data set on another and to develop a model for predictive purposes. Also, some noticeable extensions have been developed and reported, such as nonlinear extension of the above methods using the kernel technique [17] has been

proposed. Nonlinear multivariate analysis can be performed by replacing the inner products in the feature space with the kernel function without explicitly knowing the mapping in the feature space.

One of the main characteristics of these multivariate projection methods is their ability to handle large numbers of highly correlated variables. Of particular importance is that they are able to reduce the dimensionality of data space by projecting the measurement data into a low-dimensional space defined by a few latent variables that summarize all relevant information. The Fisher discriminant analysis (FDA) [18] is a classical method for feature extraction and dimension reduction jointly. It has been successfully applied in many classification problems such as image recognition [19,20], text classification [21,22] and fault diagnosis [23–26]. Classical FDA algorithm aims to find an optimal transformation to achieve maximum discrimination by maximizing the Fisher criterion function, i.e., minimizing the within-class distance and maximizing the between-class distance simultaneously. It is well known that this optimization problem can be solved by applying eigenvalue decomposition to the scatter matrices. It provides an optimal lower dimensional representation in terms of discrimination among classes of data. It derives a projection basis that separates different classes as far as possible and compresses the same class as compact as possible. Whereas PCA seeks directions that are efficient for representation, FDA seeks directions that are efficient for discrimination. However, there are some problems in FDA algorithm which may limit effectiveness of FDA. First, the within-class scatter matrix may be singular which thus causes problem during the eigenvalue decomposition. Besides, due to the rank deficiency of the between-class scatter

\* Corresponding author. Tel./fax: +86 571 87951879.  
E-mail address: [chhzhao@zju.edu.cn](mailto:chhzhao@zju.edu.cn) (C. Zhao).

matrix, the number of discriminant components is less than the number of classes, which thus may not be information-sufficient. Third, the extracted components are correlated with each other in each class, revealing overlapping information to a certain extent.

In order to avoid the singularity problem of within-class scatter matrix, several extensions of FDA algorithm, including pseudoinverse linear discriminant analysis (LDA) [27], regularized LDA [28], PCA + LDA [20] and LDA/QR [29], have been proposed in the past. Pseudoinverse LDA applies pseudoinverse to handle the singularity of matrices; regularized LDA adds a scaled identity matrix to the scatter matrix to make the perturbed scatter matrix positive definite. In contrast, PCA + LDA and LDA/QR are representatives of two-stage methods which apply an intermediate dimension reduction using two different algorithms on the original data to yield a column-orthogonal transformation matrix  $\mathbf{Z}$  in the first stage before implementing the classical FDA algorithm on the reduced space in the second stage. PCA + LDA uses PCA to do the transformation by maximizing the trace of the within-class scatter, which, however, is not compatible to the objective function of LDA so that some useful information that may be important for discrimination might be lost in the PCA preprocessing stage. The LDA/QR first maximizes the separation between different classes by applying QR decomposition [30] to a small size matrix so that the scatter matrices can be reduced. Although this objective function is more compatible to the objective function of FDA, it reduces the dimension to a too small dimension for many problems and thus a lot of information in the original data matrix might be missing. In general, two-stage method is the common way to solve the singularity issue, in which, the critical issue is how to implement the intermediate stage for dimension reduction before the actual FDA method works. Previous work has resulted in some undesirable problems while trying to solve the singularity issue.

In this paper, a new discriminant analysis algorithm, called nested-loop Fisher discriminant analysis (NeLFDA), is proposed to improve the discriminant power. A nested loop is formulated by constructing two different mathematical optimization problems with different objective functions to solve the drawbacks of the conventional FDA algorithm. First, it designs an inner-loop iterative process, which maximizes the between-class scatter but can prepare multiple preliminary components with number equal to the rank of the original within-class scatter matrix. The within-matrix scatter in fact converges to a diagonal matrix, which thus provides an analysis platform for further feature compression. Second, it further refines the preliminary components by maximizing the ratio of the between-class scatter to the within-class scatter to extract the final discriminant components which in fact works as a post-processing step on the inner-loop results. An outer-loop iterative process is implemented so that multiple final discriminant components can be obtained to solve the feature limitation caused by rank deficiency of the between-class scatter matrix. Besides, class-specific data deflation is performed whenever a final discriminant component is extracted which thus guarantees the orthogonality of discriminant components for each class. Analyses are conducted to further help to comprehend the underlying meaning of the proposed solution. The power of the proposed algorithm is evaluated regarding its ability of classification and fault diagnosis. In comparison with the conventional FDA algorithm, the promising results demonstrate the effectiveness of the proposed algorithm.

## 2. Preliminary

### 2.1. The conventional FDA algorithm

FDA algorithm [18] is one of the most popular dimension reduction techniques widely used in pattern classification. It searches for the projection directions on which the data points of different classes are separated as far as possible while the data points of the same class are kept to be close to each other.

Let  $\mathbf{X}_i$  be an  $N \times J$ -dimensional data matrix corresponding to the samples from Class  $i$  in which the rows are observations and the columns are variables.  $\bar{\mathbf{x}}_i$  is the  $J$ -dimensional sample mean for Class  $i$ . The within-class scatter matrix for Class  $i$  can be calculated by

$$\mathbf{S}_i = \sum_{\mathbf{x}_i \in \mathbf{X}_i} (\mathbf{x}_i - \bar{\mathbf{x}}_i)(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T. \quad (1)$$

Then the within-class-scatter matrix for all classes is,

$$\mathbf{S}_w = \sum_{i=1}^C \mathbf{S}_i \quad (2)$$

where,  $C$  is the number of classes.

The between-class scatter matrix is calculated as,

$$\mathbf{S}_b = \sum_{\mathbf{x}_i \in \mathbf{X}_i} n_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \quad (3)$$

where,  $n_i$  is the number of observations in Class  $i$ , and  $\bar{\mathbf{x}}$  is the total mean vector for all samples.

The FDA components are determined by maximizing the Fisher criterion, i.e., the ratio of the between-class scatter to the within-class scatter,

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \quad (4)$$

where the vector  $\mathbf{w}$  is the Fisher optimal discriminant direction.

It is easy to show that a vector  $\mathbf{w}$  that maximizes  $J(\mathbf{w})$  is equal to calculating the generalized eigenvectors of the eigenvalue problem:

$$\mathbf{S}_b \mathbf{w} = \lambda \mathbf{S}_w \mathbf{w} \quad (5)$$

where the eigenvalue  $\lambda$  indicates the degree of overall separability among the classes. If  $\mathbf{S}_w$  is non-singular, we can obtain a conventional eigenvalue problem by the following expression

$$\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{w} = \lambda \mathbf{w}. \quad (6)$$

A set of eigenvectors of  $\mathbf{S}_b$  and  $\mathbf{S}_w$  can be obtained one time corresponding to the  $m$  largest non-zero generalized eigenvalues. Note that there are at most  $C - 1$  non-zero generalized eigenvalues. Therefore, the number of discriminant components is subject to the number of classes.

### 2.2. Motivation

Although the conventional FDA algorithm has been widely used for different applications, the algorithm itself has some problems which should be addressed.

First, from Eq. (6), it requires that the within-class scatter matrix should be non-singular, which, however, may not always be satisfied in practice. Although there have been some solutions to handle this issue, some other problems are introduced as mentioned before. In general, the within-class scatter matrix is nonsingular in the case of a large number of samples, and singular in the case of a small number of samples [31]. Here, a large or small number of samples is a relative comparison between the number of samples and the number of variables. Second, it is clear that the number of discriminant components is subject to the rank of the between-class scatter matrix from Eq. (6) since the rank of product  $\mathbf{S}_w^{-1} \mathbf{S}_b$  is less than the rank of  $\mathbf{S}_b$ . However,  $C - 1$  components may not necessarily provide sufficient information to reveal the difference among different classes. Third, for discriminant components, it appears that the less correlative to each other the feature components are, the better the discriminant vectors are [31]. However, for the conventional FDA algorithm, the discriminant components

Download English Version:

<https://daneshyari.com/en/article/7563192>

Download Persian Version:

<https://daneshyari.com/article/7563192>

[Daneshyari.com](https://daneshyari.com)