Chemometrics and Intelligent Laboratory Systems 142 (2015) 49–53

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Is uncorrelated linear discriminant analysis really a new method?

S. Hou [a,*], C.B. Riley [b]

[a] Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, Prince Edward Island, C1A 4P3, Canada
[b] Massey University, Institute of Veterinary, Animal and Biomedical Sciences, Palmerston North 4442, New Zealand

ARTICLE INFO

Article history:
Received 11 November 2014
Received in revised form 6 January 2015
Accepted 10 January 2015
Available online 18 January 2015

Keywords:
Fisher's linear discriminant analysis
Linear discriminant analysis
Uncorrelated linear discriminant analysis
S-orthogonality

ABSTRACT

Fisher's linear discriminant analysis and linear discriminant analysis (LDA) are powerful methods in multivariate data analysis. Recently, a method called "uncorrelated linear discriminant analysis (ULDA)" has attracted attention in the chemometrics community. It has been stated that the major difference between ULDA and LDA is that the discriminant vectors of ULDA must satisfy an "S-orthogonality" constraint. This has led to the impression that ULDA is a different method from LDA. A number of papers published in the chemometrics field and others have generally accepted this statement. However, it can be shown that the so-called ULDA method is equivalent to Fisher's linear discriminant analysis or one simple case of LDA. There is a need to resolve the confusion surrounding ULDA in the chemometrics community. This work clarifies this confusion from a mathematical perspective and demonstrates equivalence using real experimental data sets.

1. Introduction

Discrimination and classification are important techniques of multivariate data analysis. These methods have been widely applied to different fields such as chemistry, metabolomics and medical science in the past several decades. Discriminant analysis is generally employed to separate samples (objects) from different groups (populations) in a low-dimensional space, while classification normally involves assigning new samples into groups with characteristics of interest. Nevertheless, the distinction between discrimination and classification is often unclear in solving real problems and the two terms are often used vaguely or interchangeably. Commonly used discrimination and classification methods in chemometrics include Fisher's linear discriminant analysis (Fisher's LDA) [1–3], linear discriminant analysis (LDA) [3,4], partial least squares discriminant analysis (PLSDA) [5], and support vector machines (SVM) [6].

Among the various discrimination and classification methods, Fisher's LDA is historically one of the earliest formally developed, and an important technique proposed by Fisher in the 1930s [1,2]. Fisher's LDA is often used interchangeably with LDA, but subtle differences exist. LDA was largely attributed to the efforts of Rao in the 1940s, and assumes that the data in each group follow a multivariate normal distribution [4]. In contrast, Fisher's LDA does not use such an explicit assumption. For LDA, the covariance matrices for different groups can be different, but Fisher's LDA implicitly assumes that all the covariance matrices are the same and a pooled covariance matrix is used. When LDA assumes that different groups have equal population covariance matrices, misclassification costs, and prior probabilities, this simple case of LDA becomes equivalent to Fisher's LDA [3]. As Fisher's LDA is equivalent to one simple case of LDA, Fisher's LDA and this equivalent simple case of LDA will not be distinguished and are denoted as Fisher's LDA in the remainder of this article unless clearly specified otherwise.

In the recent years, a method entitled "uncorrelated linear discriminant analysis (ULDA)" has attracted attention in the chemometrics community. It has been stated that ULDA is a "powerful tool" and "the major difference between ULDA and LDA is that the vectors in the transformation matrix obtained by ULDA have to satisfy the constraint of so-called 'S-orthogonality'" [7]. A number of papers that have been published in Chemometrics and Intelligent Laboratory System [8–10] and other journals [11–14] have followed or reiterated this statement. In particular, ULDA has been claimed to be a new method in published references [12,14]. This has given readers, especially those without strong background in mathematics, the impression that the results of LDA do not have properties such as S-orthogonality, and that ULDA is a method different from LDA. The authors of the current article posit that this is actually not true. It can be shown mathematically that ULDA is equivalent to Fisher's LDA (or one simple case of LDA). Obviously, a confusion exists about the novelty of ULDA in the chemometrics community. The work reported here aims to clarify this confusion with respect to ULDA. Mathematical proof is provided and real experimental data sets are used to demonstrate the equivalence.

In this article, a scalar is designated with an italic non-bold letter. A vector is represented by a lower case bold letter. A row vector is always written as the transpose of a column vector. The transpose operator is

Abbreviations: GSVD, generalized singular value decomposition; LDA, linear discriminant analysis; PLSDA, partial least squares discriminant analysis; SVM, support vector machines; ULDA, uncorrelated linear discriminant analysis

* Corresponding author. Tel.: +1 902 566 6080.
E-mail address: shou@upei.ca (S. Hou).

http://dx.doi.org/10.1016/j.chemolab.2015.01.005
0169-7439/

signified with the superscript "T". A data matrix is denoted by an upper case bold letter.

2. Mathematical aspect

2.1. Fisher's linear discriminant analysis

Supposing that a data matrix **X** contains measurements of samples (objects) from g groups and each sample is measured on p variables, the entire data can be arranged in matrix form:

$$\mathbf{X} = \begin{bmatrix} x_{11}^1 & \cdots & x_{1p}^1 \\ \vdots & \ddots & \vdots \\ x_{N_1 1}^1 & \cdots & x_{1p}^1 \\ \hline \cdot & \cdots & \cdot \\ \vdots & \ddots & \vdots \\ \cdot & \cdots & \cdot \\ \hline x_{11}^g & \cdots & x_{1p}^g \\ \vdots & \ddots & \vdots \\ x_{N_g 1}^g & \cdots & x_{1p}^g \end{bmatrix} = \begin{bmatrix} \left(\mathbf{x}_1^1\right)^{\mathrm{T}} \\ \vdots \\ \left(\mathbf{x}_{N_1}^1\right)^{\mathrm{T}} \\ \hline \vdots \\ \vdots \\ \vdots \\ \hline \left(\mathbf{x}_1^g\right)^{\mathrm{T}} \\ \vdots \\ \left(\mathbf{x}_{N_g}^g\right)^{\mathrm{T}} \end{bmatrix} \tag{1}$$

where the numbers of samples from different groups are denoted by N_1, N_2, …, Ng, respectively. Note that the superscript has been used to indicate the sample group information. The measurements of different variables for each sample can also be denoted in a vector form, as shown in Eq. (1). The between-group scatter matrix and the pooled within-group scatter matrix can be calculated as

$$\mathbf{S}_{\mathrm{b}} = \sum_{i=1}^{g} N_k \left(\bar{\mathbf{x}}^k - \bar{\mathbf{x}}\right)\left(\bar{\mathbf{x}}^k - \bar{\mathbf{x}}\right)^{\mathrm{T}} \text{ and} \tag{2}$$

$$\mathbf{S}_{\mathrm{w}} = \sum_{k=1}^{g} \sum_{i=1}^{N_k} \left(\mathbf{x}_i^k - \bar{\mathbf{x}}^k\right)\left(\mathbf{x}_i^k - \bar{\mathbf{x}}^k\right)^{\mathrm{T}} \tag{3}$$

respectively, where $\bar{\mathbf{x}}^k$ is the sample mean vector for the kth group, and $\bar{\mathbf{x}}$ denotes the overall mean vector for the entire data. Fisher's LDA looks for a discriminant vector **v** that maximizes the ratio

$$F = \frac{\mathbf{v}^{\mathrm{T}}\mathbf{S}_{\mathrm{b}}\mathbf{v}}{\mathbf{v}^{\mathrm{T}}\mathbf{S}_{\mathrm{w}}\mathbf{v}}, \tag{4}$$

which is referred to as Fisher's criterion. If $\lambda_1, \lambda_2, \cdots, \lambda_s > 0$ denote the s non-zero eigenvalues of $\mathbf{S}_{\mathrm{w}}^{-1}\mathbf{S}_{\mathrm{b}}$, it is shown that the solution to **v** is given as one of the corresponding eigenvectors of $\mathbf{S}_{\mathrm{w}}^{-1}\mathbf{S}_{\mathrm{b}}$. Specifically, the new latent variable as a result of the original data projected on the eigenvector corresponding to the first eigenvalue is called the first discriminant, and the second new latent variable resulted from the original data projected on the eigenvector corresponding to the second eigenvalue is termed as the second discriminant. Continuing in this way, the jth discriminant is obtained. It should be mentioned that the maximum number of discriminant vectors cannot be larger than the number of groups minus one ($s \le g - 1$), and the number of variables ($s \le p$) [3].

It should be mentioned that there are several variants to Fisher's criterion [15,16]. One of the most important variants is to maximize

$$F = \mathrm{tr}\left[\left(\mathbf{V}^{\mathrm{T}}\mathbf{S}_{\mathrm{w}}\mathbf{V}\right)^{-1}\left(\mathbf{V}^{\mathrm{T}}\mathbf{S}_{\mathrm{b}}\mathbf{V}\right)\right], \tag{5}$$

where **V** is a matrix with $\mathbf{V} = \left[\mathbf{v}_1, \cdots, \mathbf{v}_{g-1}\right]$. This generally happens when more than two groups are involved and discriminant analysis looks for a subspace denoted by **V** that maximizes Fisher's criterion in Eq. (5). Setting the derivative matrix of F with respect to **V** to zero, followed by arrangement and simplification yields

$$\mathbf{S}_{\mathrm{b}}\mathbf{V} = \mathbf{S}_{\mathrm{w}}\mathbf{V}\left[\left(\mathbf{V}^{\mathrm{T}}\mathbf{S}_{\mathrm{w}}\mathbf{V}\right)^{-1}\left(\mathbf{V}^{\mathrm{T}}\mathbf{S}_{\mathrm{b}}\mathbf{V}\right)\right]. \tag{6}$$

It can be seen that any basis of the subspace spanned by the $g - 1$ eigenvectors of $\mathbf{S}_{\mathrm{w}}^{-1}\mathbf{S}_{\mathrm{b}}$ corresponding to the $g - 1$ largest eigenvalues meets this condition. Particularly, if the columns of **V** are chosen as the $g - 1$ eigenvectors, the solution to the maximization problem in Eq. (5) will be the same as that to Eq. (4). This is to say that if searching for the discriminant vectors in a stepwise manner based on Eq. (4), the solutions will be nested in the solution to Eq. (5) where a set of discriminant vectors are searched simultaneously. Thus, the distinction between the problems in Eqs. (4) and (5) disappears in this sense.

2.2. Uncorrelated linear discriminant analysis

The term "uncorrelated linear discriminant analysis" first appeared in a paper in the field of face recognition [17]. In that study, Fisher's criterion shown in Eq. (4) was still used (the same as Fisher's LDA), but when searching for more than one discriminant vector, a constraint was imposed as follows:

$$\mathbf{v}_i^{\mathrm{T}}\mathbf{S}_{\mathrm{t}}\mathbf{v}_j = 0, \tag{7}$$

where $\mathbf{v}_i$ and $\mathbf{v}_j$ denote any two distinct discriminant vectors and $\mathbf{S}_{\mathrm{t}}$ signifies the total scatter matrix defined as

$$\mathbf{S}_{\mathrm{t}} = \sum_{k=1}^{g} \sum_{i=1}^{N_k} \left(\mathbf{x}_i^k - \bar{\mathbf{x}}\right)\left(\mathbf{x}_i^k - \bar{\mathbf{x}}\right)^{\mathrm{T}} = \mathbf{S}_{\mathrm{w}} + \mathbf{S}_{\mathrm{b}}. \tag{8}$$

The constraint in Eq. (7) is the S-orthogonality and this has been reiterated in the references [8–14]. Based on this constraint, the latent variables obtained by projecting the original data to the discriminant vectors are uncorrelated, but the discriminant vectors are generally not orthogonal (correlated). It should be mentioned that if the discriminant vectors are forced to be orthogonal (hierarchically maximizing Fisher's criterion), it is identical to the Foley–Sammon discriminant transformation [18].

2.3. Equivalence of Fisher's LDA and ULDA

At first glance, ULDA imposes an additional S-orthogonality constraint on Fisher's criterion and it is anticipated to produce a definition different from that of Fisher's LDA. However, this is not true. Since the discriminant vectors in Fisher's LDA are taken as the eigenvectors of $\mathbf{S}_{\mathrm{w}}^{-1}\mathbf{S}_{\mathrm{b}}$, then for any discriminant vector **v**,

$$\mathbf{S}_{\mathrm{w}}^{-1}\mathbf{S}_{\mathrm{b}}\mathbf{v} = \lambda\mathbf{v} \tag{9}$$