EIGENVECTOR RESEARCH INCORPORATED
Research, Training, and Software

196 Hyacinth Road
Manson, WA 98831
www.Eigenvector.com

Using Partial Least Squares Discriminant Analysis (PLSDA) for Linear Discriminant Analysis (LDA) Part II: Theory

Neal B. Gallagher

2023

Key words: PLSDA, LDA, class-centroid centering, generalized weighting

Introduction: The similarity between partial least squares discriminant analysis (PLSDA), Fisher's linear discriminant analysis (LDA) and a number of multivariate analysis methods has been discussed previously.[1,2] It is shown here that using data preprocessing with class-centroid centering followed by generalized weighting converts PLSDA to LDA for then general case of unbalanced designs.

Data Arrangement and Intra-Class Variance for the Classification Problem: In the classification problem, it is assumed that a set of measurements are made on N variables for J different sets of samples that belong to different classes. The data for a single subset can be arranged in a $M_j \times N$ data matrix $\mathbf{X}_j$ with M_j samples and N variables. The mean for the j^{th} class $\bar{\mathbf{x}}_j$ is given by

$$\bar{\mathbf{x}}_j = \mathbf{X}_j^{\text{T}} \mathbf{1}_{M_j} \frac{1}{M_j} \tag{1}$$

where $\mathbf{1}_{M_j}$ is a $M_j \times 1$ column vector of ones. The term $\mathbf{X}_j^{\text{T}} \mathbf{1}_{M_j}$ corresponds to a summation and the $\frac{1}{M_j}$ term divides the sum by the number of samples in the class to provide the class mean. The measurements can be collected into a single $M \times N$ matrix $\mathbf{X}$ as shown in Equation 2 where $M = \sum_{j=1}^{J} M_j$. Equation 3 defines $\mathbf{Y}$ as a $M \times J$ set of "dummy" variables with M_j^{-1} in each row for the j^{th} class in the corresponding j^{th} column.

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_J \end{bmatrix} \tag{2}$$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{1}_{M_1} M_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{M_2} M_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{M_J} M_J^{-1} \end{bmatrix} \tag{3}$$

Typically, each class contains multiple measurements resulting in a data set with $M > J$. The different entries in $\mathbf{Y}$ account for the potential of an unbalanced design. The intra-class variance, $\mathbf{\Sigma}_{\text{intra}}$, is calculated using:

$$\mathbf{\Sigma}_{\text{intra}} = \frac{1}{M} \sum_{j=1}^{J} M_j \left(\mathbf{X}_{\text{j}} - \mathbf{1}_{M_j} \bar{\mathbf{x}}_j^{\text{T}}\right)^{\mathbf{T}} \left(\mathbf{X}_{\text{j}} - \mathbf{1}_{M_j} \bar{\mathbf{x}}_j^{\text{T}}\right). \tag{4}$$

Equation 4 is a pooled covariance that assumes the intra-class covariance is the same for all classes.

In the special case of a balanced design, all the M_j are equal and the data mean, $\bar{\mathbf{x}} = \mathbf{X}^{\text{T}} \mathbf{1}_M \frac{1}{M}$, corresponds to the model origin. However, in the more general case where the M_j are not equal it is, the class centroid $\bar{\mathbf{x}}_{\text{cent}} = \mathbf{X}^{\text{T}} \mathbf{Y} \mathbf{1}_J \frac{1}{J}$, that corresponds to the model origin.

Definition of Inter-Class Variance: Centering the original data to the class centroid is given by:

$$\mathbf{X} - \mathbf{1}_M \bar{\mathbf{x}}_{\text{cent}}^{\text{T}} = \mathbf{X} - \tfrac{1}{J} \mathbf{1}_{\text{M}} \mathbf{1}_J^{\text{T}} \mathbf{Y}^{\text{T}} \mathbf{X} =$$
$$= \left(\mathbf{I} - \tfrac{1}{J} \mathbf{1}_M \mathbf{1}_M^{\text{T}} \mathbf{C}_{\text{mns}}\right) \mathbf{X}. \tag{5}$$

where it is recognized that $\mathbf{1}_J^{\text{T}} \mathbf{Y}^{\text{T}} = \mathbf{1}_M^{\text{T}} \mathbf{C}_{\text{mns}}$ and

$$\mathbf{C}_{\text{mns}} = \begin{bmatrix} \mathbf{I}_{M_1} M_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M_2} M_2^{-1} & & \mathbf{0} \\ & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{M_J} M_J^{-1} \end{bmatrix}. \tag{6}$$

The term $\left(\mathbf{I} - \frac{1}{J} \mathbf{1}_M \mathbf{1}_M^{\text{T}} \mathbf{C}_{\text{mns}}\right)$ can be considered a centroid-centering operator where $\mathbf{X}_{\text{C}} = \left(\mathbf{I} - \frac{1}{J} \mathbf{1}_M \mathbf{1}_M^{\text{T}} \mathbf{C}_{\text{mns}}\right) \mathbf{X}$ is the data matrix centered to the class centroid (i.e., class-centroid preprocessing). The term $\mathbf{X}_{\text{C}}^{\text{T}} \mathbf{Y}$ is the class means (mean of each class) centered to the class centroid so that the definition of the inter-class variance $\mathbf{\Sigma}_{\text{inter}}$ can be given by

$$\mathbf{\Sigma}_{\text{inter}} = \frac{1}{J-1} \mathbf{X}_{\text{C}}^{\text{T}} \mathbf{Y} \mathbf{Y}^{\text{T}} \mathbf{X}_{\text{C}}. \tag{7}$$

Linear Discriminant Analysis (LDA) via a Generalized Eigenvalue Problem: In LDA the objective is to find a set of discriminators, $\mathbf{w}$, that maximizes the separation between the classes relative to within the classes. It is proposed that the scores are given by

$$\mathbf{t} = \left(\mathbf{I} - \tfrac{1}{J} \mathbf{1}_M \mathbf{1}_M^{\text{T}} \mathbf{C}_{\text{mns}}\right) \mathbf{X} \mathbf{w} = \mathbf{X}_{\text{C}} \mathbf{w} \tag{8}$$

and the LDA objective is formalized as

$$\max_{\mathbf{w}} \frac{\mathbf{w}^{\text{T}} \mathbf{\Sigma}_{\text{inter}} \mathbf{w}}{\mathbf{w}^{\text{T}} \mathbf{\Sigma}_{\text{intra}} \mathbf{w}}, \text{ where } \lambda = \frac{\mathbf{w}^{\text{T}} \mathbf{\Sigma}_{\text{inter}} \mathbf{w}}{\mathbf{w}^{\text{T}} \mathbf{\Sigma}_{\text{intra}} \mathbf{w}} \tag{9}$$

196 Hyacinth Road
Manson, WA 98831
www.Eigenvector.com

where the Rayleigh coefficient, λ, is a scalar. Taking the derivative of Equation 9, setting to $\mathbf{0}$ and rearranging [e.g., see 3] yields

$$\boldsymbol{\Sigma}_{\text{inter}}\mathbf{w} = \lambda\boldsymbol{\Sigma}_{\text{intra}}\mathbf{w}. \tag{10}$$

Equation 10 is a generalized eigenvalue problem (GEP) with eigenvalue λ and eigenvector $\mathbf{w}$. The GEP can be solved using a Cholesky factorization or a QZ algorithm [4,5] but $\boldsymbol{\Sigma}_{\text{intra}}$ needs to be full rank. A general approach to handling ill-conditioning and rank deficiency is regularization. Two often used regularization procedures include ridging $\boldsymbol{\Sigma}_{\text{intra}} \to \boldsymbol{\Sigma}_{\text{intra}} + \theta\mathbf{I}$ where θ is a small scalar, and generalized weighting[6] $\boldsymbol{\Sigma}_{\text{intra}} \to \boldsymbol{\Sigma}_{\text{intra}} + \mathbf{A}$ where $\mathbf{A}$ is a full rank matrix.

Inspection of the $\boldsymbol{\Sigma}_{\text{inter}}$ term in Equation 7 shows a multiplication by $\mathbf{Y}$ that is rank J (because $M > J$). Therefore, it is expected that a typical GEP will include J or potentially $J-1$ eigenvectors due to class centroid centering.

Linear Discriminant Analysis (LDA) via a Symmetric Eigenvalue Problem:

Defining $\mathbf{p} = \boldsymbol{\Sigma}_{\text{intra}}^{1/2}\mathbf{w}$ allows for a variable transform of the GEP to a symmetric eigenvalue problem (SEP). Substitution of the definition into Equation 10 gives the SEP as [e.g., see 3]

$$\boldsymbol{\Sigma}_{\text{intra}}^{-1/2}\boldsymbol{\Sigma}_{\text{inter}}\boldsymbol{\Sigma}_{\text{intra}}^{-1/2}\mathbf{p} = \lambda\mathbf{p}. \tag{11}$$

where it is assumed that $\boldsymbol{\Sigma}_{\text{intra}}$ is full rank or has been regularized. It should be clear that the leading constant $\frac{1}{J}$ in $\boldsymbol{\Sigma}_{\text{inter}}$ only scales the eigenvalue and can be ignored in Equations 10 and 11 without influencing the calculation of $\mathbf{w}$ and $\mathbf{p}$. With this consideration, Equation 7 can be substituted into Equation 11 to give

$$\boldsymbol{\Sigma}_{\text{intra}}^{-1/2}\mathbf{X}_{\text{C}}^{\text{T}}\mathbf{Y}\mathbf{Y}^{\text{T}}\mathbf{X}_{\text{C}}\boldsymbol{\Sigma}_{\text{intra}}^{-1/2}\mathbf{p} = \lambda\mathbf{p}. \tag{12}$$

Because Equation 12 is an SEP, it can be considered principal components analysis (PCA) on the matrix $\mathbf{Y}^{\text{T}}\mathbf{X}_{\text{C}}\boldsymbol{\Sigma}_{\text{intra}}^{-1/2}$ or it can be considered a PLS model between $\mathbf{X}_{\text{C}}\boldsymbol{\Sigma}_{\text{intra}}^{-1/2}$ and $\mathbf{Y}$ i.e., PLSDA. The term $\mathbf{X}_{\text{C}}\boldsymbol{\Sigma}_{\text{intra}}^{-1/2}$ is the original data centered to the class centroid weighted by the inverse square root of the regularized intra-class covariance $\boldsymbol{\Sigma}_{\text{intra}}^{-1/2}$ (i.e., GLSW preprocessing [6,7]). Using the definition for the variable transform, $\mathbf{w} = \boldsymbol{\Sigma}_{\text{intra}}^{-1/2}\mathbf{p}$, the scores in Equation 8 become

$$\mathbf{t} = \mathbf{X}_{\text{C}}\boldsymbol{\Sigma}_{\text{intra}}^{-1/2}\mathbf{p} \tag{16}$$

Therefore, class centroid centering and generalized weighting preprocessing converts PLSDA to LDA.

In the special case of a balanced classification problem, then all the classes have the same number of samples so that $\left(\mathbf{I} - \frac{1}{J}\mathbf{1}_M\mathbf{1}_M^{\text{T}}\mathbf{C}_{\text{mns}}\right) \to \left(\mathbf{I} - \frac{1}{M}\mathbf{1}_M\mathbf{1}_M^{\text{T}}\right)$ and $\bar{\mathbf{x}}_{\text{cent}} \to \bar{\mathbf{x}}$. Also, in the absence of collinearity in the intra-class variance and all the variables have equal variances, σ^2, then $\boldsymbol{\Sigma}_{\text{intra}} \to \sigma^2\mathbf{I}$. When these two criteria are met, then PLSDA with mean-centering also becomes LDA.

Conclusion and Discussion: The use of class centroid centering and generalized weighting preprocessing converts partial least squares discriminant analysis (PLSDA) to linear discriminant analysis (LDA). For many problems the intra-class covariance matrix, $\boldsymbol{\Sigma}_{\text{intra}}$, may be ill-conditioned or rank deficient and require regularization for the DA models to be identifiable. Differences in results can be observed based on the regularization procedures used and the matrix decomposition method. For example, the SEP can be based on a stable and accurate singular value decomposition, and the GEP often uses an eigenvalue decomposition algorithm (e.g., QZ). Although the factors may be different depending on the decomposition algorithm (e.g., PCA, PLS or QZ), they are expected to span a similar subspace and provide similar class predictions.

References:

[1] Barker, M, Rayens, W, "Partial Least Squares for Discrimination," *J. Chemom.*, **17**, 166-173 (2003). doi: 10.1002/cem.785.
[2] Næs, T, Indahl, U., "Unified Description of Classical Classification Methods for Multicollinear Data," *J. Chemom.*, **12**, 205-220 (1998).
[3] Gallagher, NB, "A Comparison of Common Factor-Based Methods for Hyperspectral Image Exploration – PCA, MAF, MNF and MDF", *J. Spec. Imag.*, **11**, Article ID a 6 (2022). doi: 10.1255/jsi.2022.a6
[4] MATLAB Version 9.14 (2023a), Natick, MA USA, The MathWorks Inc. (see the "eig" function).
[5] Moler, CB, Stewart, GW, "An Algorithm for Generalized Matrix Eigenvalue Problems", *SIAM J. Numer. Anal.*, **10**, 2, 241-256 (1973). doi: 10.1137/0710024.
[6] Martens, H, Høy, M, Wise, BM, Bro, R, Brockhoff, PB, "Pre-Whitening of Data by Covariance-Weighted Pre-Processing," *J. Chemom.*, **17**(3), 153-165 (2003). doi: 10.1002/cem.780.
[7] PLS_Toolbox Version 9.2.1, Manson, WA USA, Eigenvector Research, Inc.