1 Discriminant Analysis

Discriminant analysis is essentially "supervised machine learning" applied to "classification" problems. One can also view discriminant analysis as a form of regression analysis in which the criterion variables are discrete.

In a typical discriminant analysis problem,

- Given:
  - A finite collection of observation units, enumerated by $\{ 1, 2, \ldots, N \}$.
  - For each observation unit, i, measurements $\mathbf{v}_i^T = (v_{i1}, v_{i2}, \ldots, v_{in}) \in \mathbb{R}^{1\times n}$on $n \in \mathbb{N}$ of the observation unit on n predictor variables $V_1, V_2, \ldots, V_n$. Thus, $\{\mathbf{v}_i\}_{i=1}^N \subset \mathbb{R}^n$ is a set of N points in $\mathbb{R}^n = \mathbb{R}^{n\times 1}$.
  - A classification of the N observational units into g groups.
- Want to determine:
  - A linear function $L = \beta_1 V_1 + \beta_2 V_2 + \cdots + \beta_n V_n$ that can be used to "predict" the given classification of observational units. L is called a *discriminant function*.

2 Principal Component Analysis

In a typical discriminant analysis problem,

- Given:
  - A finite collection of observation units, enumerated by $\{ 1, 2, \ldots, N \}$.
  - For each observation unit, i, measurements $\mathbf{v}_i^T = (x_{i1}, x_{i2}, \ldots, x_{in}) \in \mathbb{R}^{1\times n}$ of the observation unit on $n \in \mathbb{N}$ predictor variables $X_1, X_2, \ldots, X_n$. Thus, $\{\mathbf{v}_i\}_{i=1}^N \subset \mathbb{R}^n$ is a set of N points in $\mathbb{R}^n$.
- Want to determine:
  - A linear transformation $(Y_1, Y_2, \ldots, Y_n) = \mathcal{L}(X_1, X_2, \ldots, X_n)$ such that the correlation matrix with respect to the coordinates $(Y_1, \ldots, Y_n)$ is a diagonal matrix with non-negative and non-decreasing entries. The new coordinates $Y_1, \ldots, Y_n$ are called the *principal components*.

For each $k = 1, 2, \ldots, n$,

$$\mathbf{x}_k \;=\; \begin{pmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ \vdots \\ x_{Nk} \end{pmatrix}.$$

Correlation matrix: For $k, l = 1, 2, \ldots, n$,

$$R_{kl} \;:=\; \mathrm{Cor}(\, \mathbf{x}_k \,,\, \mathbf{x}_l \,) \;:=\; \frac{\sum_{i=1}^N (x_{ik} - \overline{\mathbf{x}}_k)(x_{il} - \overline{\mathbf{x}}_l)}{(N-1)\,\mathrm{SE}(\mathbf{x}_k)\mathrm{SE}(\mathbf{x}_l)} \;=\; \frac{\sum_{i=1}^N (x_{ik} - \overline{\mathbf{x}}_k)(x_{il} - \overline{\mathbf{x}}_l)}{\sqrt{\sum_{i=1}^N (x_{ik} - \overline{\mathbf{x}}_k)^2}\sqrt{\sum_{i=1}^N (x_{il} - \overline{\mathbf{x}}_l)^2}},$$

where

$$\overline{\mathbf{x}}_k \;=\; \frac{1}{N}\sum_{i=1}^N x_{ik}\,, \quad \text{and} \quad \mathrm{SE}(\mathbf{x}_k) \;:=\; \sqrt{\frac{1}{N-1}\sum_{i=1}^N (x_{ik} - \overline{\mathbf{x}}_k)^2}.$$

$R := (\, R_{kl} \,)_{k,l=1,\ldots,n} \in \mathbb{R}^{n\times n}$ is a real symmetric matrix. Note that the diagonal entries of R are all 1:

$$R_{kk} \;=\; \frac{\sum_{i=1}^{N}(x_{ik}-\overline{\mathbf{x}}_k)(x_{ik}-\overline{\mathbf{x}}_k)}{\sqrt{\sum_{i=1}^{N}(x_{ik}-\overline{\mathbf{x}}_k)^2}\sqrt{\sum_{i=1}^{N}(x_{ik}-\overline{\mathbf{x}}_k)^2}} \;=\; 1,$$

which furthermore implies that $\mathrm{trace}(R) = n$. Now, recall:

Theorem 2.1 *A real square matrix is symmetric if and only if it has an orthonormal basis of eigenvectors.*

Theorem 2.2 *Every real symmetric matrix $A \in \mathbb{R}^{n\times n}$ is orthogonally diagonalizable, i.e. there exists an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ such that*

$$D \;:=\; Q^T A\, Q$$

is a diagonal matrix. Moreover, the set of eigenvalues of A is equal to the set of diagonal entries of D. The i^{th} column of Q is an eigenvector of A corresponding to the i^{th} diagonal entry of D.

Let D be the diagonalization of R which has non-decreasing diagonal entries. Note that $\mathrm{trace}(D) = \mathrm{trace}(R) = n$. Recall also that the diagonal entries of D are necessarily non-negative and they are the eigenvalues of R. Consequently, the k^{th} diagonal entry $\lambda_k \geq 0$ of D (or eigenvalue of R) can be interpreted as the number of dimension(s) "explained" by the corresponding principal component $\mathrm{span}(Y_k)$.

3 Factor Analysis