MATH1309 – Practice Problems 9

This week we will explore the solutions to Examples 3 to 5 in the lecture notes using the procedure PROC DISCRIM

Example 3: Let $\boldsymbol{X}^T = (X_1, X_2, X_3)$ be a random vector representing important characteristics to distinguish between genuine and forged bank notes. A random sample of 50 genuine bank notes gives the mean $\overline{\boldsymbol{x}}_1^T = (2.1, 5.3, 4.0)$ and covariance matrix

$$\mathcal{S}_1 = \begin{pmatrix} 3.1 & 2.2 & 5.1 \\ 2.2 & 4.1 & 2.4 \\ 5.1 & 2.4 & 15.1 \end{pmatrix}.$$

Also the mean and covariance matrix of a random sample of 26 forged bank notes are as follows:

$$\overline{\boldsymbol{x}}_2 = \begin{pmatrix} 8.0 \\ 10.1 \\ 5.0 \end{pmatrix} \quad \text{and} \quad \mathcal{S}_2 = \begin{pmatrix} 2.9 & 2.8 & 5.1 \\ 2.8 & 4.0 & 2.6 \\ 5.1 & 2.6 & 14.9 \end{pmatrix}$$

(a) Identify the following two suspected bank notes as genuine or forged bank notes using Linear discriminant function.

$$\text{Bank note 1: } = \begin{pmatrix} 6.0 \\ 9.0 \\ 4.1 \end{pmatrix} \quad \text{and} \quad \text{Bank note 2: } = \begin{pmatrix} 2.1 \\ 4.9 \\ 4.9 \end{pmatrix}.$$

(b) List the assumptions you used for the above analysis.

Example 4: Allocate the following observations, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to most suitable population among $\Pi_1 : N_2(\boldsymbol{\mu}_1, \Sigma_1)$ and $\Pi_2 : N_2(\boldsymbol{\mu}_2, \Sigma_2)$, where

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \boldsymbol{\mu}_2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \ \Sigma_1 = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix} \text{ and } \Sigma_2 = \begin{pmatrix} 4 & -2 \\ -2 & 16 \end{pmatrix}.$$

Observations are:

$$\boldsymbol{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \boldsymbol{x}_2 = \begin{pmatrix} 2 \\ -3 \end{pmatrix}.$$

Assume misclassification costs, $c(2|1) = 2c(1|2)$ and, prior probabilities $p_1 = 0.25$ and $p_2 = 0.75$.

Example 5: Let $\boldsymbol{X}^T = (X_1, X_2)$ be a random vector representing important characteristics to distinguish between two normal populations Π_1 and Π_2. A random sample of 10 observations from Π_1, gives the mean $\overline{\boldsymbol{x}}_1^T = (-1, 3)$ and the sample covariance matrix

$$\mathcal{S}_1 = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}.$$

Also the mean and covariance matrix of a random sample of 15 from Π_2 are as follows:

$$\overline{\boldsymbol{x}}_2 = \begin{pmatrix} 0 \\ -2 \end{pmatrix} \text{ and } \mathcal{S}_2 = \begin{pmatrix} 4 & 1 \\ 1 & 9 \end{pmatrix}$$

Given the prior probability $p_1 = 0.4$, identify the following two observations assuming equal misclassification costs.

Observations are: $\boldsymbol{x}_1^T = (0.5, 1)$ and $\boldsymbol{x}_2^T = (-1, -3)$.

From a sample data set contained in Example15.dat on Canvas, perform a discriminant analysis. This file contains undergraduate grade point average (gpa) and graduate management aptitude test (gmat) results for two different groups (populations) Π1 and Π2.

(a) Use proc discrim to create the groups.
(b) Use the results in the output to determine the Apparent Error Rate and the Estimate of the Actual Error Rate.