DISCRIMINANT ANALYSIS

1 Introduction

Reference: Chapter 11, Johnson and Wichern

Discriminant analysis is also called *classification analysis*. Suppose we have number of observations from several populations and we have a new observation that is known to come from one of these populations. If the population of the new observation is unknown, the rules in discriminant analysis will enable to identify the most likely population for the new object.

2 Classification for Two Populations

Consider two p-dimensional multivariate populations as follows:

$$\text{Population 1: } \Pi_1 \text{ with pdf } f_1(\boldsymbol{x})$$
$$\text{Population 2: } \Pi_2 \text{ with pdf } f_2(\boldsymbol{x}).$$

Suppose a new observation vector $\boldsymbol{x}_0$ is known to come from either Π_1 or Π_2, we need a rule to classify $\boldsymbol{x}_0$ into population Π_1 or Π_2.

The cost of misclassification of $\boldsymbol{x}_0$ can be defined by the following table:

		Classified as:	
		Π_1	Π_2
True	Π_1	0	$c(2\|1)$
population	Π_2	$c(1\|2)$	0

where $c(j|i)$ is the cost of incorrectly classifying $\boldsymbol{x}_0$ as Π_j when it is from Π_i $(i \neq j)$ for $i, j = 1, 2$. Note that the cost of correct classification is 0.

Let us assume $p(j|i)$ be the conditional probability of incorrectly classifying $\boldsymbol{x}_0$ as Π_j when it is from Π_i $(i, j = 1, 2)$ and p_i be the prior probability of

$\boldsymbol{x}_0$ from population Π_i for $i = 1, 2$ such that $p_1 + p_2 = 1$. The expected cost of misclassification (ECM) of $\boldsymbol{x}_0$ is given by

$$\text{ECM} = c(2|1)p(2|1)p_1 + c(1|2)p(1|2)p_2.$$

The rule that minimizes the ECM are as follows:

$$\begin{aligned} &\text{Allocate } \boldsymbol{x}_0 \text{ to } \Pi_1 \text{ if } \frac{f_1(\boldsymbol{x}_0)}{f_2(\boldsymbol{x}_0)} \geq \frac{c(1|2)}{c(2|1)} \frac{p_2}{p_1} \\ &\text{otherwise allocate } \boldsymbol{x}_0 \text{ to } \Pi_2. \end{aligned} \tag{1}$$

Special Case:

In general, $c(j|i)$ and p_i's are unknown, however we can assume that the misclassification cost and also the prior probabilities are equal. That is, $c(2|1) = c(1|2)$ and $p_1 = p_2$ then the above rule becomes:

$$\text{Allocate } \boldsymbol{x}_0 \text{ to } \Pi_1 \text{ if } f_1(\boldsymbol{x}_0) \geq f_2(\boldsymbol{x}_0), \text{ otherwise allocate } \boldsymbol{x}_0 \text{ to } \Pi_2.$$

Note that the likelihood rule can be applied to non-normal populations.

Example 1: Allocate the following observations, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to most suitable exponential population among Π_1: Exp(λ_1) and Π_2: Exp(λ_2), where

$$\lambda_1 = 2, \quad \text{and } \lambda_2 = 1$$

Observations are:

$$x_1 = 2.0 \quad \text{and} \quad x_2 = 2.5.$$

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

3 Classification for Two Normal Populations When $\Sigma_1 = \Sigma_2 = \Sigma$

Consider two multivariate normal populations $\Pi_1 : N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\Pi_2 : N_p(\boldsymbol{\mu}_2, \Sigma)$. That is

$$f_i(\boldsymbol{x}) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i) \right\}, \ i = 1, 2$$

and

$$\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} = \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_1)^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_1) + \frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_2)^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_2) \right\}.$$

Now using the allocation rule given in (1):

Allocate $\boldsymbol{x}_0$ to Π_1 if

$$\exp\left\{ -\tfrac{1}{2} (\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T \Sigma^{-1} (\boldsymbol{x}_0 - \boldsymbol{\mu}_1) + \tfrac{1}{2} (\boldsymbol{x}_0 - \boldsymbol{\mu}_2)^T \Sigma^{-1} (\boldsymbol{x}_0 - \boldsymbol{\mu}_2) \right\} \geq \frac{c(1|2)}{c(2|1)} \frac{p_2}{p_1}$$

otherwise allocate $\boldsymbol{x}_0$ to Π_2.

Equivalently we can write the allocate rule as:

Allocate $\boldsymbol{x}_0$ to Π_1 if

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \Sigma^{-1} \boldsymbol{x}_0 - \tfrac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \Sigma^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq \ln\left[\frac{c(1|2)}{c(2|1)} \frac{p_2}{p_1} \right]$$

otherwise allocate $\boldsymbol{x}_0$ to Π_2.

Let $\boldsymbol{b} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ and $k = \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \Sigma^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) + \ln\left[\frac{c(1|2)}{c(2|1)} \frac{p_2}{p_1} \right]$. Now we can write the above rule as:

$$\begin{aligned} &\text{Allocate } \boldsymbol{x}_0 \text{ to } \Pi_1 \text{ if } \ \boldsymbol{b}^T \boldsymbol{x}_0 - k \geq 0, \\ &\text{otherwise allocate } \boldsymbol{x}_0 \text{ to } \Pi_2. \end{aligned} \tag{2}$$

This is also called *Fisher's linear discriminant rule.* The function $\boldsymbol{b}^T \boldsymbol{x}$ is called the *linear discriminant function* of $\boldsymbol{x}$.

Special Case: For equal misclassification costs and equal prior probabilities, that is, $c(2|1) = c(1|2)$ and $p_1 = p_2$, the above rule becomes:

$$\text{Allocate } \boldsymbol{x}_0 \text{ to } \Pi_1 \text{ if } \ \boldsymbol{b}^T \boldsymbol{x}_0 - k \geq 0, \ \text{ otherwise allocate } \boldsymbol{x}_0 \text{ to } \Pi_2.$$

where $\boldsymbol{b} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ and $k = \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \Sigma^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$.

Example 2: Allocate the following observations, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to most suitable population among $\Pi_1 : N_2(\boldsymbol{\mu}_1, \Sigma)$ and $\Pi_2 : N_2(\boldsymbol{\mu}_2, \Sigma)$, where

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \ \boldsymbol{\mu}_2 = \begin{pmatrix} 5 \\ 6 \\ 1 \end{pmatrix}, \quad \text{and } \Sigma = \begin{pmatrix} 9 & 4 & -2 \\ 4 & 4 & 3 \\ -2 & 3 & 16 \end{pmatrix}.$$

Observations are: $\boldsymbol{x}_1^T = (1, 1, 0)$ and $\boldsymbol{x}_2^T = (0, 2, -3)$.

Assume equal misclassification costs and prior probabilities.

4 Sample Discriminant Rule for Two Normal Populations When $\Sigma_1 = \Sigma_2 = \Sigma$

If any or all the parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and Σ are unknown then estimate those unknown parameters using random samples from each of the two populations and apply the rule using the estimated values of the parameters.

Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{n_1}$ be a random sample from population Π_1 and $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{n_2}$ be a random sample from population Π_2. Then estimators of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are respectively given by

$$\widehat{\boldsymbol{\mu}}_1 = \overline{\boldsymbol{x}}_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1} \boldsymbol{x}_i, \quad \text{and} \quad \widehat{\boldsymbol{\mu}}_2 = \overline{\boldsymbol{y}}_{n_2} = \frac{1}{n_2}\sum_{i=1}^{n_2} \boldsymbol{y}_i.$$

Consider the two sample covariance matrices

$$\mathcal{S}_x = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n_1})(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n_1})^T$$

and

$$\mathcal{S}_y = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2} \left(\boldsymbol{y}_i - \overline{\boldsymbol{y}}_{n_2}\right)\left(\boldsymbol{y}_i - \overline{\boldsymbol{y}}_{n_2}\right)^T.$$

Since the two populations have the same covariance matrix Σ, the estimate of Σ is given by the pooled sample covariance matrix $\mathcal{S}_{pooled}$.

$$\widehat{\Sigma} = \mathcal{S}_{pooled} = \frac{(n_1 - 1)\mathcal{S}_x + (n_2 - 1)\mathcal{S}_y}{n_1 + n_2 - 2}.$$

Using the above estimates, the allocation rule given in (2) can be written as:

Allocate $\boldsymbol{x}_0$ to Π_1 if

$$\left(\overline{\boldsymbol{x}}_{n_1} - \overline{\boldsymbol{y}}_{n_2}\right)^T \mathcal{S}_{pooled}^{-1}\boldsymbol{x}_0 - \tfrac{1}{2}\left(\overline{\boldsymbol{x}}_{n_1} - \overline{\boldsymbol{y}}_{n_2}\right)^T \mathcal{S}_{pooled}^{-1}\left(\overline{\boldsymbol{x}}_{n_1} + \overline{\boldsymbol{y}}_{n_2}\right) \geq \ln\left[\frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}\right]$$

otherwise allocate $\boldsymbol{x}_0$ to Π_2.

Equivalently

$$\text{Allocate } \boldsymbol{x}_0 \text{ to } \Pi_1 \text{ if } \quad \widehat{\boldsymbol{b}}^T\boldsymbol{x}_0 - \widehat{k} \geq 0, \text{ otherwise allocate } \boldsymbol{x}_0 \text{ to } \Pi_2 \qquad (3)$$

where $\widehat{\boldsymbol{b}} = \mathcal{S}_{pooled}^{-1}(\overline{\boldsymbol{x}}_{n_1} - \overline{\boldsymbol{y}}_{n_2})$ and

$$\widehat{k} = \frac{1}{2}(\overline{\boldsymbol{x}}_{n_1} - \overline{\boldsymbol{y}}_{n_2})^T\mathcal{S}_{pooled}^{-1}(\overline{\boldsymbol{x}}_{n_1} + \overline{\boldsymbol{y}}_{n_2}) + \ln\left[\frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}\right].$$

This is also called *Fisher's sample linear discriminant rule.* The function $\widehat{\boldsymbol{b}}^T\boldsymbol{x}$ is called the *sample linear discriminant function* of $\boldsymbol{x}$.

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.

5 Classification for Two Normal Populations When $\Sigma_1 \neq \Sigma_2$

Let $\Pi_1 : N_p(\boldsymbol{\mu}_1, \Sigma_1)$ and $\Pi_2 : N_p(\boldsymbol{\mu}_2, \Sigma_2)$ two multivariate normal populations where $\Sigma_1 \neq \Sigma_2$. That is

$$f_i(\boldsymbol{x}) = (2\pi)^{-\frac{p}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i) \right\}, \ i = 1, 2$$

and

$$\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} = \frac{|\Sigma_2|^{\frac{1}{2}}}{|\Sigma_1|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_1)^T \Sigma_1^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_1) + \frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_2)^T \Sigma_2^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_2) \right\}.$$

Now using allocation rule given in (1):

Allocate $\boldsymbol{x}_0$ to Π_1 if

$$\frac{|\Sigma_2|^{\frac{1}{2}}}{|\Sigma_1|^{\frac{1}{2}}} \exp\left\{ -\tfrac{1}{2} (\boldsymbol{x}_0 - \boldsymbol{\mu}_1)^T \Sigma_1^{-1} (\boldsymbol{x}_0 - \boldsymbol{\mu}_1) + \tfrac{1}{2} (\boldsymbol{x}_0 - \boldsymbol{\mu}_2)^T \Sigma_2^{-1} (\boldsymbol{x}_0 - \boldsymbol{\mu}_2) \right\} \geq \frac{c(1|2)}{c(2|1)} \frac{p_2}{p_1}$$

otherwise allocate $\boldsymbol{x}_0$ to Π_2.

Equivalently we can write the allocate rule as:

Allocate $\boldsymbol{x}_0$ to Π_1 if

$$-\tfrac{1}{2} \boldsymbol{x}_0^T \left(\Sigma_1^{-1} - \Sigma_2^{-1} \right) \boldsymbol{x}_0 + \left(\boldsymbol{\mu}_1^T \Sigma_1^{-1} - \boldsymbol{\mu}_2^T \Sigma_2^{-1} \right) \boldsymbol{x}_0 - K \geq 0, \tag{4}$$

otherwise allocate $\boldsymbol{x}_0$ to Π_2

where

$$K = \frac{1}{2} \left(\boldsymbol{\mu}_1^T \Sigma_1^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^T \Sigma_2^{-1} \boldsymbol{\mu}_2 \right) + \frac{1}{2} \ln \left(\frac{|\Sigma_1|}{|\Sigma_2|} \right) + \ln \left[\frac{c(1|2)}{c(2|1)} \frac{p_2}{p_1} \right]$$

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 $p_2 = 0.75$.

6 Sample Discriminant Rule for Two Normal Populations When $\Sigma_1 \neq \Sigma_2$

If any or all the parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ Σ_1 and Σ_2 are unknown then estimate those unknown parameters using random samples from each of the two populations and apply the rule using the estimated values of the parameters. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_1}$ be a random sample from population Π_1 and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_2}$ be a random sample from population Π_2. Then estimators of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, Σ_1 and Σ_2 are respectively given by

$$\widehat{\boldsymbol{\mu}}_1 = \overline{\boldsymbol{x}}_{n_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} \boldsymbol{x}_i, \quad \widehat{\boldsymbol{\mu}}_2 = \overline{\boldsymbol{y}}_{n_2} = \frac{1}{n_2} \sum_{i=1}^{n_2} \boldsymbol{y}_i,$$

$$\widehat{\Sigma}_1 = \mathcal{S}_x = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n_1}\right) \left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n_1}\right)^T$$

and

$$\widehat{\Sigma}_1 = \mathcal{S}_y = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} \left(\boldsymbol{y}_i - \overline{\boldsymbol{y}}_{n_2}\right) \left(\boldsymbol{y}_i - \overline{\boldsymbol{y}}_{n_2}\right)^T.$$

Using the above estimates, the allocation rule given in (4) can be written as:

Allocate $\boldsymbol{x}_0$ to Π_1 if

$$-\tfrac{1}{2}\boldsymbol{x}_0^T\left(\mathcal{S}_x^{-1}-\mathcal{S}_y^{-1}\right)\boldsymbol{x}_0+\left(\overline{\boldsymbol{x}}_{n_1}^T\mathcal{S}_x^{-1}-\overline{\boldsymbol{y}}_{n_2}T\mathcal{S}_y^{-1}\right)\boldsymbol{x}_0-\widehat{K}\geq 0 \qquad (5)$$

otherwise allocate $\boldsymbol{x}_0$ to Π_2

where

$$\widehat{K}=\frac{1}{2}\ln\left(\frac{|\mathcal{S}_x|}{|\mathcal{S}_y|}\right)+\frac{1}{2}\left(\overline{\boldsymbol{x}}_{n_1}^T\mathcal{S}_x^{-1}\overline{\boldsymbol{x}}_{n_1}-\overline{\boldsymbol{y}}_{n_2}^T\mathcal{S}_y^{-1}\overline{\boldsymbol{y}}_{n_2}\right)+\ln\left[\frac{c(1|2)}{c(2|1)}\frac{p_2}{p_1}\right].$$

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:

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

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 $\boldsymbol{x}_2^T=(-1,-3)$.

7 Evaluating Discriminant Functions

The way to evaluate the discriminant functions is to calculate their error rate, that is probability of misclassification.

The total probability of misclassification (TPM) is given by

$$\begin{aligned}
\text{TMP} &= \mathcal{P}(\text{misclassifiying observations}) \\
&= \mathcal{P}(\text{observations from } \Pi_1 \text{ is misclassified}) \\
&\qquad + \mathcal{P}(\text{observations from } \Pi_2 \text{ is misclassified}) \\
&= p_1\mathcal{P}(\text{misclassifing an observation from } \Pi_1) \\
&\qquad + p_2\mathcal{P}(\text{misclassifing an observation from } \Pi_2).
\end{aligned}$$

The smallest value of TPM is called the optimum error rate (OER). Thus

$$\begin{aligned}
\text{OER} &= p_1\mathcal{P}(\text{misclassifing an observation from } \Pi_1) \\
&\qquad + p_2\mathcal{P}(\text{misclassifing an observation from } \Pi_2).
\end{aligned}$$

Apparent Error Rate (APER)

The APER is the fraction of the misclassified observations in the training sample. This can be easily calculated from the following confusion matrix. Let n_1 be the number of observations in the training sample from population Π_1 and n_2 be the number of observations in the training sample from population Π_2.

Confusion Matrix

		Predicted membership		Number of
	Population	Π_1	Π_2	Observations
Actual	Π_1	n_{1c}	n_{1m}	n_1
membership	Π_2	n_{2m}	n_{2c}	n_2

where n_{1c} = number of correctly classified observations in Π_1

n_{1m}= number of misclassified observations in Π_1

n_{2c} = number of correctly classified observations in Π_2

n_{2m}= number of misclassified observations in Π_2.

Note that $n_1 = n_{1c} + n_{1m}$ and $n_2 = n_{2c} + n_{2m}$. Then, the proportion of the misclassified observations in the training sample is given by

$$\text{APER } = \frac{n_{1m} + n_{2m}}{n_1 + n_2}.$$

Example 6: Refer Example 11.5, page 602, Johnson and Wichern

Actual Error Rate(AER)

The AER indicate how the sample discriminant function will perform in the future. In general it cannot be calculated. However using cross-validation method we can estimate the expected AER.

Estimation of Expected AER

Step 1: Start with the observations in Π_1. Remove (holdout) one observation and obtain the discriminant function using remaining $(n_1 - 1)$ observations from Π_1 and n_2 from Π_2.

Step 2: Classify the removed observation.

Step 3: Repeat Steps 1 and 2 until all the Π_1 observations are classified. Let $n_{1m}^{(H)}$ be the number of misclassified observations.

Step 4: Repeat Steps 1 and 3 for the Π_2 observations. Let $n_{2m}^{(H)}$ be the number of misclassified observations.

Now the estimate of the actual error rate is given by

$$\widehat{\mathbf{E}}(\text{AER}) = \frac{n_{1m}^{(H)} + n_{2m}^{(H)}}{n_1 + n_2}.$$

Example 7: Refer Example 11.6, page 603, Johnson and Wichern