Multivariate Analysis
Chapter 11: Discrimination and Classification

Course: STAT-403

Prof. Dr. Rumana Rois
Department of Statistics and Data Science
Jahangirnagar University

Discrimination and Classification

1. **Discrimination** and *classification* are multivariate techniques concerned with **separating distinct sets of objects (or observations)** and with *allocating new objects (observations)* to previously defined groups.

2. Discrimination analysis is rather exploratory in nature. As a separative procedure, it is often employed on a onetime basis in order to investigate observed differences when causal relationships are not well understood.

3. Classification procedures are less exploratory in the sense that they lead to well-defined rules, which can be used for assigning new objects.

Goals of Discrimination and Classification

Goal 1. To describe, either graphically (in three or fewer dimensions) or algebraically, the differential features of objects (observations) from several known collections (populations). We try to find "discriminants" whose numerical values are such that the collections are separated as much as possible.

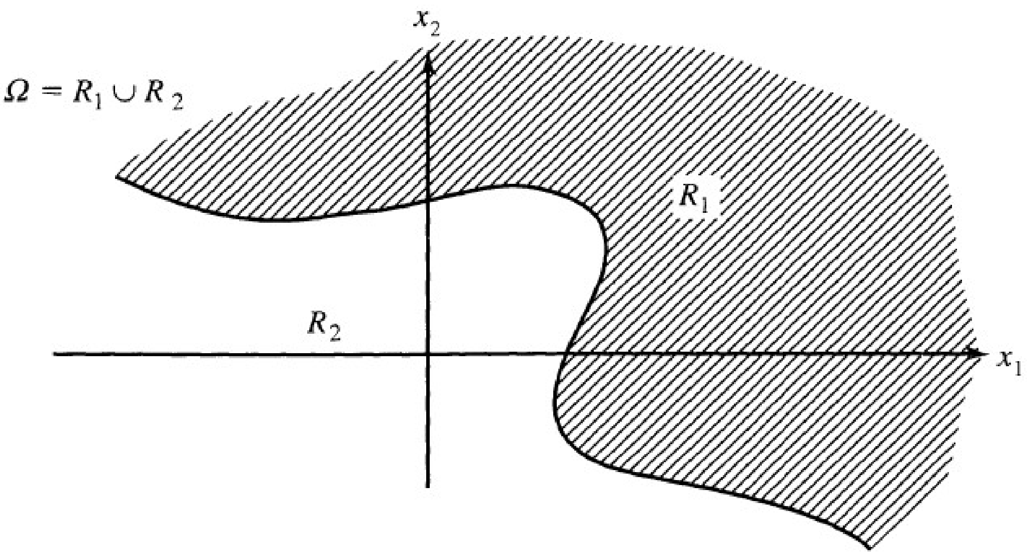
Goal 2. To sort objects (observations) into two or more labeled classes. The emphasis is on deriving a rule that can be used to optimally assign *new* objects to the labeled classes.

Separation and Classification for Two Populations

Consider that the entire population consists of two sub-populations, denoted by π_1 and π_2. The percentage of π_1 (π_2) in the entire population, which is called prior probability, is p_1 (p_2). Obviously $p_1 + p_2 = 1$. Suppose a random variable coming from π_1 has density $f_1(x)$ in p dimensional real space. In short, π_1 has density $f_1(x)$. Likewise, let π_2 has density $f_2(x)$. **The objectives are separated or classified on the basis of measurements on p associated random variables $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$.**

Populations π_1 and π_2	**Measured variable X**
Solvent and distressed property-liability insurance companies.	Total assets, cost of stocks and bonds, market value of stocks and bonds, loss expenses, surplus, amount of premiums written.
Two species of chickweed.	Sepal and petals length, petal cleft depth, bract length, scarious tip length, pollen diameter.
Purchasers of a new product and laggards (slow to purchase).	Education, income, family size, amount of previous brand switching.
Males and females.	Anthropological measurements, like circumference and volume on ancient skulls.

Rules for Assigning New Variables



$$P(2 \mid 1) = P(\mathbf{X} \in R_2 \mid \pi_1) = \int_{R_2 = \Omega - R_1} f_1(\mathbf{x})\, d\mathbf{x} \qquad (11\text{-}1)$$

$$P(1 \mid 2) = P(\mathbf{X} \in R_1 \mid \pi_2) = \int_{R_1} f_2(\mathbf{x})\, d\mathbf{x} \qquad (11\text{-}2)$$

Probability of Classification and Misclassification

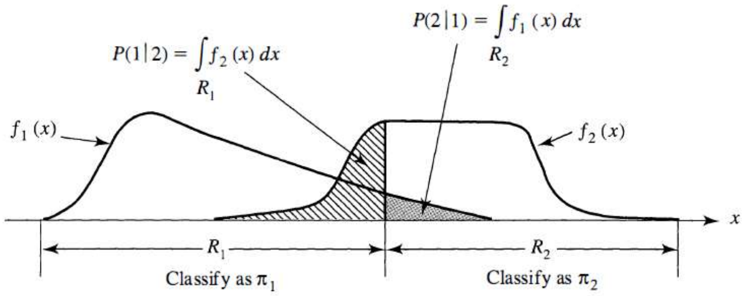


Figure 11.3 Misclassification probabilities for hypothetical classification regions when $p = 1$.

$$
\begin{aligned}
P(\text{observation is correctly classified as } \pi_1) &= P(\text{observation comes from } \pi_1 \text{ and is correctly classified as } \pi_1) \\
&= P(\mathbf{X} \in R_1 \mid \pi_1) P(\pi_1) = P(1 \mid 1) p_1 \\
P(\text{observation is misclassified as } \pi_1) &= P(\text{observation comes from } \pi_2 \text{ and is misclassified as } \pi_1) \\
&= P(\mathbf{X} \in R_1 \mid \pi_2) P(\pi_2) = P(1 \mid 2) p_2 \\
P(\text{observation is correctly classified as } \pi_2) &= P(\text{observation comes from } \pi_2 \text{ and is correctly classified as } \pi_2) \\
&= P(\mathbf{X} \in R_2 \mid \pi_2) P(\pi_2) = P(2 \mid 2) p_2
\end{aligned}
$$

$$
\begin{aligned}
P(\text{observation is misclassified as } \pi_2) &= P(\text{observation comes from } \pi_1 \text{ and is misclassified as } \pi_2) \\
&= P(\mathbf{X} \in R_2 \mid \pi_1) P(\pi_1) = P(2 \mid 1) p_1
\end{aligned} \qquad (11\text{-}3)
$$

Costs of Misclassification

The costs of misclassification can be defined by a cost matrix:

		Classify as:	
		π_1	π_2
True population:	π_1	0	$c(2 \mid 1)$
	π_2	$c(1 \mid 2)$	0

(11-4)

The costs are (1) zero for correct classification, (2) $c(1 \mid 2)$ when an observation from π_2 is incorrectly classified as π_1, and (3) $c(2 \mid 1)$ when a π_1 observation is incorrectly classified as π_2.

For any rule, the average, or *expected cost of misclassification* (ECM) is provided by multiplying the off-diagonal entries in (11-4) by their probabilities of occurrence, obtained from (11-3). Consequently,

$$\text{ECM} = c(2 \mid 1)P(2 \mid 1)p_1 + c(1 \mid 2)P(1 \mid 2)p_2 \qquad (11\text{-}5)$$

A reasonable classification rule should have an ECM as small, or nearly as small, as possible.

Rules of Classifying using ECM

Result 11.1. The regions R_1 and R_2 that minimize the ECM are defined by the values $\mathbf{x}$ for which the following inequalities hold:

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$\begin{pmatrix}\text{density}\\ \text{ratio}\end{pmatrix} \geq \begin{pmatrix}\text{cost}\\ \text{ratio}\end{pmatrix}\begin{pmatrix}\text{prior}\\ \text{probability}\\ \text{ratio}\end{pmatrix} \qquad (11\text{-}6)$$

$$R_2: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$\begin{pmatrix}\text{density}\\ \text{ratio}\end{pmatrix} < \begin{pmatrix}\text{cost}\\ \text{ratio}\end{pmatrix}\begin{pmatrix}\text{prior}\\ \text{probability}\\ \text{ratio}\end{pmatrix}$$

Rules of Classifying using ECM

SPECIAL CASES OF MINIMUM EXPECTED COST REGIONS

(a) $p_2/p_1 = 1$ (equal prior probabilities)

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{c(1\,|\,2)}{c(2\,|\,1)} \qquad R_2: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{c(1\,|\,2)}{c(2\,|\,1)}$$

(b) $c(1\,|\,2)/c(2\,|\,1) = 1$ (equal misclassification costs)

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{p_2}{p_1} \qquad R_2: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{p_2}{p_1} \qquad (11\text{-}$$

(c) $p_2/p_1 = c(1\,|\,2)/c(2\,|\,1) = 1$ or $p_2/p_1 = 1/(c(1\,|\,2)/c(2\,|\,1))$ (equal prior probabilities and equal misclassification costs)

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq 1 \qquad R_2: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < 1$$

Example 11.2

Example 11.2 (Classifying a new observation into one of the two populations)

A researcher has enough data available to estimate the density functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ associated with populations π_1 and π_2, respectively. Suppose $c(2\,|\,1) = 5$ units and $c(1\,|\,2) = 10$ units. In addition, it is known that about 20% of *all* objects (for which the measurements $\mathbf{x}$ can be recorded) belong to π_2. Thus, the prior probabilities are $p_1 = .8$ and $p_2 = .2$.

Suppose $f_1(\mathbf{x_0}) = 0.3$ and $f_2(\mathbf{x_0}) = 0.4$. Do we classify the new observation $\mathbf{x_0}$ as π_1 or π_2?

Classification with Two Multivariate Normal Populations

1. Case 1: When $\Sigma_1 = \Sigma_2 = \Sigma$

Suppose that the joint densities of $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ for populations π_1 and π_2 are given by

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right] \quad \text{for } i = 1, 2 \qquad (11\text{-}10)$$

Suppose also that the population parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\Sigma}$ are known. Then, after cancellation of the terms $(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}$ the minimum ECM regions in (11-6) become

$$R_1: \quad \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2)\right] \geq \left(\frac{c(1\,|\,2)}{c(2\,|\,1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$R_2: \quad \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2)\right] < \left(\frac{c(1\,|\,2)}{c(2\,|\,1)}\right)\left(\frac{p_2}{p_1}\right) \qquad (11\text{-}11)$$

Given these regions R_1 and R_2, we can construct the classification rule given in the following result.

Classification with Two Multivariate Normal Populations

1. Case 1: When $\Sigma_1 = \Sigma_2 = \Sigma$

Result 11.2. Let the populations π_1 and π_2 be described by multivariate normal densities of the form (11-10). Then the allocation rule that minimizes the ECM is as follows:

Allocate $\mathbf{x}_0$ to π_1 if

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\mathbf{x}_0 - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq \ln\left[\left(\frac{c(1\,|\,2)}{c(2\,|\,1)}\right)\left(\frac{p_2}{p_1}\right)\right] \qquad (11\text{-}12)$$

Allocate $\mathbf{x}_0$ to π_2 otherwise.

Classification with Two Multivariate Normal Populations

1. μ_1, μ_2, **and Σ are unknown and need to replace by their sample counterparts.**

$$\underset{(p\times 1)}{\bar{\mathbf{x}}_1} = \frac{1}{n_1}\sum_{j=1}^{n_1}\mathbf{x}_{1j}, \qquad \underset{(p\times p)}{\mathbf{S}_1} = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)'$$
$$\underset{(p\times 1)}{\bar{\mathbf{x}}_2} = \frac{1}{n_2}\sum_{j=1}^{n_2}\mathbf{x}_{2j}, \qquad \underset{(p\times p)}{\mathbf{S}_2} = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)' \qquad (11\text{-}16)$$

Since it is assumed that the parent populations have the same covariance matrix $\mathbf{\Sigma}$, the sample covariance matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ are combined (pooled) to derive a single, unbiased estimate of $\mathbf{\Sigma}$ as in (6-21). In particular, the weighted average

$$\mathbf{S}_{\text{pooled}} = \left[\frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)}\right]\mathbf{S}_1 + \left[\frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)}\right]\mathbf{S}_2 \qquad (11\text{-}17)$$

is an unbiased estimate of $\mathbf{\Sigma}$ if the data matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ contain *random* samples from the populations π_1 and π_2, respectively.

Substituting $\bar{\mathbf{x}}_1$ for $\boldsymbol{\mu}_1$, $\bar{\mathbf{x}}_2$ for $\boldsymbol{\mu}_2$, and $\mathbf{S}_{\text{pooled}}$ for $\mathbf{\Sigma}$ in (11-12) gives the "sample" classification rule:

THE ESTIMATED MINIMUM ECM RULE FOR TWO NORMAL POPULATIONS

Allocate $\mathbf{x}_0$ to π_1 if

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pooled}}^{-1}\mathbf{x}_0 - \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pooled}}^{-1}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \ge \ln\left[\left(\frac{c(1\mid 2)}{c(2\mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \qquad (11\text{-}18)$$

Allocate $\mathbf{x}_0$ to π_2 otherwise.

Classification with Two Multivariate Normal Populations

2. Case 2: When $\Sigma_1 \neq \Sigma_2$

$$R_1: \quad -\frac{1}{2}\mathbf{x}'(\mathbf{\Sigma}_1^{-1} - \mathbf{\Sigma}_2^{-1})\mathbf{x} + (\boldsymbol{\mu}_1'\mathbf{\Sigma}_1^{-1} - \boldsymbol{\mu}_2'\mathbf{\Sigma}_2^{-1})\mathbf{x} - k \ge \ln\left[\left(\frac{c(1\mid 2)}{c(2\mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$
$$R_2: \quad -\frac{1}{2}\mathbf{x}'(\mathbf{\Sigma}_1^{-1} - \mathbf{\Sigma}_2^{-1})\mathbf{x} + (\boldsymbol{\mu}_1'\mathbf{\Sigma}_1^{-1} - \boldsymbol{\mu}_2'\mathbf{\Sigma}_2^{-1})\mathbf{x} - k < \ln\left[\left(\frac{c(1\mid 2)}{c(2\mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \qquad (11\text{-}23)$$

where

$$k = \frac{1}{2}\ln\left(\frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_2|}\right) + \frac{1}{2}(\boldsymbol{\mu}_1'\mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2'\mathbf{\Sigma}_2^{-1}\boldsymbol{\mu}_2) \qquad (11\text{-}24)$$

Classification with Two Multivariate Normal Populations

2. Substituting $\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2$, $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ for $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$

QUADRATIC CLASSIFICATION RULE
(NORMAL POPULATIONS WITH UNEQUAL COVARIANCE MATRICES)

Allocate $\mathbf{x}_0$ to π_1 if

$$-\frac{1}{2}\mathbf{x}_0'(\mathbf{S}_1^{-1} - \mathbf{S}_2^{-1})\mathbf{x}_0 + (\bar{\mathbf{x}}_1'\mathbf{S}_1^{-1} - \bar{\mathbf{x}}_2'\mathbf{S}_2^{-1})\mathbf{x}_0 - k \geq \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{11-25}$$

Allocate $\mathbf{x}_0$ to π_2 otherwise.

Evaluating Classification Functions

1. Total probability of misclassification:
$$\text{TPM} = p_1 \int_{R_2} f_1(\mathbf{x})d\mathbf{x} + p_2 \int_{R_1} f_2(\mathbf{x})d\mathbf{x}$$

2. Optimum Error Rate:
$$\text{OER} = p_1 \int_{R_2} f_1(\mathbf{x})d\mathbf{x} + p_2 \int_{R_1} f_2(\mathbf{x})d\mathbf{x}$$
where R_1 and R_2 are determined for equal misclassification cost.

3. Actual Error Rate (sample classification function):
$$\text{AER} = p_1 \int_{\hat{R}_2} f_1(\mathbf{x})d\mathbf{x} + p_2 \int_{\hat{R}_1} f_2(\mathbf{x})d\mathbf{x}$$
where $\hat{R}_1$ and $\hat{R}_2$ are determined by sample size n_1 and n_2, respectively.

Evaluating Classification Functions

4. Apparent Error Rate (APER)

Confusion Matrix

		Predicted membership π_1	Predicted membership π_2		
Actual membership	π_1	n_{1C}	$n_{1M} = n_1 - n_{1C}$	n_1	(11-29)
	π_2	$n_{2M} = n_2 - n_{2C}$	n_{2C}	n_2	

where

n_{1C} = number of π_1 items correctly classified as π_1 items
n_{1M} = number of π_1 items misclassified as π_2 items
n_{2C} = number of π_2 items correctly classified
n_{2M} = number of π_2 items misclassified

The apparent error rate is then

$$\text{APER} = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \qquad (11\text{-}30)$$

which is recognized as the *proportion* of items in the training set that are misclassified.

Fisher's Linear Discriminant Function

Assume that the two populations have a common covariance matrix but does not assume that the populations are normal.

Result 11.4. The linear combination $\hat{y} = \hat{\mathbf{a}}'\mathbf{x} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pooled}}^{-1}\mathbf{x}$ maximizes the ratio

$$\frac{\begin{pmatrix}\text{Squared distance} \\ \text{between sample means of } y\end{pmatrix}}{(\text{Sample variance of } y)} = \frac{(\bar{y}_1 - \bar{y}_2)^2}{s_y^2} = \frac{(\hat{\mathbf{a}}'\bar{\mathbf{x}}_1 - \hat{\mathbf{a}}'\bar{\mathbf{x}}_2)^2}{\hat{\mathbf{a}}'\mathbf{S}_{\text{pooled}}\hat{\mathbf{a}}} = \frac{(\hat{\mathbf{a}}'\mathbf{d})^2}{\hat{\mathbf{a}}'\mathbf{S}_{\text{pooled}}\hat{\mathbf{a}}} \qquad (11\text{-}33)$$

over all possible coefficient vectors $\hat{\mathbf{a}}$ where $\mathbf{d} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$. The maximum of the ratio (11-33) is $D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pooled}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$.

Proof. The maximum of the ratio in (11-33) is given by applying (2-50) directly. Thus, setting $\mathbf{d} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$, we have

$$\max_{\hat{\mathbf{a}}} \frac{(\hat{\mathbf{a}}'\mathbf{d})^2}{\hat{\mathbf{a}}'\mathbf{S}_{\text{pooled}}\hat{\mathbf{a}}} = \mathbf{d}'\mathbf{S}_{\text{pooled}}^{-1}\mathbf{d} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pooled}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = D^2$$

where D^2 is the sample squared distance between the two means. ■

Note that s_y^2 in (11-33) may be calculated as

$$s_y^2 = \frac{\sum_{j=1}^{n_1}(y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2}(y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} \qquad (11\text{-}34)$$

with $y_{1j} = \hat{\mathbf{a}}'\mathbf{x}_{1j}$ and $y_{2j} = \hat{\mathbf{a}}'\mathbf{x}_{2j}$.

Fisher's Linear Discriminant Function

AN ALLOCATION RULE BASED ON FISHER'S DISCRIMINANT FUNCTION[8]

Allocate $\mathbf{x}_0$ to π_1 if

$$\hat{y}_0 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pooled}}^{-1}\mathbf{x}_0 \geq \hat{m} = \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pooled}}^{-1}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)$$

or (11-35)

$$\hat{y}_0 - \hat{m} \geq 0$$

Allocate $\mathbf{x}_0$ to π_2 if

$$\hat{y}_0 < \hat{m}$$

or

$$\hat{y}_0 - \hat{m} < 0$$

Real Data Example

1. In a survey, the cost of transporting milk from farms to dairy plants by gasoline trucks and diesel trucks was observed. Cost data on X_1= fuel, X_2= repair, and X_3= capital, all measures on a per-mile basis are presented in Table 1 for n_1= 36 gasoline and n_2= 23 diesel trucks. This data is available in the file **multivar5th/T6-10.dat**.

Table 1: Milk Transportation-Cost Data

Gasoline Trucks			Diesel Trucks		
x_1	x_2	x_3	x_1	x_2	x_3
16.44	12.43	11.23	8.50	12.26	9.11
⋮	⋮	⋮	⋮	⋮	⋮
17.32	6.86	4.44	12.03	9.22	23.09

(i) Calculate Fisher's linear discriminant function for classifying the cost data into gasoline or diesel trucks.

(ii) Classify the cost data with characteristics $x_1 = 16$, $x_2 = 10$, and $x_3 = 10$ into gasoline or diesel trucks?

Real Data Example

```
> ###### Using lda function from MASS package
> ###### we can get the linear discriminant function.
> ###### Linear Discriminant Analysis
> x=read.table('T6-10.dat',header = FALSE)
> library(MASS)
> fit_lda=lda(x[,4]~x[,1]+x[,2]+x[,3])
> fit_lda
Call:
lda(x[, 4] ~ x[, 1] + x[, 2] + x[, 3])

Prior probabilities of groups:
   diesel  gasoline
0.3898305 0.6101695

Group means:
           x[, 1]   x[, 2]    x[, 3]
diesel   10.10565 10.76217 18.167826
gasoline 12.21861  8.11250  9.590278

Coefficients of linear discriminants:
               LD1
x[, 1]  0.13374629
x[, 2] -0.07030203
x[, 3] -0.16739189
```


Real Data Example

```
> x=read.table('T6-10.dat',header = FALSE)
>
> group1 <- x[x[,4] == 'gasoline',][,1:3]
> group2 <- x[x[,4] == 'diesel',][,1:3]
> n1 <- nrow(group1)
> n2 <- nrow(group2)
>
> group1.means <- apply(group1, 2, mean)
> group2.means <- apply(group2, 2, mean)
>
> w1 <- (n1 - 1) * var(group1)
> w2 <- (n2 - 1) * var(group2)
>
> sp1 <- 1 / (n1 + n2 - 2) * (w1 + w2)
> cutoff <- .5 * (group1.means - group2.means) %*% solve(sp1) %*% (group1.means +
group2.means)
> cutoff
          [,1]
[1,] -2.845247
>
> x0=matrix(c(16,10,10), 3, 1,byrow = T)
> y_hat=(group1.means - group2.means) %*% solve(sp1) %*% x0
> if (y_hat>=cutoff) {
+   predict_class='gasoline'
+   print ( predict_class)
+ } else { predict_class= 'diesel'
+ print (predict_class) }
[1] "gasoline"
```

A Test for Significant Separation of two populations

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ vs. } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

Test statistic:

$$\left(\frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p}\right)\left(\frac{n_1 n_2}{n_1 + n_2}\right) D^2 \sim F_{p, n_1+n_2-p-1}.$$

If H_0 is rejected, we can conclude that the separation between the two populations π_1 and π_2 is significant.

Classification with Several Populations

The Minimum Expected Cost of Misclassification Method

Let $f_i(\mathbf{x})$ be the density associated with population $\pi_i, i = 1, 2, \ldots, g$.

$$p_i = \text{the prior probability of population } \pi_i, \quad i = 1, 2, \ldots, g$$
$$c(k \mid i) = \text{the cost of allocating an item to } \pi_k \text{ when, in fact, it belongs to } \pi_i, \quad \text{for } k, i = 1, 2, \ldots, g$$

For $k = i$, $c(i \mid i) = 0$. Finally, let R_k be the set of $\mathbf{x}$'s classified as π_k and

$$P(k \mid i) = P(\text{classifying item as } \pi_k \mid \pi_i) = \int_{R_k} f_i(\mathbf{x})\, d\mathbf{x}$$

for $k, i = 1, 2, \ldots, g$ with $P(i \mid i) = 1 - \sum_{\substack{k=1 \\ k \neq i}}^{g} P(k \mid i)$.

The conditional expected cost of misclassifying an $\mathbf{x}$ from π_1 into π_2, or $\pi_3, \ldots,$ or π_g is

$$\begin{aligned} \text{ECM}(1) &= P(2 \mid 1)c(2 \mid 1) + P(3 \mid 1)c(3 \mid 1) + \cdots + P(g \mid 1)c(g \mid 1) \\ &= \sum_{k=2}^{g} P(k \mid 1)c(k \mid 1) \end{aligned}$$

Classification with Several Populations

1. Allocate $\mathbf{x}$ to that population π_k for which $\sum_{i=1, i \neq k} p_i f_i(\mathbf{x}) c(k|i)$ is **smallest**.

2. Allocate $\mathbf{x}$ to π_k if $p_k f_k(\mathbf{x}) > p_i f_i(\mathbf{x})$ for all $i \neq k$ (with equal misclassification cost).

3. Allocate $\mathbf{x}$ to π_k if $P(\pi_k|\mathbf{x}) = \frac{p_k f_k(\mathbf{x})}{\sum_{i=1}^{g} p_i f_i(\mathbf{x})}$ for $k = 1, \ldots, g$ is **largest**.

Classification with Several Populations

Example 11.9 (Classifying a new observation into one of three known populations)

Let us assign an observation $\mathbf{x}_0$ to one of the $g = 3$ populations π_1, π_2, or π_3, given the following hypothetical prior probabilities, misclassification costs, and density values:

		True population		
		π_1	π_2	π_3
Classify as:	π_1	$c(1\|1) = 0$	$c(1\|2) = 500$	$c(1\|3) = 100$
	π_2	$c(2\|1) = 10$	$c(2\|2) = 0$	$c(2\|3) = 50$
	π_3	$c(3\|1) = 50$	$c(3\|2) = 200$	$c(3\|3) = 0$
Prior probabilities:		$p_1 = .05$	$p_2 = .60$	$p_3 = .35$
Densities at $\mathbf{x}_0$:		$f_1(\mathbf{x}_0) = .01$	$f_2(\mathbf{x}_0) = .85$	$f_3(\mathbf{x}_0) = 2$