4. Repeat Steps 1 through 3 for the $\pi_2$ observations. Let $n_2^{(H)}$ be the number of holdout observations misclassified in this group.

$$
\hat{P}(2|1) = \frac{n_1^{(H)}}{n_1}, \quad \hat{P}(1|2) = \frac{n_2^{(H)}}{n_2},
$$

$$
\hat{E}(\text{AER}) = \frac{n_1^{(H)} + n_2^{(H)}}{n_1 + n_2}.
$$

**Reading**: textbook, 11.1, 11.2, 11.3, 11.4

- Classification with Several Populations
  - Generalization of classification procedure from 2 to $g \geq 2$ groups
  - Minimum expected cost of misclassification method
    - Let $f_i(x)$ be the density associated with population $\pi_i$, $i = 1, 2, \ldots, g$.
    - Let $p_i = \text{the prior probability of population } \pi_i$, $i = 1, 2, \ldots, g$.
    - Let $c(k|i)$ = the cost of allocating an item to $\pi_k$ when, in fact, it belongs to $\pi_i$, for $k, i = 1, 2, \ldots, g$.
    - Let $R_k$ be the set of x’s classified as $\pi_k$.

Then,

$$
P(k|i) = P(\text{classifying item as } \pi_k | i) = \int_{R_k} f_i(x) \, dx \quad \text{for } k, i = 1, 2, \ldots, g.
$$

The overall ECM is:

$$
\text{ECM}(1) = P(2|1)c(2|1) + P(3|1)c(3|1) + \cdots + P(g|1)c(g|1) = \sum_{k=2}^{g} P(k|1)c(k|1)
$$

In a similar manner, we can obtain the conditional expected costs of misclassification $\text{ECM}(2), \ldots, \text{ECM}(g)$.

The overall ECM is:

$$
\text{ECM} = p_1 \text{ECM}(1) + p_2 \text{ECM}(2) + \cdots + p_g \text{ECM}(g)
$$

$$
= p_1 \left( \sum_{k=2}^{g} P(k|1)c(k|1) \right) + p_2 \left( \sum_{k=1, k \neq 2}^{g} P(k|2)c(k|2) \right) + \cdots + p_g \left( \sum_{k=1}^{g-1} P(k|g)c(k|g) \right)
$$

$$
= \sum_{i=1}^{g} p_i \left( \sum_{k=1, k \neq i}^{g} P(k|i)c(k|i) \right)
$$

**Result 11.5.** The classification regions that minimize the ECM are defined by allocating $x$ to that population $\pi_k$, $k = 1, 2, \ldots, g$, for which

$$
\sum_{i=1}^{g} p_i f_i(x) c(k|i)
$$

is smallest. If a tie occurs, $x$ can be assigned to any of the tied populations.
Suppose all the misclassification costs are equal. The minimum ECM is the minimum total probability of misclassification. In which case, we would allocate $x$ to that population $\pi_k$, $k = 1, 2, \ldots, g$, for which

$$\sum_{i=1}^{g} p_i f_i(x)$$

is smallest. It will be smallest when the omitted term, $p_k f_k(x)$, is largest.

Minimum ECM Classification Rule with equal misclassification costs

Allocate $x_0$ to $\pi_k$ if $p_k f_k(x) > p_i f_i(x)$ for all $i \neq k$

Notice that the classification rule is identical to the one that maximize the posterior probability

$$P(\pi_k | x) = P(x \text{ comes from } \pi_k \text{ given that } x \text{ was observed}) = \frac{p_k f_k(x)}{\sum_{i=1}^{g} p_i f_i(x)} \frac{(\text{prior}) \times (\text{likelihood})}{\sum [(\text{prior}) \times (\text{likelihood})]} \text{ for } k = 1, 2, \ldots, g$$

Classification with Normal Populations

Under normality assumption,

$$f_i(x) = \frac{1}{(2\pi)^{p/2} |\Sigma_i|^{1/2}} \exp \left( -\frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i) \right), \ i = 1, 2, \ldots, g$$

Allocate $x$ to $\pi_k$ if

$$\ln p_k f_k(x) = \ln p_k - \left( \frac{p}{2} \right) \ln (2\pi) - \frac{1}{2} \ln |\Sigma_k| - \frac{1}{2} (x - \mu_k)' \Sigma_k^{-1} (x - \mu_k)$$

$$= \max_i \ln p_i f_i(x)$$

define quadratic discrimination score for $i$th population

$$d_i^Q(x) = -\frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i) + \ln p_i \quad i = 1, 2, \ldots, g$$

Minimum Total Probability of Misclassification (TPM) rule for normal populations with unequal $\Sigma_i$

Allocate $x$ to $\pi_k$ if the quadratic score $d_k^Q(x)$ is largest of $d_1^Q(x), d_2^Q(x), \ldots, d_g^Q(x)$

In practice, the $\mu_i$ and $\Sigma_i$ are unknown ⇒ replaced by their sample quantities

$$\hat{d}_i^Q(x) = -\frac{1}{2} \ln |S_i| - \frac{1}{2} (x - \bar{x}_i)' S_i^{-1} (x - \bar{x}_i) + \ln p_i, \quad i = 1, 2, \ldots, g$$

Estimated Minimum TPM rule for normal population with unequal $\Sigma_i$

Allocate $x$ to $\pi_k$ if the quadratic score $\hat{d}_k^Q(x)$ is largest of $\hat{d}_1^Q(x), \hat{d}_2^Q(x), \ldots, \hat{d}_g^Q(x)$

When $\Sigma_i = \Sigma$, for $i = 1, 2, \ldots, g$,

$$d_i^Q(x) = -\frac{1}{2} \ln |\Sigma| - \frac{1}{2} x' \Sigma^{-1} x + \mu_i' \Sigma^{-1} x - \frac{1}{2} \mu_i' \Sigma^{-1} \mu_i + \ln p_i$$

The first two terms are the same for $d_1^Q(x), d_2^Q(x), \ldots, d_g^Q(x)$.
- Define the linear discriminant score

\[ d_i(x) = \mu_i^\prime \Sigma^{-1} x - \frac{1}{2} \mu_i^\prime \Sigma^{-1} \mu_i + \ln p_i \quad \text{for } i = 1, 2, \ldots, g \]

An estimate \( \hat{d}_i(x) \) of the linear discriminant score \( d_i(x) \) is based on the pooled estimate:

\[ S_{\text{pooled}} = \frac{1}{n_1 + n_2 + \cdots + n_g - g} \left( (n_1 - 1) S_1 + (n_2 - 1) S_2 + \cdots + (n_g - 1) S_g \right) \]

and is given by

\[ \hat{d}_i(x) = \bar{x}_i^\prime S_{\text{pooled}}^{-1} x - \frac{1}{2} \bar{x}_i^\prime S_{\text{pooled}}^{-1} \bar{x}_i + \ln p_i \quad \text{for } i = 1, 2, \ldots, g \]

- Estimated Minimum TPM rule for normal populations with equal covariance

Allocate \( x \) to \( \pi_k \) if

the linear discriminant score \( \hat{d}_k(x) \) is the largest of \( \hat{d}_1(x), \hat{d}_2(x), \ldots, \hat{d}_g(x) \).

An equivalent classifier for the equal-covariance case is to use

\[ D_i^2(x) = (x - \bar{x}_i)^\prime S_{\text{pooled}}^{-1} (x - \bar{x}_i) \]

It measures the squared distances from \( x \) to the sample mean vector \( \bar{x}_i \).

The allocatory rule is then

Assign \( x \) to the population \( \pi_i \) for which \(-\frac{1}{2} D_i^2(x) + \ln p_i \) is largest.

* If the prior probabilities are unknown, the usual procedure is to set \( p_1 = p_2 = \cdots = p_g = 1/g \). An observation is then assigned to the closest population.