Remark: For all the 3 examples ECM (or TPM) of the ECM minimizing classification rules can be calculated once we calculate $P(1|2)$ & $P(2|1)$

e.g: Consider "Example 1" of $\pi_1 : N_p(\underset{\sim}{\mu}_1, \Sigma_*)$

$\pi_2 : N_p(\underset{\sim}{\mu}_2, \Sigma)$

$$P(1|2) = P_{\pi_2}\left(\underset{\sim}{X} \in R_1^*\right)$$

$$= P_{\pi_2}\left((\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \Sigma^{-1} \underset{\sim}{X} \geq \log\left(\frac{p_2 C(1|2)}{p_1 C(2|1)}\right) + \frac{1}{2}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \Sigma^{-1} (\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2)\right)$$

Note that under π_2

$$Z_1 = (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \Sigma^{-1} \underset{\sim}{X} \sim N_1\left((\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \Sigma^{-1} \underset{\sim}{\mu}_2, \underbrace{(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \Sigma^{-1} (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)}_{= \Delta^2 \leftarrow \text{Sq of Mahalanobis distance}}\right)$$

$$\Rightarrow P(1|2) = P_{\pi_2}\Big(\underbrace{\frac{Z_1 - (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \Sigma^{-1} \underset{\sim}{\mu}_2}{\Delta}}_{\text{under } \pi_2 \sim N(0,1)} \geq \Big[\frac{1}{2}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \Sigma^{-1} (\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2) - (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \Sigma^{-1} \underset{\sim}{\mu}_2 + \log\left(\frac{p_2 C(1|2)}{p_1 C(2|1)}\right)\Big] \Delta^{-1}\Big)$$

Let $Y = \dfrac{Z_1 - (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \Sigma^{-1} \underset{\sim}{\mu}_2}{\Delta}$

$$P(1|2) = P_{\pi_2}\left(Y \geq \frac{1}{2}\Delta + \frac{1}{\Delta} \log\left(\frac{p_2 C(1|2)}{p_1 C(2|1)}\right)\right)$$

$$\Rightarrow P(1|2) = 1 - \Phi\left(\frac{\Delta}{2} + \frac{1}{\Delta}\log\left(\frac{p_2 C(1|2)}{p_1 C(2|1)}\right)\right)$$

Sly $P(2|1)$ can be calculated

Note: TPM / ECM can be estimated by estimating Δ from learning set $\mathcal{L}$.

Remark: Performance measure for classifier

Apparent error rate (APER)

- Calculate the "confusion matrix"

Actual class \ Predicted class	π_1	π_2	
π_1	n_{1C}	n_{1M}	$(n_{1C} + n_{1M} = n_1)$
π_2	n_{2M}	n_{2C}	$(n_{2M} + n_{2C} = n_2)$

n_{iM} : # of cases from π_i misclassified by classifier

n_{iC} : # of - - - - π_i correctly classified

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2}$$

Classification in multiclass problem

Let $\pi_1, \pi_2, \ldots, \pi_C$ be C classes with prior probabilities $p(\pi_1), p(\pi_2), \ldots p(\pi_C)$, respectively.

Bayes classifier

Assign $\underset{\sim}{x}$ to π_j if the posterior prob of the class π_j given the obsn $\underset{\sim}{x}$, i.e., $p(\pi_j | \underset{\sim}{x})$ is the highest over all classes $\pi_1, \ldots, \pi_C$

i.e. Assign $\underset{\sim}{x}$ to π_j if

$$p(\pi_j | \underset{\sim}{x}) \geq p(\pi_K | \underset{\sim}{x}) \; ; \; K = 1, \ldots, C, \; K \neq j$$

$$\Leftrightarrow \quad \frac{p(\pi_j)\, p(\underset{\sim}{x} | \pi_j)}{\sum_{i=1}^{C} p(\pi_i) p(\underset{\sim}{x} | \pi_i)} \geq \frac{p(\pi_K)\, p(\underset{\sim}{x} | \pi_K)}{\sum_{i=1}^{C} p(\pi_i)\, p(\underset{\sim}{x} | \pi_i)}$$

$$\Leftrightarrow \quad p(\pi_j)\, p(\underset{\sim}{x} | \pi_j) \geq p(\pi_K)\, p(\underset{\sim}{x} | \pi_K)$$

$\Rightarrow$ as in the 2-class problem (done in class), Bayes classifier for a $C (\geq 2)$-class problem can be expressed in terms of prior and class conditional densities

TPM minimizing classifier

TPM for a general c-class problem can be defined as

$$TPM = \sum_{i=1}^{c} p(\pi_i)\, P(\text{error} \mid \pi_i)$$

↑ misclassification

i.e. $P(\text{error} \mid \pi_i)$: prob of misclassifying a pattern from π_i

Let $\{R_1, R_2, \ldots, R_c\}$ be the classification partition, then

$$P(\text{error} \mid \pi_i) = \int_{R_i^c} f(\underset{\sim}{x} \mid \pi_i)\, d\underset{\sim}{x}$$

$$\Rightarrow TPM = \sum_{i=1}^{c} \int_{\mathcal{X}-R_i} f(\underset{\sim}{x} \mid \pi_i)\, p(\pi_i)\, d\underset{\sim}{x}$$

$$= \sum_{i=1}^{c} p(\pi_i) \int_{\mathcal{X}-R_i} f(\underset{\sim}{x} \mid \pi_i)\, d\underset{\sim}{x}$$

$$= \sum_{i=1}^{c} p(\pi_i) \left(\int_{\mathcal{X}} f(\underset{\sim}{x} \mid \pi_i)\, d\underset{\sim}{x} - \int_{R_i} f(\underset{\sim}{x} \mid \pi_i)\, d\underset{\sim}{x} \right)$$

$$= \sum_{i=1}^{c} p(\pi_i) \left(1 - \int_{R_i} f(\underset{\sim}{x} \mid \pi_i)\, d\underset{\sim}{x} \right.$$

i.e.

$$TPM = 1 - \sum_{i=1}^{C} \int_{R_i} p(\pi_i) f(\underset{\sim}{x} | \pi_i) \, d\underset{\sim}{x}$$

$\Rightarrow$ Minimizing the TPM w.r.t. partition $\{R_1, \ldots R_C\}$ is equivalent to maximizing

$$\sum_{i=1}^{C} \int_{R_i} p(\pi_i) f(\underset{\sim}{x} | \pi_i) \, d\underset{\sim}{x} \quad \text{w.r.t. } \{R_1, \ldots, R_C\}$$

This can be viewed as prob of correct classification

Maximization of the above is achieved by selecting R_i to be the region for which $p(\pi_i) f(\underset{\sim}{x} | \pi_i)$ is the largest among all classes (i.e among all $i = 1(1)C$)

i.e. the TPM minimizing classification rule is

Assign $\underset{\sim}{x}$ to π_i if

$$p(\pi_i) f(\underset{\sim}{x} | \pi_i) \geq p(\pi_K) f(\underset{\sim}{x} | \pi_K) \quad \forall K \neq i$$

Remark: As in the 2-class problem, the TPM minimizing rule is equivalent to the one obtained by maximizing the "posterior prob", i.e. the Bayes classifier

Minimum ECM classification rule

Let $p(\pi_i) = p_i \quad i = 1(1)C$

class conditional densities $f(\underset{\sim}{x} | \pi_i) \; ; \; i = 1(1)C$

$c(K|i)$: cost of misclassifying an item/pattern from π_i to π_K ; $K, i = 1(1)C$

$c(i|i) = 0 \quad \forall \; i = 1(1)C$

$\{R_1, \ldots, R_C\}$: classification partition

$$P(K|i) = P(\text{classifying an item to } \pi_K | \pi_i)$$

$$= \int_{R_K} f(\underset{\sim}{x} | \pi_i) \, d\underset{\sim}{x}$$

$$P(i|i) = \int_{R_i} f(\underset{\sim}{x} | \pi_i) \, d\underset{\sim}{x} = \int_{\mathcal{X} - \cup_{K \neq i} R_K} f(\underset{\sim}{x} | \pi_i) \, d\underset{\sim}{x}$$

$$= 1 - \sum_{\substack{K=1 \\ K \neq i}}^{C} P(K|i)$$

Note that the conditional ECM of $\underset{\sim}{x}$ from π_1 into π_2 or π_3 or ... or π_R is

$$ECM_1 = P(2|1)\,c(2|1) + P(3|1)\,c(3|1) + \ldots + P(C|1)\,c(C|1)$$

$$= \sum_{K=2}^{C} P(K|1)\,c(K|1)$$

The above ECM_1 is with prob $p(\pi_1) = p_1$

Thus for other conditional ECM's ECM_i we can write similar expressions w.p. p_i (as below)

$\Rightarrow$

$$ECM = p_1 ECM_1 + p_2 ECM_2 + \cdots + p_c ECM_c$$

i.e. $$ECM = p_1 \left(\sum_{\substack{K=2}}^{c} P(K|1) C(K|1) \right)$$

$$+ p_2 \left(\sum_{\substack{K=1 \\ K \neq 2}}^{c} P(K|2)\, C(K|2) \right) + \cdots$$

$$\cdots + p_c \left(\sum_{K=1}^{c-1} P(K|c)\, C(K|c) \right)$$

i.e. $$ECM = \sum_{i=1}^{c} p_i \sum_{\substack{K=1 \\ K \neq i}}^{c} P(K|i)\, C(K|i)$$

$$= \sum_{K=1}^{c} \sum_{\substack{i=1 \\ i \neq K}}^{c} p_i\, P(K|i)\, C(K|i)$$

$$= \sum_{K=1}^{c} \sum_{\substack{i=1 \\ i \neq K}}^{c} \int_{R_K} p_i\, C(K|i)\, f(\underset{\sim}{x}|\pi_i)\, d\underset{\sim}{x}$$

$$\text{i.e. } ECM = \sum_{K=1}^{C} \int_{R_K} \left(\sum_{\substack{i=1 \\ i \neq K}}^{C} p_i \, c(K|i) \, f(\underset{\sim}{x}|\pi_i) \right) d\underset{\sim}{x}$$

$$= \sum_{K=1}^{C} \int_{R_K} h_K(\underset{\sim}{x}) \, d\underset{\sim}{x}$$

Note that

$$\sum_{K=1}^{C} \int_{R_K} \left(h_K(\underset{\sim}{x}) - \min_j h_j(\underset{\sim}{x}) \right) d\underset{\sim}{x} \geq 0$$

with equality only if

$$h_K(\underset{\sim}{x}) = \min_j h_j(\underset{\sim}{x}) \quad \forall \underset{\sim}{x} \text{ in } R_K$$

Thus the ECM minimizing classification rule is

Assign $\underset{\sim}{x}$ to π_K if

$$\sum_{\substack{i=1 \\ i \neq K}}^{C} p_i \, c(K|i) \, f(\underset{\sim}{x}|\pi_i)$$ is minimum among all such c expressions

Remark: under equal cost setup, i.e. all misclassification costs are same

we assign $\underset{\sim}{x}$ to π_K if

$$\sum_{\substack{i=1 \\ i \neq K}}^{C} p_i \, f(\underset{\sim}{x}|\pi_i)$$ is the smallest

i.e. $$\sum_{\substack{i=1 \\ i \neq K}}^{c} p_i f(\underset{\sim}{x}|\pi_i) \leq \sum_{\substack{i=1 \\ i \neq j}}^{c} p_i f(\underset{\sim}{x}|\pi_i) \quad -(*) \qquad \forall\, j \neq K$$

Subtracting $\sum_{\substack{i=1 \\ i \neq K,j}}^{c} p_i f(\underset{\sim}{x}|\pi_i)$ from both sides of $(*)$, we get

$$\sum_{\substack{i=1 \\ i \neq K}}^{c} p_i f(\underset{\sim}{x}|\pi_i) - \sum_{\substack{i=1 \\ i \neq K,j}}^{c} p_i f(\underset{\sim}{x}|\pi_i)$$

$$\leq \sum_{\substack{i=1 \\ i \neq j}}^{c} p_i f(\underset{\sim}{x}|\pi_i) - \sum_{\substack{i=1 \\ i \neq K,j}}^{c} p_i f(\underset{\sim}{x}|\pi_i)$$

i.e. $p_j f(\underset{\sim}{x}|\pi_j) \leq p_K f(\underset{\sim}{x}|\pi_K) \quad \forall\, j \neq K$

i.e the rule is

assign $\underset{\sim}{x}$ to π_K if

$p_K f(\underset{\sim}{x}|\pi_K)$ is largest

This is the same as TPM minimizing rule (what we expect under equal misclassification cost).

Further, the above is (under equal cost) equivalent to Bayes classifier.

Multiclass classification problem

Example 1: 3-class problem

Consider the following cost-prior table

	True membership π_1	π_2	π_3
π_1	$c(1\|1)=0$	$c(1\|2)=500$	100
Classified as π_2	10	0	50
π_3	50	200	0
prior prob	$p_1 = 0.05$	$p_2 = 0.60$	$p_3 = 0.35$

← misclassify, cost matrix

Let $\underset{\sim}{x}_0$ be a new obsn →

$$f(\underset{\sim}{x}_0 | \pi_1) = 0.01 \; ; \; f(\underset{\sim}{x}_0 | \pi_2) = 0.85$$
$$f(\underset{\sim}{x}_0 | \pi_3) = 2$$

ECM minimizing classifier

Compute $\sum_{\substack{i=1 \\ i \neq K}}^{3} p_i \, c(K|i) \, f(\underset{\sim}{x}_0 | \pi_i) \quad \forall K = 1, 2, 3$

$K = 1$: $\sum_{i=2}^{3} p_i \, c(1|i) \, f(\underset{\sim}{x}_0 | \pi_i)$

$$= p_2 \, c(1|2) \, f(\underset{\sim}{x}_0 | \pi_2) + p_3 \, f(\underset{\sim}{x}_0 | \pi_3) \, c(1|3)$$
$$= 325$$

$K=2$: $\sum_{\substack{i=1 \\ i\neq 2}}^{3} p_i\, c(2|i)\, f(\underset{\sim}{x}_0|\pi_i)$

$= p_1\, c(2|1)\, f(\underset{\sim}{x}_0|\pi_1) + p_3\, c(2|3)\, f(\underset{\sim}{x}_0|\pi_3)$

$= 35.06$

$K=3$: $\sum_{i=1}^{2} p_i\, c(3|i)\, f(\underset{\sim}{x}_0|\pi_i) = 102.03$

Since $\sum_{\substack{i=1 \\ i\neq 2}}^{3} p_i\, c(2|i)\, f(\underset{\sim}{x}_0|\pi_i)$ is the smallest

we assign $\underset{\sim}{x}_0$ to π_2

Note: Suppose in the same example, we take equal cost $\Rightarrow$ TPM minimizing rule

Calculate $p_1 f(\underset{\sim}{x}_0|\pi_1) = 0.0005$

$p_2 f(\underset{\sim}{x}_0|\pi_2) = 0.510$

$p_3 f(\underset{\sim}{x}_0|\pi_3) = 0.7$

Since $p_3 f(\underset{\sim}{x}_0|\pi_3) > p_i f(\underset{\sim}{x}_0|\pi_i)$ $i=1,2$

$\underset{\sim}{x}_0$ is allocated to π_3

Bayes classifier would also have done the same assignment

Example 3 : Discrete populations

Consider 3 bivariate discrete populations with the following j.p.m.f.s :

Π_1

x_2 \ x_1	1	2
1	0.5	0.2
2	0.1	0.2

Π_2

x_2 \ x_1	1	2
1	0.2	0.1
2	0.3	0.4

Π_3

x_2 \ x_1	1	2
1	0.25	0.25
2	0.25	0.25

Prior probabilities are equal, i.e. $p(\pi_i) = \frac{1}{3}$; $i = 1,2,3$ $(= p_i)$

Misclassification costs are

$$c(1|i) = 1 \; ; \; i = 2,3$$

$$c(2|i) = 2 \; ; \; i = 1,3$$

$$c(3|i) = 3 \; ; \; i = 1,2$$

Let us denote by $f_i(\underset{\sim}{x})$ to be class conditional prob for class π_i ; $i = 1,2,3$.

ECM classification rule is :

Assign $\underset{\sim}{x}$ to π_K if

$$\sum_{\substack{i=1 \\ i \neq K}}^{3} p_i \, f_i(\underset{\sim}{x}) \, c(K|i) \quad \text{is smallest}$$

Based on the above find rule for all pairs $\underset{\sim}{x}$

Pair (1,1)

$\underset{\sim}{x} = (1,1)$

K = 1

$$\sum_{i=2}^{3} p_i f_i(\underset{\sim}{x}) c(1|i)$$

$$= p_2 f_2(\underset{\sim}{x}) c(1|2) + p_3 f_3(\underset{\sim}{x}) c(1|3)$$

$$= \frac{1}{3}\left(f_2(\underset{\sim}{x}) + f_3(\underset{\sim}{x})\right)$$

$$= \frac{1}{3}(0.2 + 0.25) = 0.15$$

K = 2

$\underset{\sim}{x} = (1,1)$

$$\sum_{\substack{i=1 \\ i \neq 2}}^{3} p_i f_i(\underset{\sim}{x}) c(2|i)$$

$$= p_1 f_1(\underset{\sim}{x}) c(2|1) + p_3 f_3(\underset{\sim}{x}) c(2|3)$$

$$= \frac{2}{3}\left(f_1(\underset{\sim}{x}) + f_3(\underset{\sim}{x})\right) = \frac{2}{3}(0.5 + 0.25)$$

$$= 0.5$$

K = 3

$$\sum_{i=1}^{2} p_i f_i(\underset{\sim}{x}) c(3|i)$$

$$= p_1 f_1(\underset{\sim}{x}) c(3|1) + p_2 f_2(\underset{\sim}{x}) c(3|2)$$

$$= \frac{3}{3}\left(f_1(\underset{\sim}{x}) + f_2(\underset{\sim}{x})\right) = 0.7$$

$\sum_{\substack{i=1 \\ i \neq 1}}^{3} p_i f_i(\underset{\sim}{x}) c(1|i)$ is smallest

$\Rightarrow$ Assign $\underset{\sim}{x} = (1,1)$ to π_1

i.e. $(1,1) \in R_1$ partition

Pair (1,2)

$K=1 \rightarrow \frac{1}{3}\left(f_2(1,2) + f_3(1,2)\right) = \frac{1}{3}[0.55)$ ← smallest

$K=2 \rightarrow \frac{2}{3}\left(f_1(\overset{(1,2)}{\underset{\swarrow}{x}}) + f_3(1,2)\right) = \frac{2}{3}[0.35)$

$K=3 \rightarrow \frac{3}{3}\left(f_1(x,2) + f_2(1,2)\right) = 0.4$

$K=1$ gives smallest

$\Rightarrow (1,2) \in R_1$ partition

Pair (2,1)

$K=1 \rightarrow \frac{1}{3}\left(f_2(2,1) + f_3(2,1)\right) = \frac{1}{3}(0.35)$

$K=2 \rightarrow \frac{2}{3}\left(f_1(2,1) + f_3(2,1)\right) = \frac{2}{3}(0.45)$

$K=3 \rightarrow \frac{3}{3}\left(f_1(2,1) + f_2(2,1)\right) = 0.3$

$K=1$ gives smallest

$\Rightarrow (2,1) \in R_1$ partition

Pair (2,2)

$K=1 \rightarrow \frac{1}{3}\left(f_2(2,2) + f_3(2,2)\right) = \frac{1}{3}(0.65)$

$K=2 \rightarrow \frac{2}{3}\left(f_1(2,2) + f_3(2,2)\right) = \frac{2}{3}(0.45)$

$K=3 \rightarrow \frac{3}{3}\left(f_1(2,2) + f_2(2,2)\right) = 0.6$

$\Rightarrow (2,2) \in R_1$

$\Rightarrow$ ECM minimizing classification rule is

x_2 \ x_1	1	2
1	π_1	π_1
2	π_1	π_1.

Suppose we have a set $\mathcal{L}$ of preclassified examples

$$\mathcal{L} = \{((1,2),\pi_1), ((1,1),\pi_1), ((1,1),\pi_1), ((1,1),\pi_3), ((2,2),\pi_2), ((2,1),\pi_3), ((1,2),\pi_3), ((2,2),\pi_3)\}$$

If we apply ECM rule on $\mathcal{L}$, we get

classified into \ coming from	π_1	π_2	π_3
π_1	3	1	4
π_2	0	0	0
π_3	0	0	0

Misclassification rate would be $\frac{5}{8}$!!

↑ not good at all!

(i) Try to find the TPM opt rule

(ii) Try to calculate ECM/TPM of the two optimum rule

Example 2: Multiclass Gaussian

$$\pi_i \equiv N_p(\underset{\sim}{\mu}_i, \Sigma_i)$$

Assume for simplicity that $C(k|i)$ s are all same

We know that under this equal cost, the rule is

Assign $\underset{\sim}{x}$ to π_k if

$$p_k f(\underset{\sim}{x}|\pi_k) = \max_i p_i f(\underset{\sim}{x}|\pi_i)$$

i.e if $\log(p_k f(\underset{\sim}{x}|\pi_k)) = \max_i \log(p_i f(\underset{\sim}{x}|\pi_i))$

with $f(\underset{\sim}{x}|\pi_k) = (2\pi)^{-p/2}|\Sigma_k|^{-1/2}\exp\left(-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu}_k)'\Sigma_k^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}_k)\right)$

Define quadratic discriminant score for the ith popn, π_i, as

$$\delta_i^{QDS}(\underset{\sim}{x}) = -\frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu}_i)'\Sigma_i^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}_i) + \log p_i$$

Classification Rule

Assign $\underset{\sim}{x}$ to π_k if

$$\delta_k^{QDS}(\underset{\sim}{x}) = \max_i \delta_i^{QDS}(\underset{\sim}{x})$$

Note: With a given learning sample $\mathcal{L}$ of preclassified examples, we calculate the quadratic discriminant score for each class

$$\hat{\delta}_i^{QDS}(\underset{\sim}{x}) = -\frac{1}{2}\log|S_i| - \frac{1}{2}(\underset{\sim}{x} - \bar{\underset{\sim}{x}}_i)' S_i^{-1} (\underset{\sim}{x} - \bar{\underset{\sim}{x}}_i) + \log p_i$$

S_i : sample variance-covariance matrix based on preclassified examples corresponding to class π_i in $\mathcal{L}$

$\bar{\underset{\sim}{x}}_i$: sample mean vector based on preclassified examples corresponding to class π_i in $\mathcal{L}$.

Remark: Sp case: suppose π_i is $N_p(\underset{\sim}{\mu}_i, \Sigma)$ $i = 1(1)3$ $\Sigma > 0$

$$\log p_i f_i(\underset{\sim}{x}) = \underbrace{-\frac{1}{2}|\Sigma| - \frac{1}{2}\underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{x}}_{\text{same } \forall \text{ pop's}} + \underset{\sim}{\mu}_i'\Sigma^{-1}\underset{\sim}{x} - \frac{1}{2}\underset{\sim}{\mu}_i'\Sigma^{-1}\underset{\sim}{\mu}_i + \log p_i$$

Define the linear discriminant score as

$$\delta_i^{LDS}(\underset{\sim}{x}) = \underset{\sim}{\mu}_i'\Sigma^{-1}\underset{\sim}{x} - \frac{1}{2}\underset{\sim}{\mu}_i'\Sigma^{-1}\underset{\sim}{\mu}_i + \log p_i$$

Assignment rule is:

Assign $\underset{\sim}{x}$ to π_K if

$$\delta_K^{LDS}(\underset{\sim}{x}) = \max_i \delta_i^{LDS}(\underset{\sim}{x})$$

We have to use $\mathcal{L}$ to get the following estimated LDS

$$\hat{\delta}_i^{LDS}(\underset{\sim}{x}) = \bar{\underset{\sim}{x}}_i' S^{-1} \underset{\sim}{x} - \frac{1}{2} \bar{\underset{\sim}{x}}_i' S^{-1} \bar{\underset{\sim}{x}}_i + \log p_i$$

$\bar{\underset{\sim}{x}}_i$ for $i = 1(1)C$ calculated from pre-classified examples of π_i class in $\mathcal{L}$

$$\& \left(\sum_{i=1}^{C} n_i - C \right) S = \sum_{i=1}^{C} (n_i - 1) S_i$$

S_i for $i = 1(1)C$ calculated from pre-classified examples of π_i class in $\mathcal{L}$.

(divisor $(n_i - 1)$ sample covariance matrix)

Estimated classification rule is:

Assign $\underset{\sim}{x}$ to π_K if

$$\hat{\delta}_K^{LDS}(\underset{\sim}{x}) = \max_i \hat{\delta}_i^{LDS}(\underset{\sim}{x})$$

Note: The above rule is equivalent to

Assign $\underset{\sim}{x}$ to π_K if

$$-\frac{1}{2} (\underset{\sim}{x} - \bar{\underset{\sim}{x}}_K)' S^{-1} (\underset{\sim}{x} - \bar{\underset{\sim}{x}}_K) + \log p_K$$

i.e assign $\underset{\sim}{x}$ to pop^n closest in terms of sqd Mahalanobis distance penalised by $\log p_K$.

Note: We may also need to estimate p_i's from $\mathcal{L}$.