Discrimination and Classification

Discrimination

Situation:

We have two or more populations π_1, π_2, *etc* (possibly p-variate normal).

The populations are known – that is, we have data from each population.

We have data for a new case (population unknown) and we want to identify the population for which the new case is a member.

The Basic Problem for Two Populations

Suppose that the data from a new case

$\boldsymbol{X} = x_1, \ldots , x_p$ has joint density function either :

$\pi_1 : f_1(x_1, \ldots , x_p)$ or

$\pi_2 : f_2(x_1, \ldots , x_p)$

We want to make the decision to

D_1: Classify the case in π_1 (f_1 is the correct distribution) or

D_2: Classify the case in π_2 (f_2 is the correct distribution)

The Two Types of Errors

1. Misclassifying the case in π_1 when it lies in π_2.

Let $P(1|2) = P[D_1|\pi_2]$ = probability of this type of error

2. Misclassifying the case in π_2 when it lies in π_1.

Let $P(2|1) = P[D_2|\pi_1]$ = probability of this type of error

This is like **Type I** and **Type II** errors in hypothesis testing.

Note:

A discrimination scheme is defined by splitting p – dimensional space into two regions.

1. R_1 = the region were we make the decision D_1. (the decision to classify the case in π_1)

2. R_2 = the region were we make the decision D_2. (the decision to classify the case in π_2)

There can be several approaches to determining the regions R_1 and R_2. All concerned with considering the probabilities of misclassification $P(2|1)$ and $P(1|2)$

1. Set up the regions R_1 and R_2 so that one of the probabilities of misclassification, say $P(2|1)$, is at some low acceptable value α. The level of the other probability of misclassification is then just computed $P[1|2] = \beta$ (without trying to minimize it)

2. Set up the regions R_1 and R_2 so that the total probability of misclassification:

$$P[Misclassification] = P(1)P(2|1) + P(2)P(1|2)$$

is minimized, where $P(1)$ & $P(2)$ are prior probabilities of π_1 & π_2:

$$P[1] = P[\text{the case belongs to } \pi_1]$$

$$P[2] = P[\text{the case belongs to } \pi_2]$$

3. Set up the regions R_1 and R_2 so that the total expected *cost* of misclassification:

$$E[\text{Cost of } \mathit{Misclassification}] = \text{ECM}$$

$$= c_{2|1} P(1)\, P(2|1) + c_{1|2}\, P(2) P(1|2)$$

is minimized

$P[1] = P[\text{the case belongs to } \pi_1]$

$P[2] = P[\text{the case belongs to } \pi_2]$

$c_{2|1}$= the cost of misclassifying the case in π_2 when the case belongs to π_1.

$c_{1|2}$= the cost of misclassifying the case in π_1 when the case belongs to π_2.

The Optimal Classification Rule
(aka Neyman-Pearson Lemma)

Suppose that the data $x_1, \ldots, x_p$ has joint density function

$$f(x_1, \ldots, x_p; \theta)$$

where θ is either θ_1 or θ_2.

Let

$$f_1(x_1, \ldots, x_p) = f(x_1, \ldots, x_p; \theta_1) \text{ and}$$
$$f_2(x_1, \ldots, x_p) = f(x_1, \ldots, x_p; \theta_2)$$

We want to make the decision

D_1: $\theta = \theta_1$ (f_1 is the correct distribution) against

D_2: $\theta = \theta_2$ (f_2 is the correct distribution)

Result 11.1 on p. 581: The optimal regions (minimizing ECM, expected cost of misclassification) for making the decisions D_1 and D_2 respectively are R_1 and R_2 such that:

$$R_1 = \left\{(x_1, \dots, x_p): \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} \geq k\right\}$$

and

$$R_2 = \left\{(x_1, \dots, x_p): \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} < k\right\}$$

where

$$k = \frac{c_{1|2} P[2]}{c_{2|1} P[1]}$$

Fishers Linear Discriminant Function

Suppose that $x_1, \ldots, x_p$ is from a p-variate Normal distribution with mean vector:

$$\vec{\mu}_1 \text{ or } \vec{\mu}_2$$

The covariance matrix Σ is the same for both populations π_1 and π_2.

$$g(\vec{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)'\Sigma^{-1}(\vec{x}-\vec{\mu}_1)}$$

$$h(\vec{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_2)'\Sigma^{-1}(\vec{x}-\vec{\mu}_2)}$$

The Neymann-Pearson Lemma states that we should classify into populations π_1 and π_2 using:

$$\lambda = \frac{g(\vec{x})}{h(\vec{x})} = \frac{\dfrac{1}{(2\pi)^{p/2}\left|\Sigma\right|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)'\Sigma^{-1}(\vec{x}-\vec{\mu}_1)}}{\dfrac{1}{(2\pi)^{p/2}\left|\Sigma\right|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_2)'\Sigma^{-1}(\vec{x}-\vec{\mu}_2)}}$$

$$= e^{\frac{1}{2}(\vec{x}-\vec{\mu}_2)'\Sigma^{-1}(\vec{x}-\vec{\mu}_2)-\frac{1}{2}(\vec{x}-\vec{\mu}_1)'\Sigma^{-1}(\vec{x}-\vec{\mu}_1)}$$

That is make the decision

$$D_1 : \text{population is } \pi_1$$

if $\lambda \geq k$

or $\ln \lambda = \frac{1}{2}(\vec{x}-\vec{\mu}_2)'\Sigma^{-1}(\vec{x}-\vec{\mu}_2)-\frac{1}{2}(\vec{x}-\vec{\mu}_1)'\Sigma^{-1}(\vec{x}-\vec{\mu}_1) > \ln k$

or $$\left(\vec{x}-\vec{\mu}_2\right)'\Sigma^{-1}\left(\vec{x}-\vec{\mu}_2\right)-\left(\vec{x}-\vec{\mu}_1\right)'\Sigma^{-1}\left(\vec{x}-\vec{\mu}_1\right)>2\ln k$$

or $$\begin{aligned}&\vec{x}'\Sigma^{-1}\vec{x}-2\vec{\mu}_2'\Sigma^{-1}\vec{x}+\vec{\mu}_2'\Sigma^{-1}\vec{\mu}_2\\&\qquad-\vec{x}'\Sigma^{-1}\vec{x}+2\vec{\mu}_1'\Sigma^{-1}\vec{x}-\vec{\mu}_1'\Sigma^{-1}\vec{\mu}_1>2\ln k\end{aligned}$$

and

$$\left(\vec{\mu}_1-\vec{\mu}_2\right)'\Sigma^{-1}\vec{x}>\ln k+\tfrac{1}{2}\left(\vec{\mu}_1'\Sigma^{-1}\vec{\mu}_1-\vec{\mu}_2'\Sigma^{-1}\vec{\mu}_2\right)$$

Result 11.2 on p. 585: We make the decision

$$D_1 : \text{population is } \pi_1$$

if $\quad \vec{a}'\vec{x} > K$

where

$$\vec{a} = \Sigma^{-1}\left(\vec{\mu}_1 - \vec{\mu}_2\right) \text{ and } K = \ln k + \tfrac{1}{2}\left(\vec{\mu}_1'\Sigma^{-1}\vec{\mu}_1 - \vec{\mu}_2'\Sigma^{-1}\vec{\mu}_2\right)$$

and $\quad k = \dfrac{c_{1|2}P[2]}{c_{2|1}P[1]}$

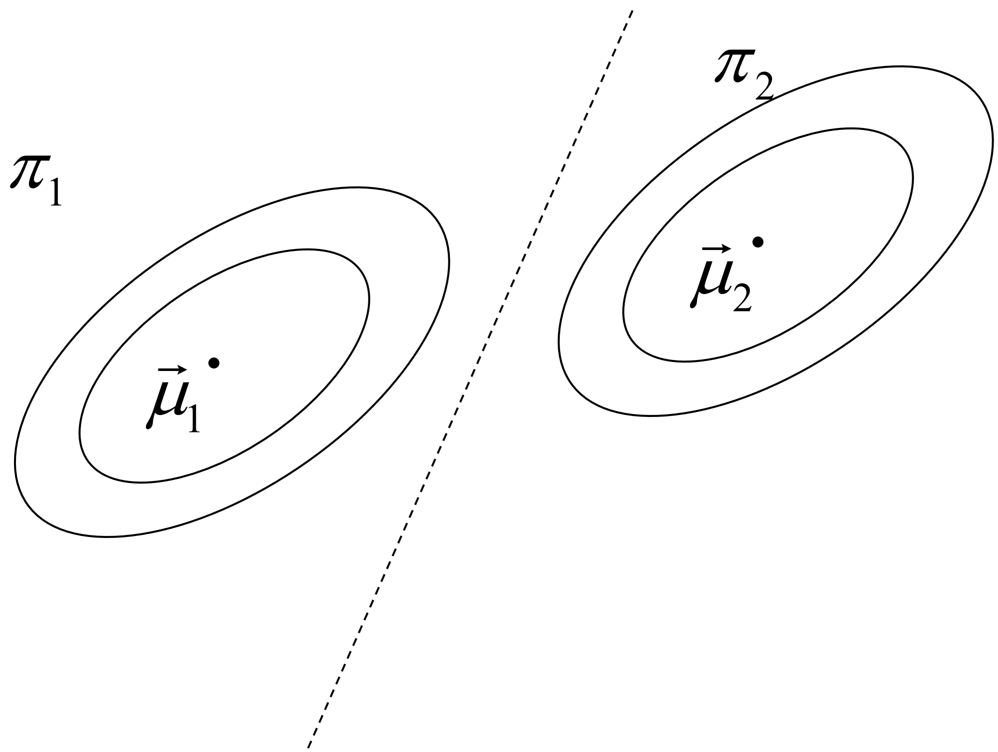
Note: $k = 1$ and $\ln k = 0$ if $c_{1|2} = c_{2|1}$ and $P[1] = P[2]$.

and $K = \tfrac{1}{2}\left(\vec{\mu}_1'\Sigma^{-1}\vec{\mu}_1 - \vec{\mu}_2'\Sigma^{-1}\vec{\mu}_2\right) = \tfrac{1}{2}\left(\vec{\mu}_1 - \vec{\mu}_2\right)'\Sigma^{-1}\left(\vec{\mu}_1 + \vec{\mu}_2\right)$

The function

$$\vec{a}'\vec{x} = \left(\vec{\mu}_1 - \vec{\mu}_2\right)' \Sigma^{-1}\vec{x}$$

Is called ***Fisher's linear discriminant function***

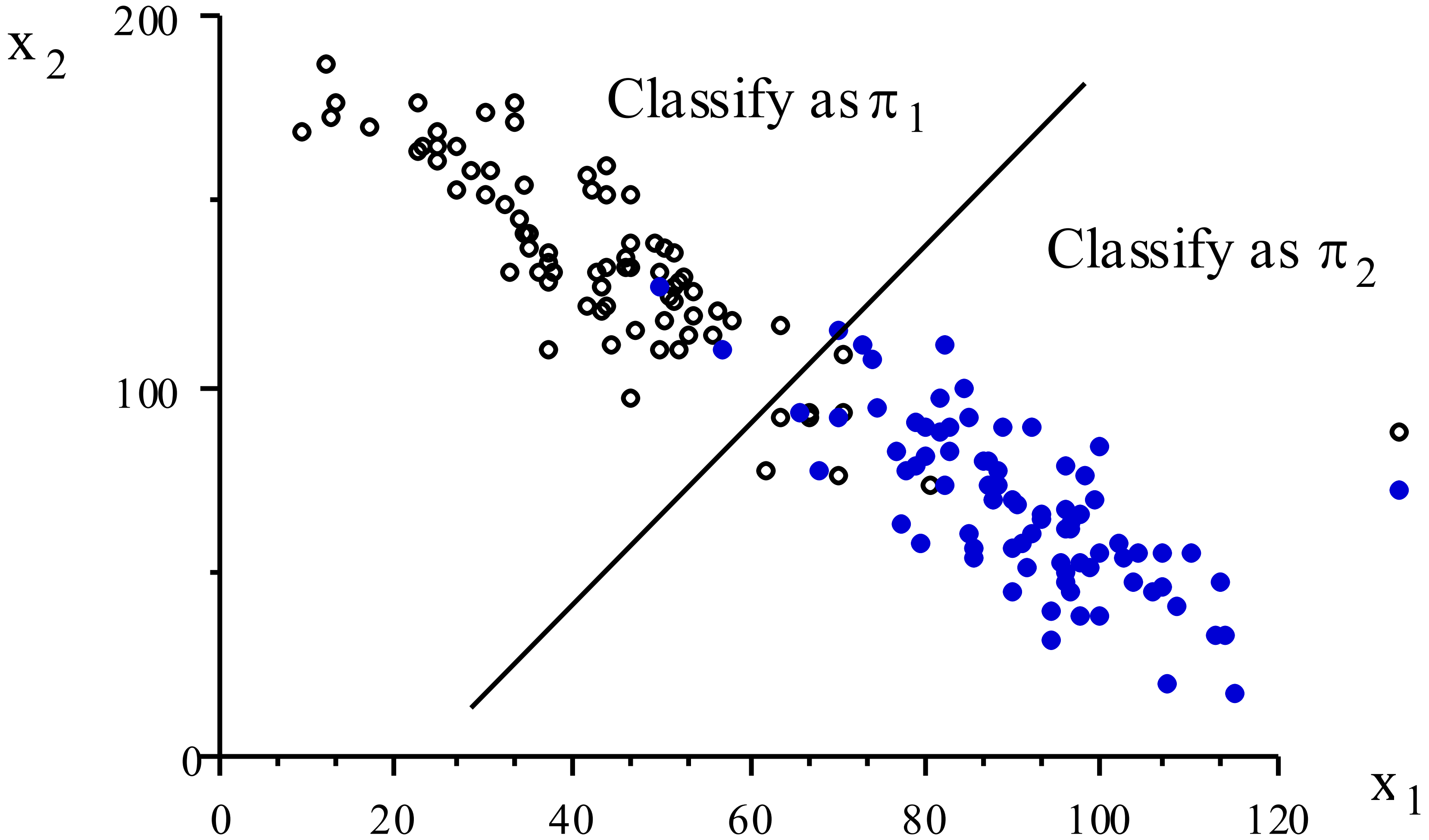


$$\vec{a}'\vec{x} = \left(\vec{\mu}_1 - \vec{\mu}_2\right)' \Sigma^{-1}\vec{x} = K_\alpha$$

In the case where the populations are unknown but estimated from data

Fisher's linear discriminant function

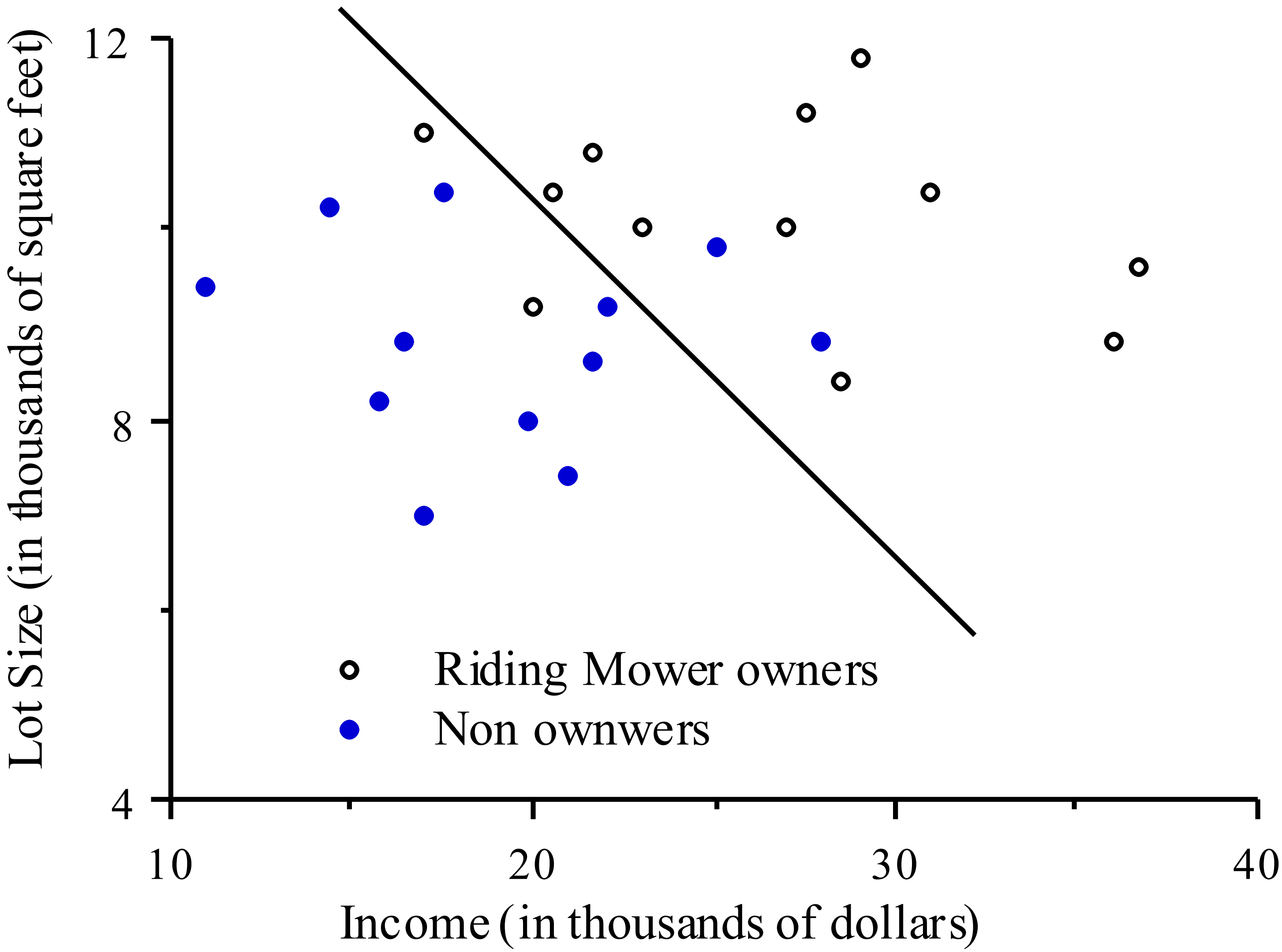
$$\hat{\vec{a}}'\vec{x} = \left(\bar{\vec{x}}_1 - \bar{\vec{x}}_2\right)' S^{-1}\vec{x}$$



A Pictorial representation of Fisher's procedure for two populations

Example 11.1 on p. 578

π_1 : Riding-mower owners		π_2 : Nonowners	
x_1 (Income in \$1000s)	x_2 (Lot size in 1000 sq ft)	x_1 (Income in \$1000s)	x_2 (Lot size in 1000 sq ft)
20.0	9.2	25.0	9.8
28.5	8.4	17.6	10.4
21.6	10.8	21.6	8.6
20.5	10.4	14.4	10.2
29.0	11.8	28.0	8.8
36.7	9.6	16.4	8.8
36.0	8.8	19.8	8.0
27.6	11.2	22.0	9.2
23.0	10.0	15.8	8.2
31.0	10.4	11.0	9.4
17.0	11.0	17.0	7.0
27.0	10.0	21.0	7.4



Lot Size (in thousands of square feet)
12
8
4
10
20
30
40
Income (in thousands of dollars)
Riding Mower owners
Non ownwers

Example 2

Annual financial data are collected for firms approximately 2 years prior to bankruptcy and for financially sound firms at about the same point in time. The data on the four variables

- x_1 = CF/TD = (cash flow)/(total debt),
- x_2 = NI/TA = (net income)/(Total assets),
- x_3 = CA/CL = (current assets)/(current liabilties, and
- x_4 = CA/NS = (current assets)/(net sales) are given in the following table.

The data are given in the following table:

	Bankrupt Firms				Nonbankrupt Firms			
	x_1	x_2	x_3	x_4		x_1	x_2	x_3	x_4
Firm	CF/TD	NI/TA	CA/CL	CA/NS	Firm	CF/TD	NI/TA	CA/CL	CA/NS
1	-0.4485	-0.4106	1.0865	0.4526	1	0.5135	0.1001	2.4871	0.5368
2	-0.5633	-0.3114	1.5314	0.1642	2	0.0769	0.0195	2.0069	0.5304
3	0.0643	0.0156	1.0077	0.3978	3	0.3776	0.1075	3.2651	0.3548
4	-0.0721	-0.0930	1.4544	0.2589	4	0.1933	0.0473	2.2506	0.3309
5	-0.1002	-0.0917	1.5644	0.6683	5	0.3248	0.0718	4.2401	0.6279
6	-0.1421	-0.0651	0.7066	0.2794	6	0.3132	0.0511	4.4500	0.6852
7	0.0351	0.0147	1.5046	0.7080	7	0.1184	0.0499	2.5210	0.6925
8	-0.6530	-0.0566	1.3737	0.4032	8	-0.0173	0.0233	2.0538	0.3484
9	0.0724	-0.0076	1.3723	0.3361	9	0.2169	0.0779	2.3489	0.3970
10	-0.1353	-0.1433	1.4196	0.4347	10	0.1703	0.0695	1.7973	0.5174
11	-0.2298	-0.2961	0.3310	0.1824	11	0.1460	0.0518	2.1692	0.5500
12	0.0713	0.0205	1.3124	0.2497	12	-0.0985	-0.0123	2.5029	0.5778
13	0.0109	0.0011	2.1495	0.6969	13	0.1398	-0.0312	0.4611	0.2643
14	-0.2777	-0.2316	1.1918	0.6601	14	0.1379	0.0728	2.6123	0.5151
15	0.1454	0.0500	1.8762	0.2723	15	0.1486	0.0564	2.2347	0.5563
16	0.3703	0.1098	1.9914	0.3828	16	0.1633	0.0486	2.3080	0.1978
17	-0.0757	-0.0821	1.5077	0.4215	17	0.2907	0.0597	1.8381	0.3786
18	0.0451	0.0263	1.6756	0.9494	18	0.5383	0.1064	2.3293	0.4835
19	0.0115	-0.0032	1.2602	0.6038	19	-0.3330	-0.0854	3.0124	0.4730
20	0.1227	0.1055	1.1434	0.1655	20	0.4875	0.0910	1.2444	0.1847
21	-0.2843	-0.2703	1.2722	0.5128	21	0.5603	0.1112	4.2918	0.4443
					22	0.2029	0.0792	1.9936	0.3018
					23	0.4746	0.1380	2.9166	0.4487
					24	0.1661	0.0351	2.4527	0.1370
					25	0.5808	0.0371	5.0594	0.1268

Note: $k = 1$ and $\ln k = 0$ if $c_{1|2} = c_{2|1}$ and $P[1] = P[2]$.

and $K = \frac{1}{2}\left(\vec{\mu}_1'\Sigma^{-1}\vec{\mu}_1 - \vec{\mu}_2'\Sigma^{-1}\vec{\mu}_2\right) = \frac{1}{2}\left(\vec{\mu}_1 - \vec{\mu}_2\right)'\Sigma^{-1}\left(\vec{\mu}_1 + \vec{\mu}_2\right)$

Thus $\vec{a}'\vec{x} > K$ with

$\vec{a} = \Sigma^{-1}\left(\vec{\mu}_1 - \vec{\mu}_2\right)$ and $K = \ln k + \frac{1}{2}\left(\vec{\mu}_1'\Sigma^{-1}\vec{\mu}_1 - \vec{\mu}_2'\Sigma^{-1}\vec{\mu}_2\right)$

is equivalent to

$$\left(\vec{\mu}_1 - \vec{\mu}_2\right)'\Sigma^{-1}\vec{x} > \frac{1}{2}\left[\vec{\mu}_1'\Sigma^{-1}\vec{\mu}_1\right] - \frac{1}{2}\left[\vec{\mu}_2'\Sigma^{-1}\vec{\mu}_2\right]$$

or

$$\frac{1}{2}\left[\vec{\mu}_2'\Sigma^{-1}\vec{\mu}_2\right] - \vec{\mu}_2'\Sigma^{-1}\vec{x} > \frac{1}{2}\left[\vec{\mu}_1'\Sigma^{-1}\vec{\mu}_1\right] - \vec{\mu}_1'\Sigma^{-1}\vec{x}$$

$$\frac{1}{2}\left[\vec{\mu}_2'\Sigma^{-1}\vec{\mu}_2\right] - \vec{\mu}_2'\Sigma^{-1}\vec{x} + \frac{1}{2}\vec{x}'\Sigma^{-1}\vec{x} > \frac{1}{2}\left[\vec{\mu}_1'\Sigma^{-1}\vec{\mu}_1\right] - \vec{\mu}_1'\Sigma^{-1}\vec{x} + \frac{1}{2}\vec{x}'\Sigma^{-1}\vec{x}$$

$$\left(\vec{x} - \vec{\mu}_2\right)'\Sigma^{-1}\left(\vec{x} - \vec{\mu}_2\right) > \left(\vec{x} - \vec{\mu}_1\right)'\Sigma^{-1}\left(\vec{x} - \vec{\mu}_1\right)$$

$$\text{Mahalanobis distance}\left(\vec{x}, \vec{\mu}_2\right) > \text{Mahalanobis distance}\left(\vec{x}, \vec{\mu}_1\right)$$

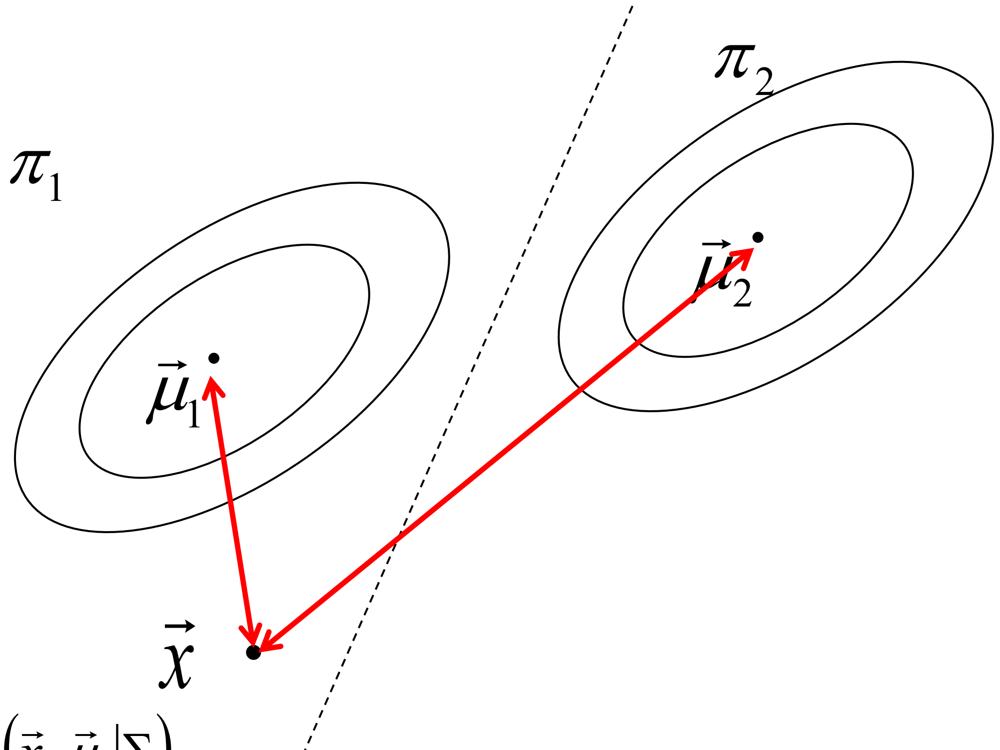
$$d_M^2\left(\vec{x}, \vec{\mu}_2 \middle| \Sigma\right) > d_M^2\left(\vec{x}, \vec{\mu}_1 \middle| \Sigma\right)$$

Thus we make the decision

$$D_1 : \text{population is } \pi_1$$

if

$$\text{Mahalanobis distance}(\vec{x}, \vec{\mu}_2) > \text{Mahalanobis distance}(\vec{x}, \vec{\mu}_1)$$



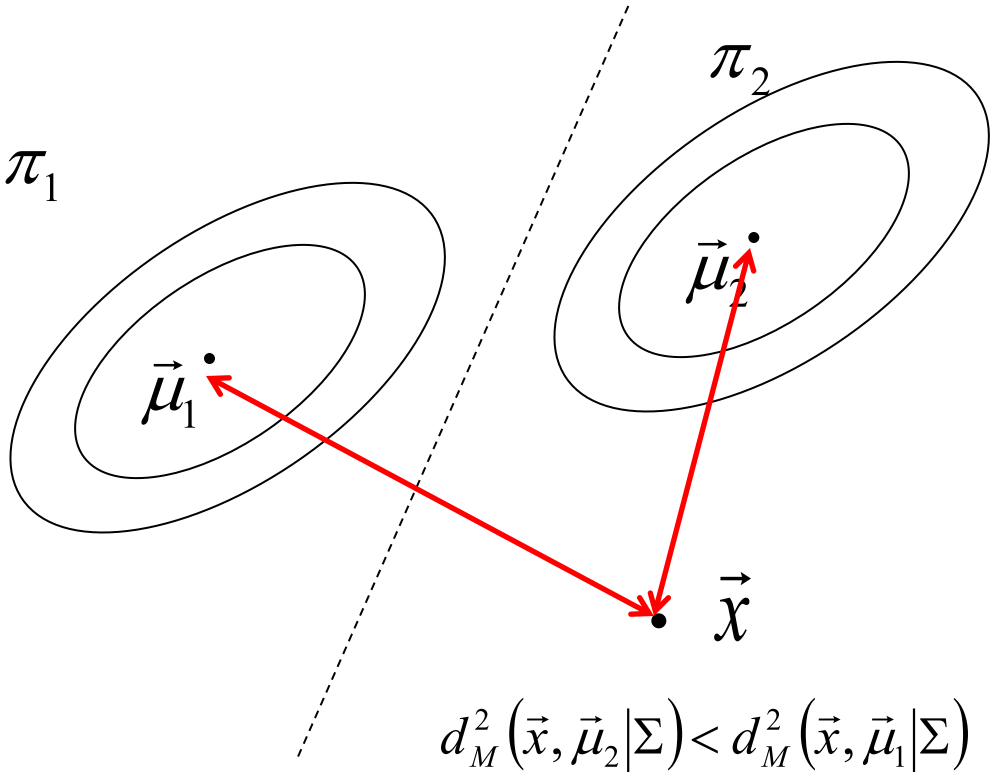
$$d_M^2(\vec{x}, \vec{\mu}_2 | \Sigma) > d_M^2(\vec{x}, \vec{\mu}_1 | \Sigma)$$

Thus we make the decision

$$D_2 : \text{population is } \pi_2$$

if

$$\text{Mahalanobis distance}\left(\vec{x}, \vec{\mu}_2\right) < \text{Mahalanobis distance}\left(\vec{x}, \vec{\mu}_1\right)$$



$$d_M^2\left(\vec{x}, \vec{\mu}_2 \middle| \Sigma\right) < d_M^2\left(\vec{x}, \vec{\mu}_1 \middle| \Sigma\right)$$

Thus we make the decision

$$D_1 : \text{population is } \pi_1$$

if

$$\text{Mahalanobis distance}\left(\vec{x}, \vec{\mu}_2\right) > \text{Mahalanobis distance}\left(\vec{x}, \vec{\mu}_1\right)$$

where

$$\vec{a} = \Sigma^{-1}\left(\vec{\mu}_1 - \vec{\mu}_2\right) \text{ and } K = \ln k + \tfrac{1}{2}\left(\vec{\mu}_1'\Sigma^{-1}\vec{\mu}_1 - \vec{\mu}_2'\Sigma^{-1}\vec{\mu}_2\right)$$

and $$k = \frac{c_{1|2}P[2]}{c_{2|1}P[1]}$$

Note: $k = 1$ and $\ln k = 0$ if $c_{1|2} = c_{2|1}$ and $P[1] = P[2]$.

$$\text{and } K = \tfrac{1}{2}\left(\vec{\mu}_1'\Sigma^{-1}\vec{\mu}_1 - \vec{\mu}_2'\Sigma^{-1}\vec{\mu}_2\right) = \tfrac{1}{2}\left(\vec{\mu}_1 - \vec{\mu}_2\right)'\Sigma^{-1}\left(\vec{\mu}_1 + \vec{\mu}_2\right)$$

Classification of *p*-variate Normal Distributions When $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$

Suppose that $x_1, \ldots, x_p$ are data from either of a p-variate normal distribution with mean vectors

$$\boldsymbol{\mu}_1 \text{ or } \boldsymbol{\mu}_2$$

and covariance matrices, Σ_1 and Σ_2 respectively.

That is,

$$f_1(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}_1|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)'\mathbf{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}$$

$$f_2(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}_2|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)'\mathbf{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)}$$

Thus, the covariance matrices, as well as the mean vectors, are different from one another for the two populations.

The optimal rule states that we should classify into populations π_1 and π_2 using:

$$\lambda = \frac{f(\vec{x})}{g(\vec{x})} = \frac{\dfrac{1}{(2\pi)^{p/2}\left|\Sigma_1\right|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_1)'\Sigma_1^{-1}(\vec{x}-\vec{\mu}_1)}}{\dfrac{1}{(2\pi)^{p/2}\left|\Sigma_2\right|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_2)'\Sigma_2^{-1}(\vec{x}-\vec{\mu}_2)}}$$

$$= \frac{\left|\Sigma_2\right|^{1/2}}{\left|\Sigma_1\right|^{1/2}} e^{\frac{1}{2}(\vec{x}-\vec{\mu}_2)'\Sigma_2^{-1}(\vec{x}-\vec{\mu}_2)-\frac{1}{2}(\vec{x}-\vec{\mu}_1)'\Sigma_1^{-1}(\vec{x}-\vec{\mu}_1)}$$

That is make the decision

$$D_1 : \text{population is } \pi_1$$

if $\lambda \geq k$

$$\ln\lambda = \tfrac{1}{2}\left[(\vec{x}-\vec{\mu}_2)'\Sigma_2^{-1}(\vec{x}-\vec{\mu}_2)-(\vec{x}-\vec{\mu}_1)'\Sigma_1^{-1}(\vec{x}-\vec{\mu}_1)+\ln\left|\Sigma_2\right|-\ln\left|\Sigma_1\right|\right] \geq \ln k$$

or

$$\tfrac{1}{2}\vec{x}'\left(\Sigma_2^{-1}-\Sigma_1^{-1}\right)\vec{x}+\left(\vec{\mu}_1'\Sigma_1^{-1}-\vec{\mu}_2'\Sigma_2^{-1}\right)\vec{x}-K\geq\ln k$$

where

$$K=\frac{1}{2}\ln\frac{\left|\Sigma_2\right|}{\left|\Sigma_1\right|}+\frac{1}{2}\left(\vec{\mu}_1'\Sigma_1^{-1}\vec{\mu}_1-\vec{\mu}_2'\Sigma_2^{-1}\vec{\mu}_2\right)'$$

and

$$k=\frac{c_{1|2}P\left[2\right]}{c_{2|1}P\left[1\right]}$$

Summarizing we make the decision to classify in population p_1 if:

$$\vec{x}' A\vec{x} + \vec{b}'\vec{x} + c \geq 0$$

where

$$A = \tfrac{1}{2}\left(\Sigma_2^{-1} - \Sigma_1^{-1}\right)$$

$$\vec{b} = \Sigma_1^{-1}\vec{\mu}_1 - \Sigma_2^{-1}\vec{\mu}_2$$

and

$$c = -\frac{1}{2}\ln\frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2}\left(\vec{\mu}_1'\Sigma_1^{-1}\vec{\mu}_1 - \vec{\mu}_2'\Sigma_2^{-1}\vec{\mu}_2\right) - \ln\frac{c_{1|2}P[2]}{c_{2|1}P[1]}$$

Discrimination of p-variate Normal distributions

(unequal Covariance matrices)