S1 Sains Data
FMIPA UNESA

Analisis Multivariat – Pertemuan 10

Discriminant Analysis

Prodi S1 Sains Data
Universitas Negeri Surabaya
15 April 2025

Introduction

Table 11.1

π_1: Riding-mower owners		π_2: Nonowners	
x_1 (Income in \$1000s)	x_2 (Lot size in 1000 ft^2)	x_1 (Income in \$1000s)	x_2 (Lot size in 1000 ft^2)
90.0	18.4	105.0	19.6
115.5	16.8	82.8	20.8
94.8	21.6	94.8	17.2
91.5	20.8	73.2	20.4
117.0	23.6	114.0	17.6
140.1	19.2	79.2	17.6
138.0	17.6	89.4	16.0
112.8	22.4	96.0	18.4
99.0	20.0	77.4	16.4
123.0	20.8	63.0	18.8
81.0	22.0	81.0	14.0
111.0	20.0	93.0	14.8

A good classification procedure should result in few misclassifications

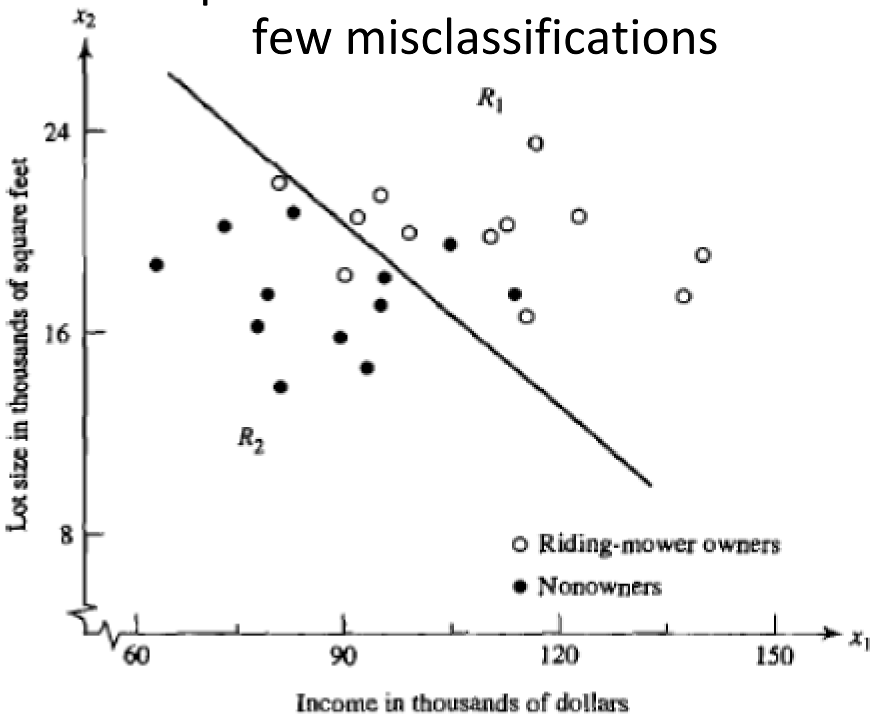
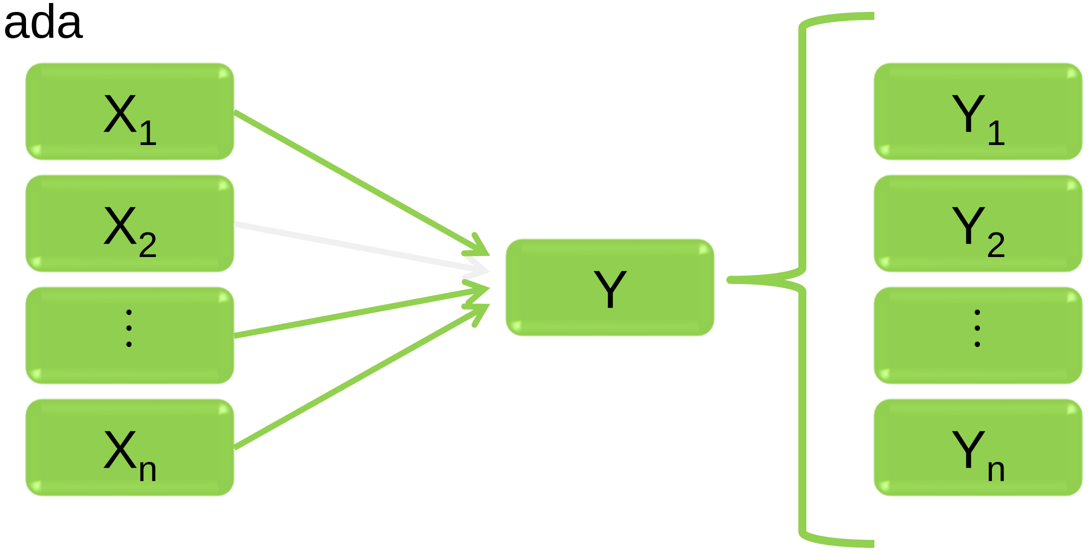


Figure 11.1 Income and lot size for riding-mower owners and nonowners.

Analisis Diskriminan

satu metode statistika multivariat yang digunakan untuk mengklasifikasikan beberapa variabel prediktor ke dalam observasi pada kelompok khusus dan yang sudah ada



$y \longrightarrow$ Kategori, nominal

$x \longrightarrow$ Interval, rasio

Analisis Diskriminan

Metode dependensi

Tujuan: Mendapatkan suatu fungsi yang disebut dengan fungsi diksrikminan yang dapat memisahkan objek sesuai dengan grupnya. Fungsi juga dapat digunakan untuk memprediksi grup dari suatu objek baru yang diamati (Sharma, 1996)

k populasi: $\pi_1, \pi_2, \ldots, \pi_k$

Probabilitas densitas populasi ke-i

$f_i(\mathbf{x})$

Ruang sekat yang memisahkan populasi

$$R_1, R_2, \ldots., R_k$$

$(x_1, x_2, \ldots, x_p)$

Setiap observasi hanya Dikelompokkan dalam satu kelas

$$R_i \longrightarrow \pi_i$$

Asumsi

Asumsi

- Populasi berdistribusi Multivariat Normal
- Kehomogenan varians

Bayes' Rule Applied to the Classification Problem

- Suppose that we have *g* populations (groups) and that the *i*th population is denoted as πi.
- We are interested in P(πi|x), the conditional probability that observation came from population πi given that the observed values of the multivariate vector of variables x.
- **We will classify an observation to the population for which the value of** *P*(πi|x) **is greatest**. This is the most probable group given the observed values of x.

$$P(\text{member of } \pi_i | \text{ we observed } \mathbf{x}) = \frac{P(\text{member of } \pi_i \text{ and we observe } \mathbf{x})}{P(\text{we observe } \mathbf{x})}$$

$$p(\pi_i|\mathbf{x}) = \frac{p_i f(\mathbf{x}|\pi_i)}{\sum_{j=1}^{g} p_j f(\mathbf{x}|\pi_j)}$$

Decision Rule – Two Population

We would classify to population 1 when

$$\frac{p_1 f(\mathbf{x}|\pi_1)}{p_2 f(\mathbf{x}|\pi_2)} > 1$$

$$\frac{f(\mathbf{x}|\pi_1)}{f(\mathbf{x}|\pi_2)} > \frac{p_2}{p_1}$$

Linear Discriminant Analysis

We assume that in population $\boldsymbol{\pi_i}$ the probability density function of $\boldsymbol{x}$ is multivariate normal with mean vector $\boldsymbol{\mu_i}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ (same for all populations). As a formula, this is...

$$f_x = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$f(\mathbf{x}|\pi_i) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu_\mathrm{i})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu_\mathrm{i})\right)$$

Simple Example

s.no	Height (cm)	Gender
1	140	F
2	145	F
3	135	F
4	169	M
5	165	F
6	142	M
7	168	M
8	141	F
9	159	F
10	160	M
11	172	M

Predicting the gender of a person through his/her height

The question ?

What will be gender of a person whose height is 152 cm?

Simple Example – Solution

$$p(\pi_i|\mathbf{x}) = \frac{p_i f(\mathbf{x}|\pi_i)}{\sum_{j=1}^{g} p_j f(\mathbf{x}|\pi_j)}$$

$$\boldsymbol{P(\, gender = male \mid height = 152\,)}$$

$$= \frac{P(\,height = 152 \mid gender = male\,)P(gender = male)}{P(\,height = 152 \mid gender = male\,)P(gender = male) + P(\,height = 152 \mid gender = female\,)P(gender = female)}$$

$$\boldsymbol{P(\, gender = female \mid height = 152\,)}$$

$$= \frac{P(\,height = 152 \mid gender = female\,)P(gender = female)}{P(\,height = 152 \mid gender = male\,)P(gender = male) + P(\,height = 152 \mid gender = female\,)P(gender = female)}$$

s.no	Height (cm)	Gender
1	140	F
2	145	F
3	135	F
4	169	M
5	165	F
6	142	M
7	168	M
8	141	F
9	159	F
10	160	M
11	172	M

$$f_x = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Male Class	Female Class
$\sigma^2 = 117$ ($\sigma = 10.8$)	$\sigma^2 = 117$ ($\sigma = 10.8$)
$\mu_M = 162.2$	$\mu_F = 147.33$
Probability distribution = $0.037e^{-\frac{1}{2}\frac{(x-162.2)^2}{117}}$	Probability distribution = $0.037e^{-\frac{1}{2}\frac{(x-147.33)^2}{117}}$
Probability when x = 152 = 0.0237 (substitute x = 152 in the dist. equation)	Probability when x = 152 = 0.033 (substitute x = 152 in the dist. equation)

$$P(\text{gender} = \text{male} \mid \text{height} = x) = \frac{0.037e^{\frac{(x-162.2)^2}{117}} * 0.454}{0.037e^{\frac{(x-162.2)^2}{117}} * 0.454 + 0.037e^{\frac{(x-147.33)^2}{117}} * 0.545}$$

$$P(\,gender = male \mid height = 152\,) = \frac{0.0237x0.454}{0.0237x0.454 + 0.033x0.545} = 0.37$$

$$P(\,gender = female \mid height = 152\,) = \frac{0.033x0.545}{0.0237x0.454 + 0.033x0.545} = 0.625$$

we can classify the height of 152 cm in the female class.

Expected Cost Misclassification (ECM)

- There are additional features that an “optimal” classification rule should possess.
- Classification schemes are often evaluated in terms of their misclassification.
  - A good classification procedure should result in few misclassifications.
- The idea is to create a rule that minimizes the chances of making these mistakes.
  - The **probabilities of misclassification** should be small
  - An optimal classification procedure should, whenever possible, account for the **costs** associated with misclassification
- An optimal classification rule should take “prior probabilities of occurrence” into account.

Bayes' Rule (Review)

Consider any two events *A* and *B*. To find P(B|A), the probability that *B* occurs given that *A* has occurred, Bayes' Rule states the following.

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

$$P(A \, and \, B) = P(A)P(B|A)$$

Probability of B occurring given evidence A has already occurred

Probability of A occurring

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{B}|A) \cdot P(A)}{P(B)}$$

Probability of A occurring given evidence B has already occurred

Probability of B occurring

Expected Cost Misclassification (ECM)

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

$$ECM = c(2|1)P(2|1)p1 + c(1|2)P(1|2)p2$$

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

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

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

Decision Rule that minimize the ECM

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

$$\left(\begin{matrix}\text{density}\\\text{ratio}\end{matrix}\right) \geq \left(\begin{matrix}\text{cost}\\\text{ratio}\end{matrix}\right)\left(\begin{matrix}\text{prior}\\\text{probability}\\\text{ratio}\end{matrix}\right)$$

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

$$\left(\begin{matrix}\text{density}\\\text{ratio}\end{matrix}\right) < \left(\begin{matrix}\text{cost}\\\text{ratio}\end{matrix}\right)\left(\begin{matrix}\text{prior}\\\text{probability}\\\text{ratio}\end{matrix}\right)$$

Example – ECM Rule

Goal : Classifying a new observation into one of the two populations

- Suppose c(2|1) = 5 units and c(112) = 10 units.
- It is known that about 20% of all objects (for which the measurements x can be recorded) belong to population 2.
    - Thus, the prior probabilities are p1 = 0.8 and p2 = 0.2.
- Suppose the density functions evaluated at a new observation Xo give f1(x0) = 0.3 and f2(x0) = 0.4.

Do we classify the new observation as population 1 or 2?

$$\frac{f_1(\mathbf{x}_0)}{f_2(\mathbf{x}_0)} = .75 > \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right) = .5$$

classify it as belonging to population 1

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)

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

(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)

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

The Estimated Minimum ECM Rule

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

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

Decision Rule that minimize the ECM

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

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

Classification Rule

Allocate $\mathbf{x}_0$ to π_1 if

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

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

Allocate $\mathbf{x}_0$ to π_1 if

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

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

Fisher's Approach

Fisher's rule is equivalent to the minimum ECM rule with equal prob and equal cost of misclassification. Two normal populations have the same covariance matrix.

Allocate $\mathbf{x}_0$ to π_1 if

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

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

If, in (11-18),

$$\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right) = 1$$

$$\underset{(p\times 1)}{\bar{\mathbf{x}}_1} = \frac{1}{n_1}\sum_{j=1}^{n_1}\mathbf{x}_{1j}, \quad \underset{(p\times p)}{\mathbf{S}_1} = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)'$$

$$\underset{(p\times 1)}{\bar{\mathbf{x}}_2} = \frac{1}{n_2}\sum_{j=1}^{n_2}\mathbf{x}_{2j}, \quad \underset{(p\times p)}{\mathbf{S}_2} = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)'$$

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

Discriminant Function

$$\boxed{\hat{y} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pooled}}^{-1}\mathbf{x} = \hat{\mathbf{a}}'\mathbf{x}}$$

Critical cutting score

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

$$\bar{y}_1 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pooled}}^{-1}\bar{\mathbf{x}}_1 = \hat{\mathbf{a}}'\bar{\mathbf{x}}_1$$

$$\bar{y}_2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pooled}}^{-1}\bar{\mathbf{x}}_2 = \hat{\mathbf{a}}'\bar{\mathbf{x}}_2$$

Classification Rule

Allocate $\mathbf{x}_0$ to π_1 if $\hat{y}_0 = \hat{\mathbf{a}}'\mathbf{x}_0 \geq \hat{m}$

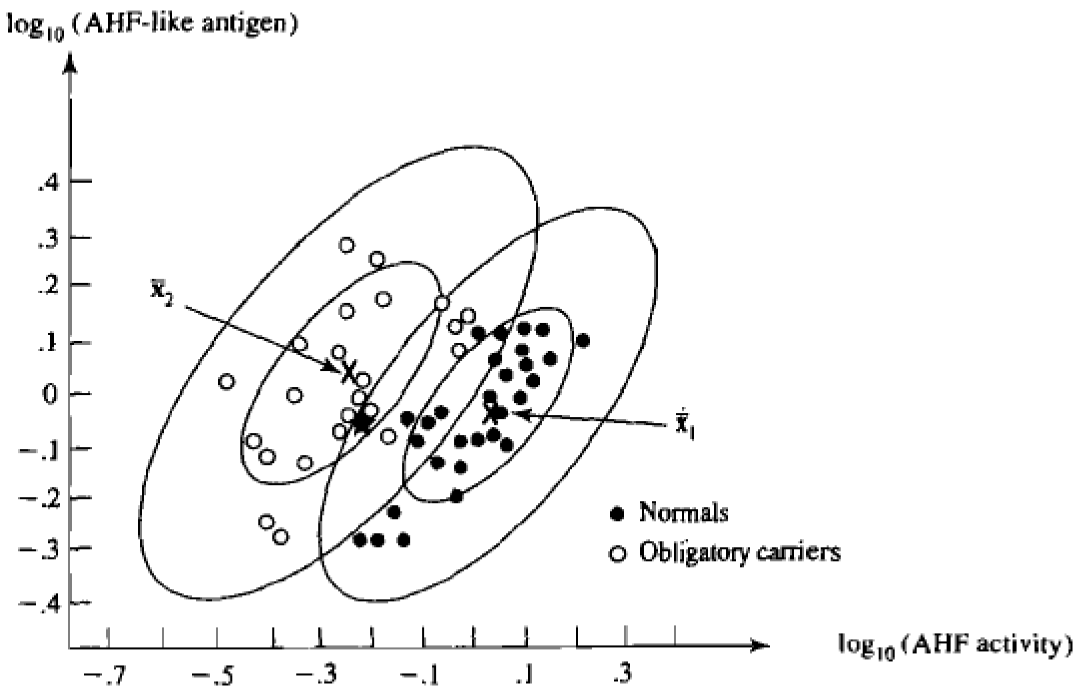
Allocate $\mathbf{x}_0$ to π_2 if $\hat{y}_0 = \hat{\mathbf{a}}'\mathbf{x}_0 < \hat{m}$

Example - Fisher

Detection of hemophilia A carriers by two variables $\log_{10}(\text{AHF activity})$, $\log_{10}(\text{AHF-like antigen})$

normal group carrier group

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} -.0065 \\ -.0390 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} -.2483 \\ .0262 \end{bmatrix} \quad \mathbf{S}_{\text{pooled}}^{-1} = \begin{bmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{bmatrix}$$



discriminant function

$$\hat{y} = \hat{\mathbf{a}}'\mathbf{x} = [\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2]'\mathbf{S}_{\text{pooled}}^{-1}\mathbf{x}$$
$$= [.2418 \quad -.0652]\begin{bmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
$$= 37.61x_1 - 28.92x_2$$

critical cutting score

$$\bar{y}_1 = \hat{\mathbf{a}}'\bar{\mathbf{x}}_1 = [37.61 \quad -28.92]\begin{bmatrix} -.0065 \\ -.0390 \end{bmatrix} = .88$$

$$\bar{y}_2 = \hat{\mathbf{a}}'\bar{\mathbf{x}}_2 = [37.61 \quad -28.92]\begin{bmatrix} -.2483 \\ .0262 \end{bmatrix} = -10.10$$

$$\boxed{\hat{m} = \tfrac{1}{2}(\bar{y}_1 + \bar{y}_2) = \tfrac{1}{2}(.88 - 10.10) = -4.61}$$

Allocate $\mathbf{x}_0$ to π_1 if $\hat{y}_0 = \hat{\mathbf{a}}'\mathbf{x}_0 \geq \hat{m} = -4.61$

Allocate $\mathbf{x}_0$ to π_2 if $\hat{y}_0 = \hat{\mathbf{a}}'\mathbf{x}_0 < \hat{m} = -4.61$

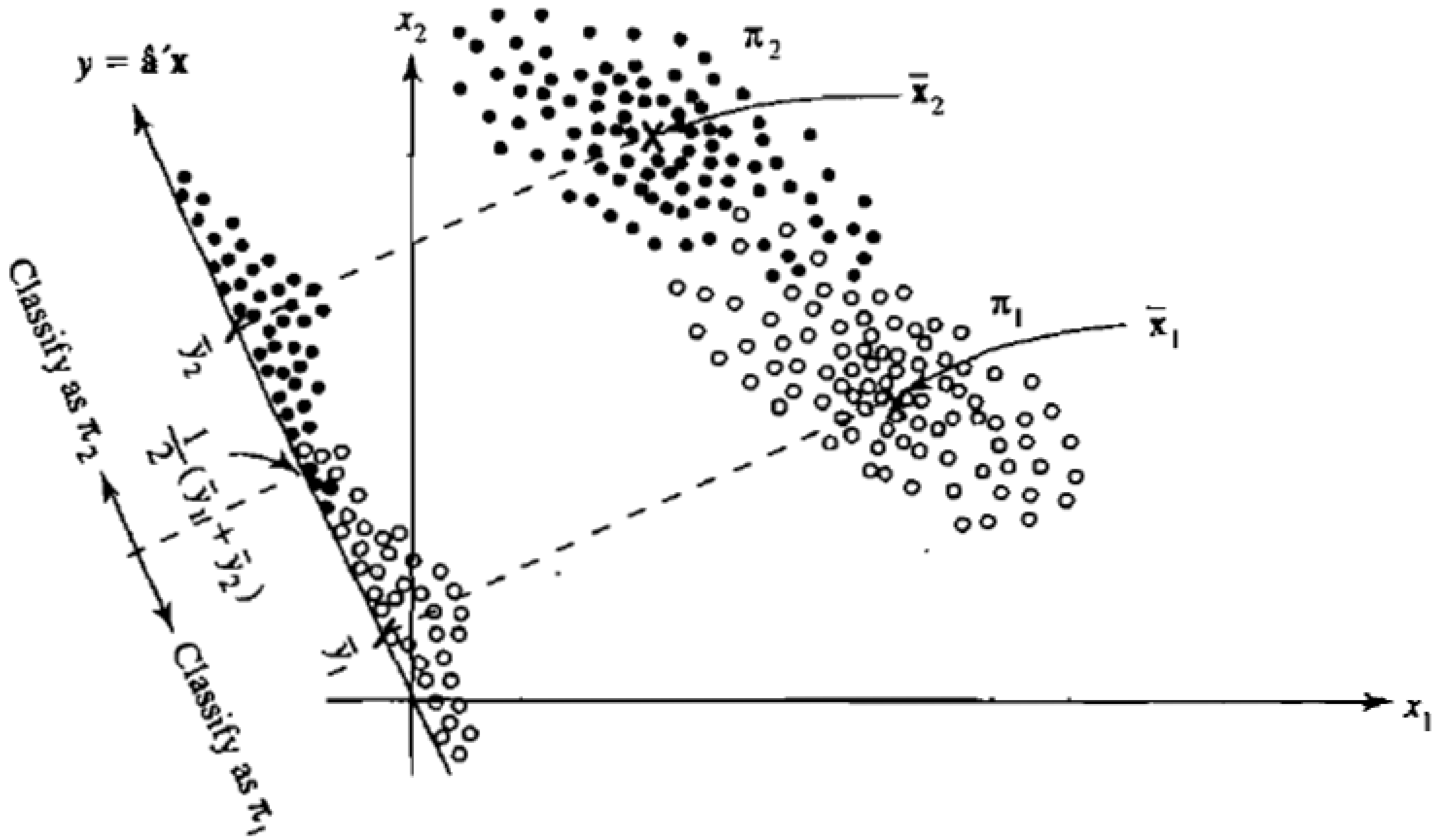
Example – Classification New Data

Measurements of AHF activity and AHF-like antigen on a woman who may be a hemophilia A carrier give X1 = -.210 and *X2* = - .044. Should this woman be classified as normal or obligatory carrier?

$$\hat{y}_0 = \hat{\mathbf{a}}'\mathbf{x}_0 = [37.61 \quad -28.92]\begin{bmatrix} -.210 \\ -.044 \end{bmatrix} = -6.62 < -4.61$$

obligatory carrier

A pictorial representation of Fisher's procedure for two populations with p = 2.



Evaluating Classification Functions

Apparent error rate (APER), is defined as the fraction of observations in the training sample that are misclassified by the sample classification function

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

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

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

Calculating the APER

		Predicted membership		
		π_1: riding-mower owners	π_2: nonowners	
Actual membership	π_1: riding-mower owners	$n_{1C} = 10$	$n_{1M} = 2$	$n_1 = 12$
	π_2: nonowners	$n_{2M} = 2$	$n_{2C} = 10$	$n_2 = 12$

$$\text{APER} = \left(\frac{2+2}{12+12}\right)100\% = \left(\frac{4}{24}\right)100\% = 16.7\%$$

Assignment (Classifying Alaskan and Canadian salmon)

- Lakukan klasifikasi pada data berikut menggunakan EXCEL.
- Buat plot visualisasi (scatter plot) terlebih dahulu
- Gunakan hanya variabel **freshwater** dan **marine**
- Lakukan perhitungan detail untuk mendapatkan fungsi diskriminan, critical cutting score, confusion matrix, dan APER

Table 11.2 Salmon Data (Growth-Ring Diameters)

Alaskan			Canadian		
Gender	Freshwater	Marine	Gender	Freshwater	Marine
2	108	368	1	129	420
1	131	355	1	148	371
1	105	469	1	179	407
2	86	506	2	152	381
1	99	402	2	166	377
2	87	423	2	124	389
1	94	440	1	156	419
2	117	489	2	131	345
2	79	432	1	140	362
1	99	403	2	144	345
1	114	428	2	149	393
2	123	372	1	108	330

Link dataset here

The salmon data contains two measurements of the growth rings on the scale of Alaskan and Canadian salmon

Assignment (Classifying The Gender of Snake)

- Lakukan klasifikasi pada data berikut menggunakan EXCEL.
- Buat plot visualisasi (scatter plot) terlebih dahulu
- Gunakan hanya variabel **tail length** dan **snout to vent length**
- Lakukan perhitungan detail untuk mendapatkan fungsi diskriminan, critical cutting score, confusion matrix, dan APER

Table 11.10 Concho Water Snake Data

	Gender	Age	TailLength	Snto VnLength		Gender	Age	TailLength	Snto VnLength
1	Female	2	127	441	1	Male	2	126	457
2	Female	2	171	455	2	Male	2	128	466
3	Female	2	171	462	3	Male	2	151	466
4	Female	2	164	446	4	Male	2	115	361
5	Female	2	165	463	5	Male	2	138	473
6	Female	2	127	393	6	Male	2	145	477
7	Female	2	162	451	7	Male	3	145	507
8	Female	2	133	376	8	Male	3	145	493
9	Female	2	173	475	9	Male	3	158	558
10	Female	2	145	398	10	Male	3	152	495
11	Female	2	154	435	11	Male	3	159	521
12	Female	3	165	491	12	Male	3	138	487

Link dataset here

Classification with Several Populations

Minimum ECM Rule

KEEP IN MIND.
Minimum ECM have 3 components: prior probabilities, misclassification costs, and density functions.

Allocate $\mathbf{x}_0$ to π_k if

$$p_k f_k(\mathbf{x}) > p_i f_i(\mathbf{x}) \quad \text{for all } i \neq k$$

Allocate xo to the one that maximizes the posterior probability

$$P(\pi_k \mid \mathbf{x}) = \frac{p_k f_k(\mathbf{x})}{\sum_{i=1}^{g} p_i f_i(\mathbf{x})} = \frac{(\text{prior}) \times (\text{likelihood})}{\sum [(\text{prior}) \times (\text{likelihood})]}$$

$$\text{for } k = 1, 2, \ldots, g$$

Example

		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 $\mathbf{x}_0$:		$f_1(\mathbf{x}_0) = .01$	$f_2(\mathbf{x}_0) = .85$	$f_3(\mathbf{x}_0) = 2$

Assign this observation to one of 3 population

$$p_1 f_1(\mathbf{x}_0) = (.05)(.01) = .0005$$
$$p_2 f_2(\mathbf{x}_0) = (.60)(.85) = .510$$
$$p_3 f_3(\mathbf{x}_0) = (.35)(2) = .700$$

$$p_3 f_3(\mathbf{x}_0) = .700 \geq p_i f_i(\mathbf{x}_0), i = 1, 2$$

Allocate xo to population 3

$$P(\pi_1 | \mathbf{x}_0) = \frac{p_1 f_1(\mathbf{x}_0)}{\sum_{i=1}^{3} p_i f_i(\mathbf{x}_0)} = \frac{(.05)(.01)}{(.05)(.01) + (.60)(.85) + (.35)(2)} = \frac{.0005}{1.2105} = .0004$$

$$P(\pi_2 | \mathbf{x}_0) = \frac{p_2 f_2(\mathbf{x}_0)}{\sum_{i=1}^{3} p_i f_i(\mathbf{x}_0)} = \frac{(.60)(.85)}{1.2105} = \frac{.510}{1.2105} = .421$$

$$P(\pi_3 | \mathbf{x}_0) = \frac{p_3 f_3(\mathbf{x}_0)}{\sum_{i=1}^{3} p_i f_i(\mathbf{x}_0)} = \frac{(.35)(2)}{1.2105} = \frac{.700}{1.2105} = .578$$

Allocate xo to population 3

Fisher Approach

Calculating fisher discriminant

1. Define the *sample between groups* matrix (B)

$$\mathbf{B} = \sum_{i=1}^{g} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'$$

2. Define the *sample within groups* matrix (B)

$$\mathbf{W} = \sum_{i=1}^{g} (n_i - 1)\mathbf{S}_i = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'$$

3. Find eigenvector of $\boldsymbol{W}^{-1}\boldsymbol{B}$

Dasar untuk LDA ekstraksi fitur

Classification rule

Allocate $\mathbf{x}$ to π_k if

$$\sum_{j=1}^{r} (\hat{y}_j - \bar{y}_{kj})^2 = \sum_{j=1}^{r} [\hat{\mathbf{a}}_j'(\mathbf{x} - \bar{\mathbf{x}}_k)]^2 \le \sum_{j=1}^{r} [\hat{\mathbf{a}}_j'(\mathbf{x} - \bar{\mathbf{x}}_i)]^2 \quad \text{for all } i \ne k \tag{11-67}$$

where $\hat{\mathbf{a}}_j$ is defined in (11-62), $\bar{y}_{kj} = \hat{\mathbf{a}}_j'\bar{\mathbf{x}}_k$ and $r \le s$.

Example

Consider the observations on $p = 2$ variables from $g = 3$ populations

$$\pi_1\ (n_1 = 3) \qquad \pi_2\ (n_2 = 3) \qquad \pi_3\ (n_3 = 3)$$

$$\mathbf{X}_1 = \begin{bmatrix} -2 & 5 \\ 0 & 3 \\ -1 & 1 \end{bmatrix}; \qquad \mathbf{X}_2 = \begin{bmatrix} 0 & 6 \\ 2 & 4 \\ 1 & 2 \end{bmatrix}; \qquad \mathbf{X}_3 = \begin{bmatrix} 1 & -2 \\ 0 & 0 \\ -1 & -4 \end{bmatrix}$$

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}; \qquad \bar{\mathbf{x}}_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}; \qquad \bar{\mathbf{x}}_3 = \begin{bmatrix} 0 \\ -2 \end{bmatrix} \qquad \bar{\mathbf{x}} = \begin{bmatrix} 0 \\ \frac{5}{3} \end{bmatrix}$$

Assuming that the populations have a common covariance matrix let us obtain the Fisher discriminants

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}; \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}; \quad \bar{\mathbf{x}}_3 = \begin{bmatrix} 0 \\ -2 \end{bmatrix} \quad \bar{\mathbf{x}} = \begin{bmatrix} 0 \\ \frac{5}{3} \end{bmatrix}$$

$$\mathbf{B} = \sum_{i=1}^{3} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' = \begin{bmatrix} 2 & 1 \\ 1 & 62/3 \end{bmatrix}$$

$$\mathbf{W} = \sum_{i=1}^{3} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' = (n_1 + n_2 + n_3 - 3)\mathbf{S}_{\text{pooled}}$$

$$= \begin{bmatrix} 6 & -2 \\ -2 & 24 \end{bmatrix}$$

$$\mathbf{W}^{-1} = \frac{1}{140}\begin{bmatrix} 24 & 2 \\ 2 & 6 \end{bmatrix}; \qquad \mathbf{W}^{-1}\mathbf{B} = \begin{bmatrix} .3571 & .4667 \\ .0714 & .9000 \end{bmatrix}$$

Find eigenvector of $W^{-1}B$

$$\hat{\mathbf{a}}_1' = [.386 \quad .495]$$

$$\hat{\mathbf{a}}_2' = [.938 \quad -.112]$$

The two discriminants are

$$\hat{y}_1 = \hat{\mathbf{a}}_1'\mathbf{x} = [.386 \quad .495]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = .386x_1 + .495x_2$$

$$\hat{y}_2 = \hat{\mathbf{a}}_2'\mathbf{x} = [.938 \quad -.112]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = .938x_1 - .112x_2$$

$$\hat{y}_1 = \hat{\mathbf{a}}_1'\mathbf{x} = .386x_1 + .495x_2$$
$$\hat{y}_2 = \hat{\mathbf{a}}_2'\mathbf{x} = .938x_1 - .112x_2$$

classify the new observation $\mathbf{x}_0' = [1 \quad 3]$

$$\hat{y}_1 = .386x_{01} + .495x_{02} = .386(1) + .495(3) = 1.87$$
$$\hat{y}_2 = .938x_{01} - .112x_{02} = .938(1) - .112(3) = .60$$

$$\bar{y}_{kj} = \hat{\mathbf{a}}_j'\bar{\mathbf{x}}_k$$

$$\bar{y}_{11} = \hat{\mathbf{a}}_1'\bar{\mathbf{x}}_1 = [.386 \quad .495]\begin{bmatrix} -1 \\ 3 \end{bmatrix} = 1.10$$
$$\bar{y}_{12} = \hat{\mathbf{a}}_2'\bar{\mathbf{x}}_1 = [.938 \quad -.112]\begin{bmatrix} -1 \\ 3 \end{bmatrix} = -1.27$$

$$\bar{y}_{21} = \hat{\mathbf{a}}_1'\bar{\mathbf{x}}_2 = 2.37$$
$$\bar{y}_{22} = \hat{\mathbf{a}}_2'\bar{\mathbf{x}}_2 = .49$$
$$\bar{y}_{31} = \hat{\mathbf{a}}_1'\bar{\mathbf{x}}_3 = -.99$$
$$\bar{y}_{32} = \hat{\mathbf{a}}_2'\bar{\mathbf{x}}_3 = .22$$

$$\sum_{j=1}^{2}(\hat{y}_j - \bar{y}_{1j})^2 = (1.87 - 1.10)^2 + (.60 + 1.27)^2 = 4.09$$
$$\sum_{j=1}^{2}(\hat{y}_j - \bar{y}_{2j})^2 = (1.87 - 2.37)^2 + (.60 - .49)^2 \boxed{= .26}$$
$$\sum_{j=1}^{2}(\hat{y}_j - \bar{y}_{3j})^2 = (1.87 + .99)^2 + (.60 - .22)^2 = 8.32$$

We allocate x0 to Population 2, the smallest value

Semangat guys ☺