Analisis Deskriminan

Ide Dasar

1. Sudah ada pengelompokan objek
2. Mencari fungsi yang bisa dijadikan dasar membedakan (mendiskriminankan) objek ke dalam kelompok-kelompok
3. Menentukan ke kelompok mana suatu objek baru
4. Peubah pembeda adalah Peubah yang ragamnya besar
5. Pembedaan seringkali memerlukan kombinasi beberapa Peubah (satu peubah tidak cukup)

Memperoleh fungsi diskriminan, yaitu fungsi yang mampu digunakan membedakan suatu objek masuk ke dalam populasi tertentu berdasarkan pengamatan terhadap objek tersebut

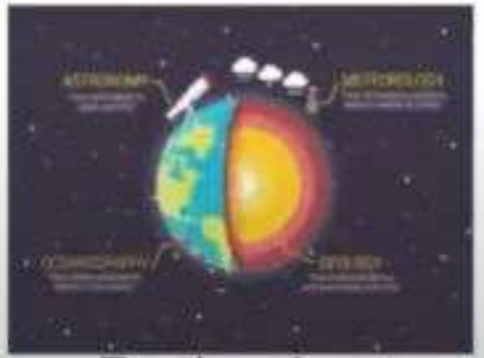
Penerapan

Bankruptcy prediction

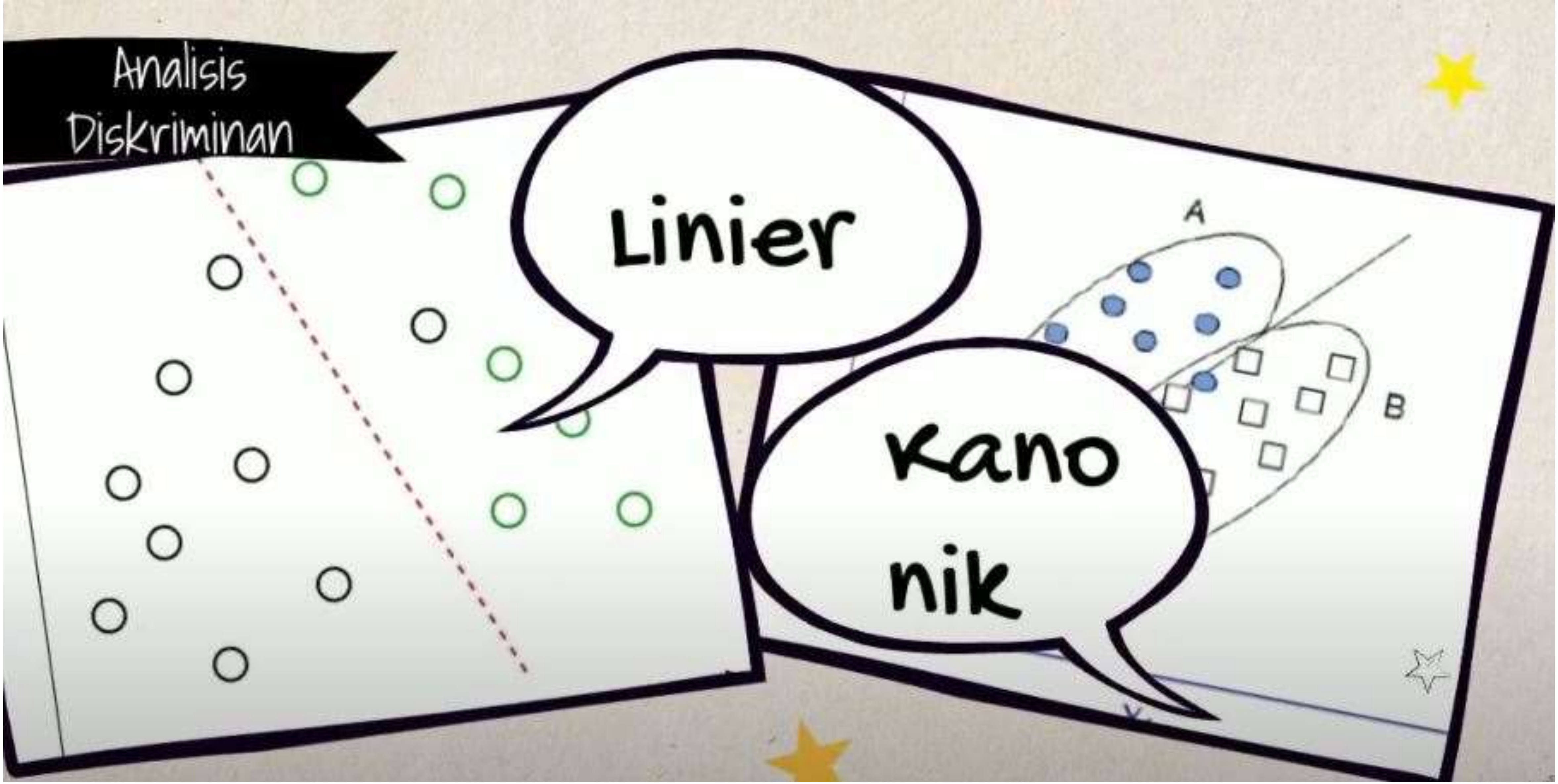
Face recognition

Marketing

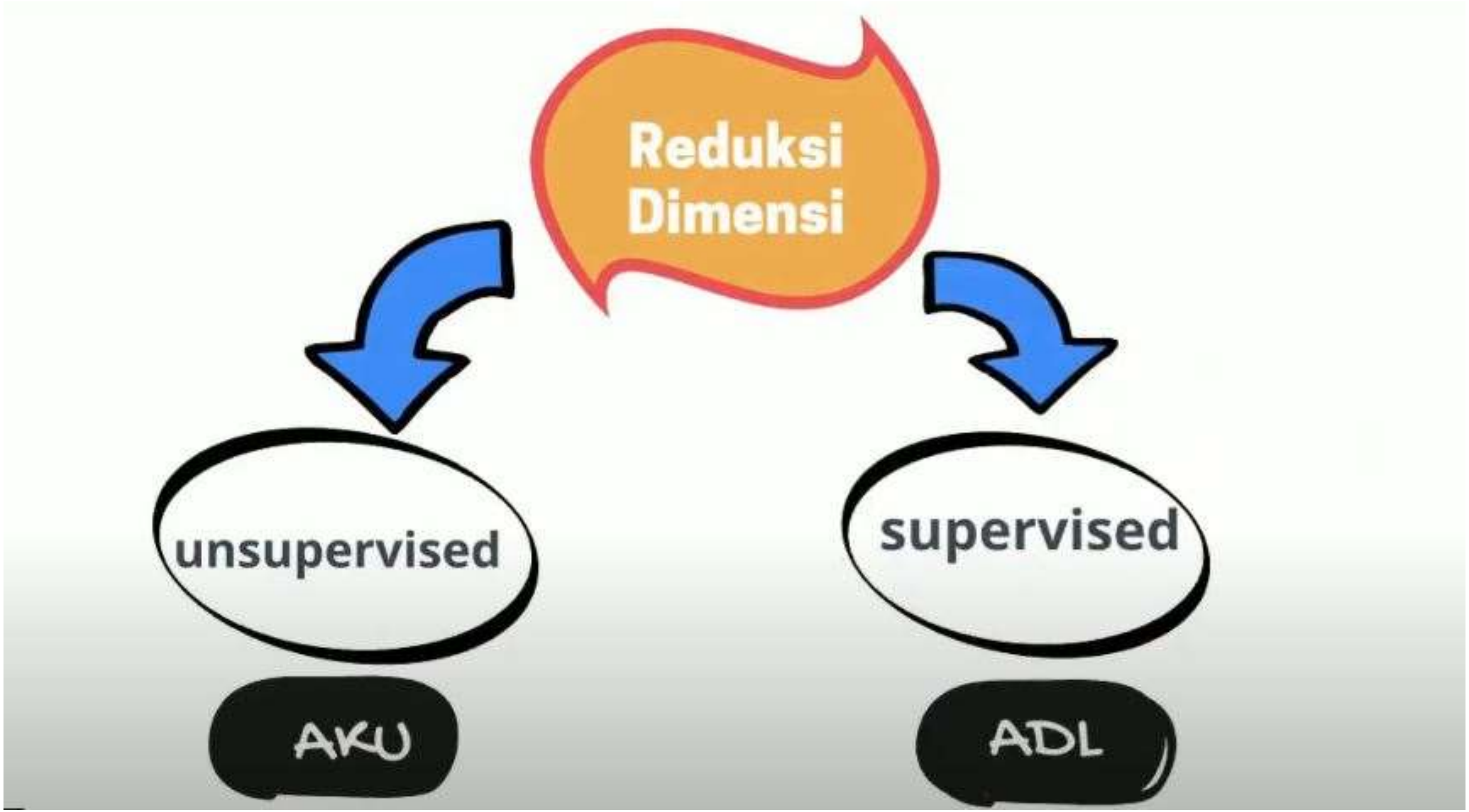
Biomedical studies



Earth science



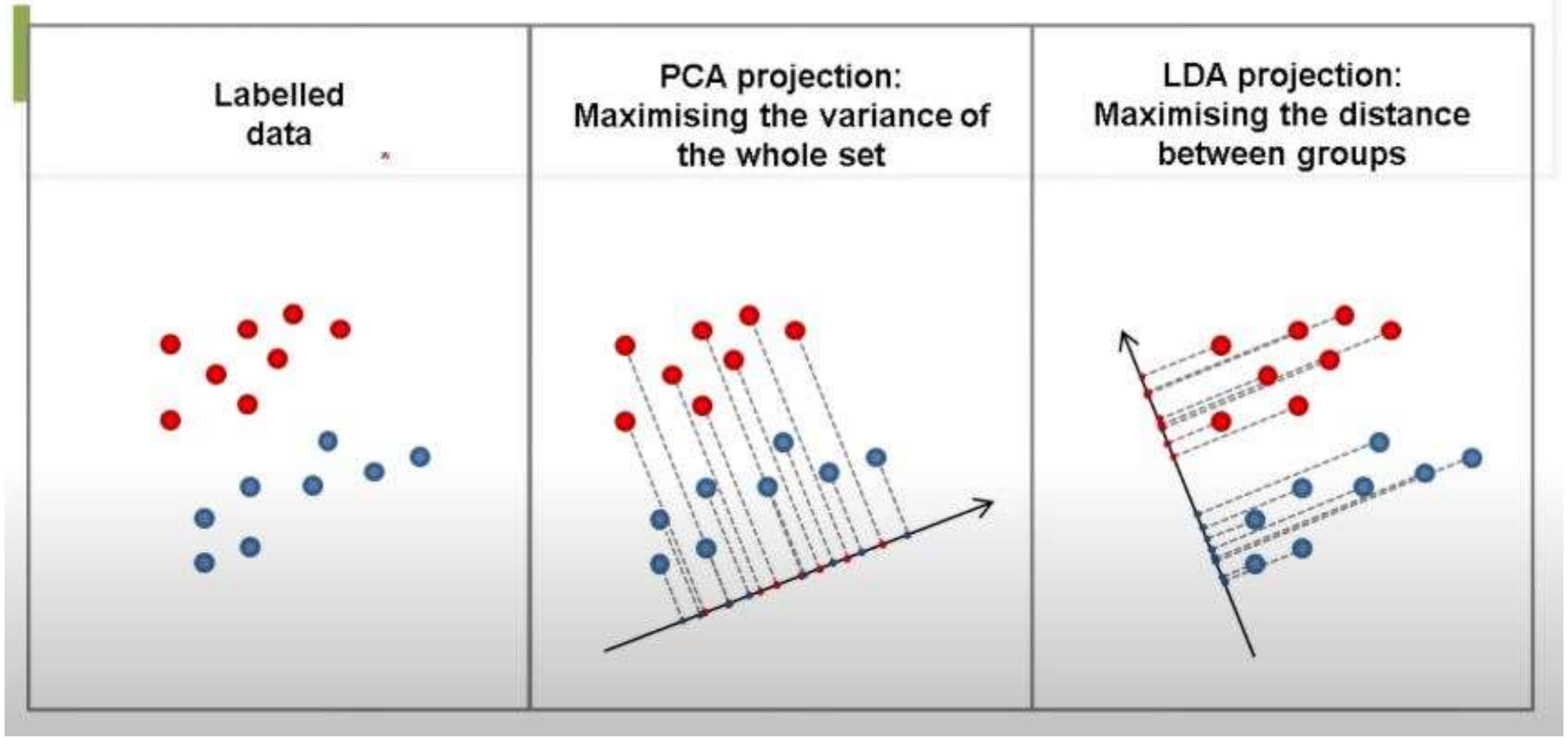
Analisis
Diskriminan
Linier
Kano
nik
A
B


Reduksi Dimensi
unsupervised
supervised
AKU
ADL

berfokus pada memaksimalkan keterpisahan di antara kategori yang diketahui dengan membuat sumbu linier baru dan memproyeksikan titik data pada sumbu tersebut.

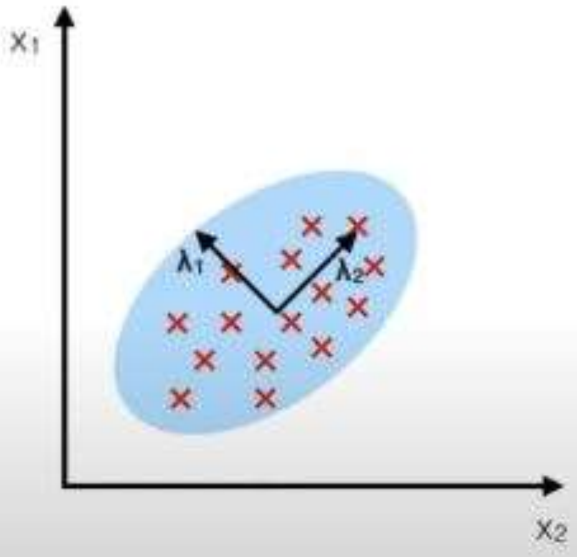
ANALISIS DISKRIMINAN

LINIER


Labelled data
PCA projection: Maximising the variance of the whole set
LDA projection: Maximising the distance between groups

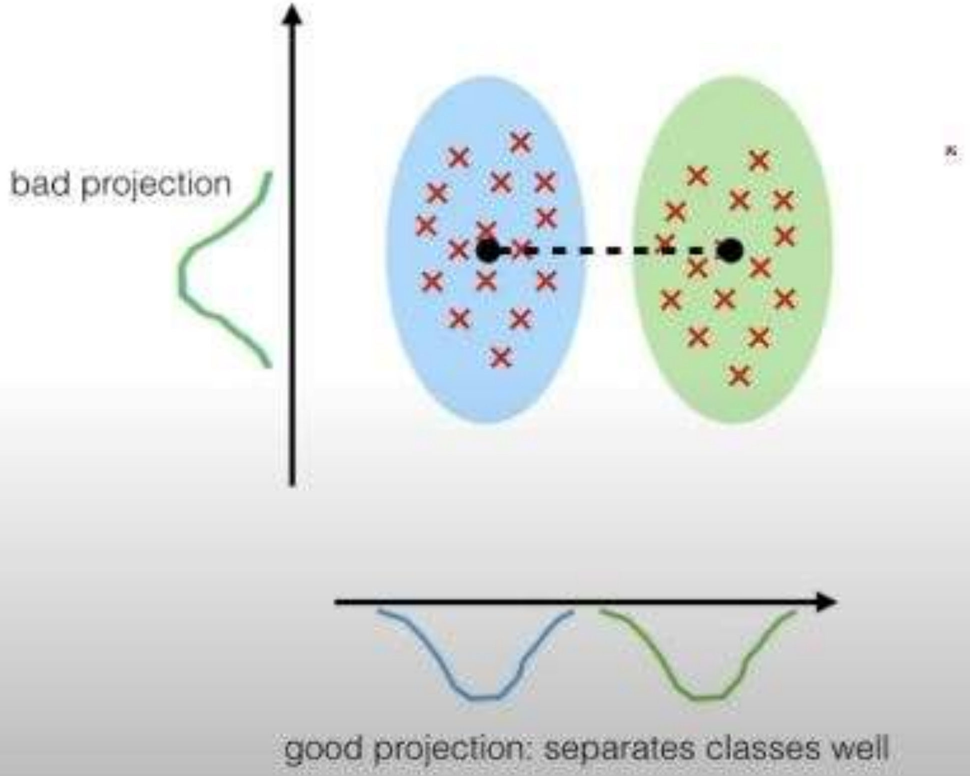
PCA:

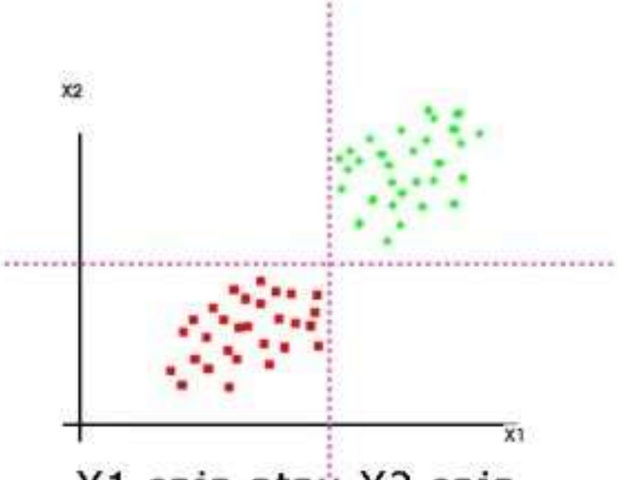
component axes that
maximize the variance



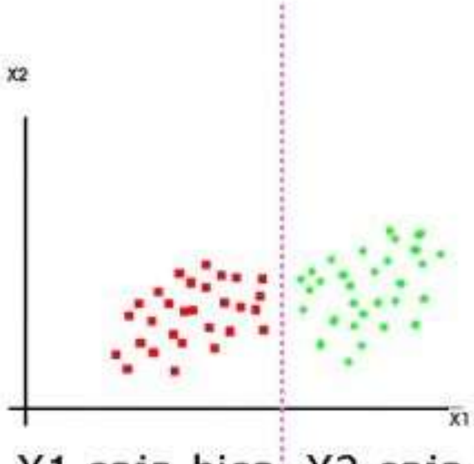
LDA:

maximizing the component
axes for class-separation

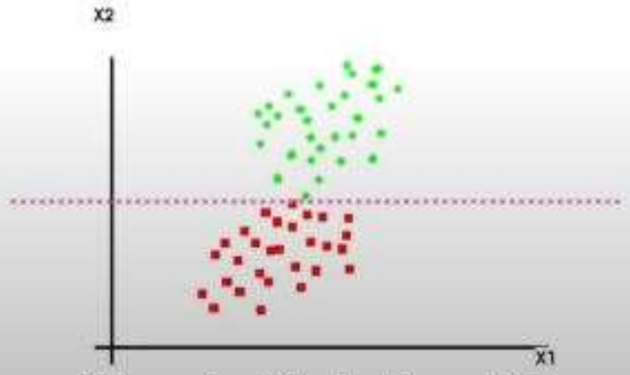




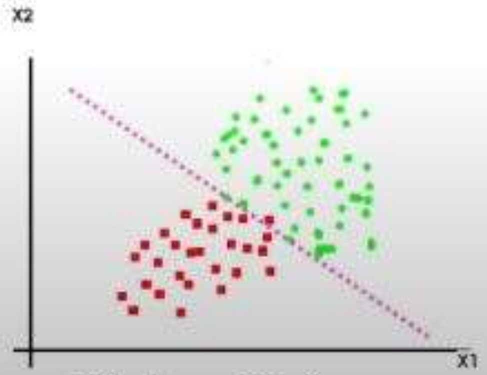
X1 saja atau X2 saja bisa membedakan



X1 saja bisa, X2 saja tidak bisa



X1 saja tidak bisa, X2 saja bisa



X1 dan X2 harus dikombinasikan

Fungsi Diskriminan

- Bisa memisahkan kelompok-kelompok dengan salah klasifikasi paling kecil
- Mirip fungsi model regresi
- Bisa juga menggunakan analisis regresi logistik

$$D = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_kX_k$$

Dimana:

D = skor diskriminan

b_i = koefisien diskriminan atau bobot

X_i = predictor atau variable independen

Aturan Fungsi Diskriminan Fisher

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

Allocate $\mathbf{x}_0$ to π_1 if

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

or (11-35)

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

Allocate $\mathbf{x}_0$ to π_2 if

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

or

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

Ilustrasi

Exercises 11.19

a) Using the original data sets $\mathbf{X}_1$ and $\mathbf{X}_2$ given in Example 11.6, calculate $\bar{\mathbf{x}}_i$, $\mathbf{S}_i$, $i = 1,2$, and $\mathbf{S}_{\text{pooled}}$, verifying the results provided for these quantities in the example.

Example 11.6.

Consider the following data matrices. We shall assume that the $n_1 = n_2 = 3$ bivariate observations were selected randomly from two populations π_1 and π_2 with a common covariate matrix.

$$\mathbf{X}_1 = \begin{bmatrix} 2 & 12 \\ 4 & 10 \\ 3 & 8 \end{bmatrix} \quad \text{dan} \quad \mathbf{X}_2 = \begin{bmatrix} 5 & 7 \\ 3 & 9 \\ 4 & 5 \end{bmatrix}$$

$$\mathbf{X}_1 = \begin{bmatrix} 2 & 12 \\ 4 & 10 \\ 3 & 8 \end{bmatrix}; \quad \bar{\mathbf{x}}_1 = \begin{bmatrix} 3 \\ 10 \end{bmatrix}, \quad \mathbf{S}_1 = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

$$\mathbf{X}_2 = \begin{bmatrix} 5 & 7 \\ 3 & 9 \\ 4 & 5 \end{bmatrix}; \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 7 \end{bmatrix},$$

$$\mathbf{S}_2 = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

Menghitung $\mathbf{S}_{\text{pooled}}$ dapat menggunakan persamaan berikut:

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

sehingga matriks kovarian gabungannya adalah

$$\begin{aligned} &\mathbf{S}_{\text{pooled}} \\ &= \left[\frac{3 - 1}{(3 - 1) + (3 - 1)}\right]\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} \\ &+ \left[\frac{3 - 1}{(3 - 1) + (3 - 1)}\right]\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} \end{aligned}$$

$$\mathbf{S}_{\text{pooled}} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

b) Using calculation in Part a,compute Fisher's linear discriminant function, and use it to classify the sample observation according to Rule (11.25).

Diketahui nilai $\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2$ dan $\mathbf{S}_{\text{pooled}}$ dibagian (a). Kemudian hitung $\mathbf{S}_{pooled}^{-1}$ dimana

$$\mathbf{S}_{pooled}^{-1} = \begin{bmatrix} 1.333 & 0.333 \\ 0.333 & 0.333 \end{bmatrix}$$

Fungsi diskriminan prior yang sama adalah

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

$$= [-1 \quad 3]\left[\begin{bmatrix} 1.333 & 0.333 \\ 0.333 & 0.333 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right]$$

$$= -0.333x_1 + 0.667x_2$$

dengan

$$\bar{y}_1 = \hat{\mathbf{a}}'\bar{\mathbf{x}}_1 = [-0.333 \quad 0.667]\begin{bmatrix} 3 \\ 10 \end{bmatrix} = 5.667$$

$$\bar{y}_2 = \hat{\mathbf{a}}'\bar{\mathbf{x}}_2 = [-0.333 \quad 0.667]\begin{bmatrix} 4 \\ 7 \end{bmatrix} = 3.333$$

dan titik tengah rata-ratanya

$$\bar{m} = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) = \frac{1}{2}(5.667 + 3.333) = 4.5$$

- Alokasikan x_0 ke π_1 jika $-0.333x_1 + 0.667x_2 - 4.5 \geq 0$ dan
- alokasikan x_0 ke π_2 jika $-0.333x_1 + 0.667x_2 - 4.5 < 0$ sehingga diperoleh

	π_1		π_2	
	$\hat{\mathbf{a}}'\mathbf{x} - \bar{m}$	klasifikasi	$\hat{\mathbf{a}}'\mathbf{x} - \bar{m}$	klasifikasi
=(-0.333*2)+(0.667*12)-4.5	2.838	π_1	-1.496	π_2
=(-0.333*4)+(0.667*10)-4.5	0.838	π_1	0.505	π_1
=(-0.333*3)+(0.667*8)-4.5	-0.163	π_2	-2.497	π_2

dan titik tengah rata-ratanya

$$\bar{m} = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) = \frac{1}{2}(5.667 + 3.333) = 4.5$$

- Alokasikan x_0 ke π_1 jika $-0.333x_1 + 0.667x_2 - 4.5 \geq 0$ dan
- alokasikan x_0 ke π_2 jika $-0.333x_1 + 0.667x_2 - 4.5 < 0$ sehingga diperoleh

	π_1		π_2		
	$\hat{\mathbf{a}}'\mathbf{x} - \bar{m}$	klasifikasi	$\hat{\mathbf{a}}'\mathbf{x} - \bar{m}$	klasifikasi	
=(-0.333*2)+(0.667*12)-4.5	2.838	π_1	-1.496	π_2	=(-0.333*5)+(0.667*7)-4.5
=(-0.333*4)+(0.667*10)-4.5	0.838	π_1	0.505	π_1	=(-0.333*3)+(0.667*9)-4.5
=(-0.333*3)+(0.667*8)-4.5	-0.163	π_2	-2.497	π_2	=(-0.333*4)+(0.667*5)-4.5

ATURAN *smallest squared distance* $D_i^2(x)$

- Hitung $\mathrm{D}_{\mathrm{i}}^2(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}}_{\mathbf{i}})'\mathbf{S}_{\mathrm{pooled}}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_{\mathbf{i}}), \mathrm{i} = 1{,}2$ untuk semua kelompok
- Nilai Di yang terkecil merupakan lokasi dari kelompoknya

c) Classify the sample observation on the basis of smallest squared distance $D_i^2(x)$ of the observation from the group means $\bar{x}_1$ and $\bar{x}_2$. [See (11.54).] Compare the results with those in Part b.

$$D_i^2(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}}_i)' \mathbf{S}_{\text{pooled}}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i), i = 1,2$$

Dari persamaan diatas untuk $\bar{\mathbf{x}}_1$ diperoleh hasil

$$\begin{aligned} & \left(\begin{matrix}2\\12\end{matrix} - \begin{matrix}3\\10\end{matrix}\right)' \begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix} \left(\begin{matrix}2\\12\end{matrix} - \begin{matrix}3\\10\end{matrix}\right) = 1.333 \\ D_1^2(\mathbf{x}) = & \left(\begin{matrix}4\\10\end{matrix} - \begin{matrix}3\\10\end{matrix}\right)' \begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix} \left(\begin{matrix}4\\10\end{matrix} - \begin{matrix}3\\10\end{matrix}\right) = 1.333 \\ & \left(\begin{matrix}3\\8\end{matrix} - \begin{matrix}3\\10\end{matrix}\right)' \begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix} \left(\begin{matrix}3\\8\end{matrix} - \begin{matrix}3\\10\end{matrix}\right) = 1.332 \end{aligned}$$

$$\begin{aligned} & \left(\begin{matrix}5\\7\end{matrix} - \begin{matrix}3\\10\end{matrix}\right)' \begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix} \left(\begin{matrix}5\\7\end{matrix} - \begin{matrix}3\\10\end{matrix}\right) = 4.333 \\ D_2^2(\mathbf{x}) = & \left(\begin{matrix}3\\9\end{matrix} - \begin{matrix}3\\10\end{matrix}\right)' \begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix} \left(\begin{matrix}3\\9\end{matrix} - \begin{matrix}3\\10\end{matrix}\right) = 0.333 \\ & \left(\begin{matrix}4\\5\end{matrix} - \begin{matrix}3\\10\end{matrix}\right)' \begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix} \left(\begin{matrix}5\\7\end{matrix} - \begin{matrix}3\\10\end{matrix}\right) = 6.328 \end{aligned}$$

untuk $\bar{\mathbf{x}}_2$ diperoleh hasil

$$\begin{pmatrix}2 & - & 4\\12 & & 7\end{pmatrix}'\begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix}\begin{pmatrix}2 & - & 4\\12 & & 7\end{pmatrix} = 6{,}997$$

- $D_1^2(\mathbf{x}) = \begin{pmatrix}4 & - & 4\\10 & & 7\end{pmatrix}'\begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix}\begin{pmatrix}4 & - & 4\\10 & & 7\end{pmatrix} = 2.997$

$$\begin{pmatrix}3 & - & 4\\8 & & 7\end{pmatrix}'\begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix}\begin{pmatrix}3 & - & 4\\8 & & 7\end{pmatrix} = 1$$

$$\begin{pmatrix}5 & - & 4\\7 & & 7\end{pmatrix}'\begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix}\begin{pmatrix}5 & - & 4\\7 & & 7\end{pmatrix} = 1.333$$

- $D_2^2(\mathbf{x}) = \begin{pmatrix}3 & - & 4\\9 & & 7\end{pmatrix}'\begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix}\begin{pmatrix}3 & - & 4\\9 & & 7\end{pmatrix} = 1.333$

$$\begin{pmatrix}4 & - & 4\\5 & & 7\end{pmatrix}'\begin{bmatrix}1.333 & 0.333\\0.333 & 0.333\end{bmatrix}\begin{pmatrix}4 & - & 4\\5 & & 7\end{pmatrix} = 0.333$$

Hasil tersebut dapat dikelompokan dengan membuat tabel berikut:

π_1			π_2		
$D_1^2(\mathbf{x})$	$D_2^2(\mathbf{x})$	klasifikasi	$D_1^2(\mathbf{x})$	$D_2^2(\mathbf{x})$	klasifikasi
1.333	6.997	π_1	4.333	1.333	π_2
1.333	2.997	π_1	0.333	1.333	π_2
1.332	1	π_2	0.628	0.333	π_1

π_1		π_2	
$\hat{\mathbf{a}}'\mathbf{x} - \bar{m}$	klasifikasi	$\hat{\mathbf{a}}'\mathbf{x} - \bar{m}$	klasifikasi
2.838	π_1	-1.496	π_2
0.838	π_1	0.505	π_1
-0.163	π_2	-2.497	π_2

Fisher

π_1			π_2		
$D_1^2(\mathbf{x})$	$D_2^2(\mathbf{x})$	klasifikasi	$D_1^2(\mathbf{x})$	$D_2^2(\mathbf{x})$	klasifikasi
1.333	6.997	π_1	4.333	1.333	π_2
1.333	2.997	π_1	0.333	1.333	π_1
1.332	1	π_2	0.628	0.333	π_2

Smallest squared distance