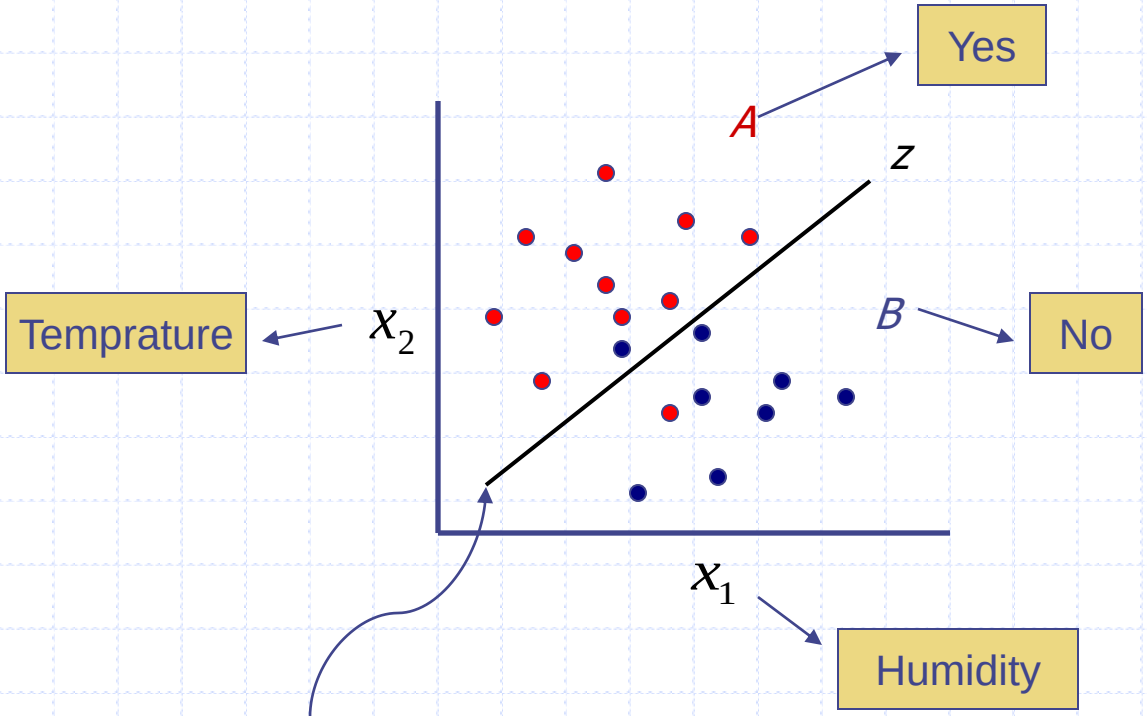
LINEAR DISCRIMINANT ANALYSIS (LDA)

Dr. Saed Sayad

Linear Discriminant Analysis



$$Z = b_1 x_1 + b_2 x_2$$

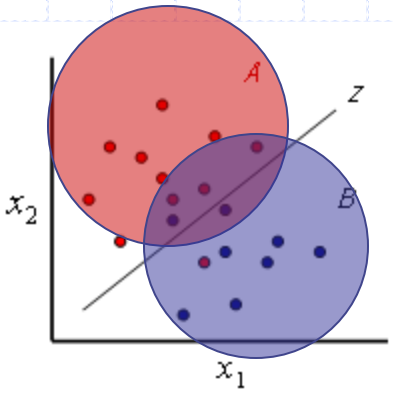
Linear Discriminant Analysis

$$\Delta^2 = \frac{(\bar{z}_A - \bar{z}_B)^2}{\text{Varianceof Z within groups}}$$

maximize

$$\bar{z}_A = b_1\bar{x}_{A1} + b_2\bar{x}_{A2}$$

$$\bar{z}_B = b_1\bar{x}_{B1} + b_2\bar{x}_{B2}$$



$$Z = b_1x_1 + b_2x_2$$

Linear Discriminant Analysis

$$b_1 = c_{11}(\bar{x}_{A1} - \bar{x}_{B1}) + c_{12}(\bar{x}_{A2} - \bar{x}_{B2})$$

$$b_2 = c_{21}(\bar{x}_{A1} - \bar{x}_{B1}) + c_{22}(\bar{x}_{A2} - \bar{x}_{B2})$$

$$\begin{bmatrix} S_{x1x1} & S_{x1x2} \\ S_{x2x1} & S_{x2x2} \end{bmatrix}^{-1} \Rightarrow \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

Linear Discriminant Analysis

$$z_0 = b_1\left(\frac{\bar{x}_{A1} + \bar{x}_{B1}}{2}\right) + b_2\left(\frac{\bar{x}_{A2} + \bar{x}_{B2}}{2}\right)$$

If $\bar{z}_A > \bar{z}_B$ and $z > z_0$ Then *A*

If $\bar{z}_A < \bar{z}_B$ and $z > z_0$ Then *B*

LINEAR DISCRIMINANT ANALYSIS

TEMP	HUMIDITY	PLAY
85	85	no
80	90	no
83	78	yes
70	96	yes
68	80	yes
65	70	no
64	65	yes
72	95	no
69	70	yes
75	80	yes
75	70	yes
72	90	yes
81	75	yes
71	80	no

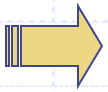
$$S_{xy} = \sum (x - \bar{x})(y - \bar{y})$$

	S_{x1x1}	S_{x1x2}	S_{x2x2}
No	249.2	163	370
Yes	304	37	781.6
Pooled	553.2	200	1151.6
Var or Covar	39.5	14.3	82.6

$$\begin{bmatrix} S_{x1x1} & S_{x1x2} \\ S_{x2x1} & S_{x2x2} \end{bmatrix}^{-1} \Rightarrow \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

39.5	14.3
14.3	82.6

Covariance Matrix



0.0270	-0.0047
-0.0047	0.0129

Inverse Matrix

LINEAR DISCRIMINANT ANALYSIS

	Mean X1	Mean X2
No	74.6	84
Yes	73	78.22
Difference	1.6	5.78
Mean=0.5(No+Yes)*	73.8	81.11

$$b_1 = c_{11}(\bar{x}_{A1} - \bar{x}_{B1}) + c_{12}(\bar{x}_{A2} - \bar{x}_{B2})$$

b1 = (0.0270)(1.6)+(-0.0047)(5.78) = 0.016

$$b_2 = c_{21}(\bar{x}_{A1} - \bar{x}_{B1}) + c_{22}(\bar{x}_{A2} - \bar{x}_{B2})$$

b2 = (-0.0047)(1.6)+(0.0129)(5.78) = 0.067

Linear Discriminant Analysis

$$Z = b_1 x_1 + b_2 x_2$$

$$Z = 0.016 x_1 + 0.067 x_2$$

$$z_0 = b_1\left(\frac{\bar{x}_{A1} + \bar{x}_{B1}}{2}\right) + b_2\left(\frac{\bar{x}_{A2} + \bar{x}_{B2}}{2}\right)$$

Z_0 = (0.016)(73.8)+(0.067)(81.11) = 6.615

If $\bar{z}_{yes} > \bar{z}_{no}$ and $z > z_0$ Then *yes*

If $\bar{z}_{yes} < \bar{z}_{no}$ and $z > z_0$ Then *no*

LINEAR DISCRIMINANT ANALYSIS

TEMP	HUMIDITY	PLAY
85	85	no
80	90	no
83	78	yes
70	96	yes
68	80	yes
65	70	no
64	65	yes
72	95	no
69	70	yes
75	80	yes
75	70	yes
72	90	yes
81	75	yes
71	80	no

$z = (0.016)(85)+(0.067)(85) = 7.055$

$z_0 = 6.615$

Correct

If $\bar{z}_{yes} < \bar{z}_{no}$ and $z > z_0$ Then *no*

(0.016 x 74.6 + 0.067 x 84) = 6.82

(0.016 x 73 + 0.067 x 78.22) = 6.41

Linear Discriminant Analysis

Mahalanobis Distance

$$\Delta^2 = b_1(\mu_1^{No} - \mu_1^{Yes}) + b_2(\mu_2^{No} - \mu_2^{Yes})$$

$$\Delta^2 = 0.016 \times 1.6 + 0.067 \times 5.78 = 0.4129$$

$$\Delta = 0.64$$

Two normal distributions (populations) with equal variance, separated by 0.64 standard deviation.