(4) Eigen Vector & Eigen Value :-

- The covariance matrix defines both the variance (spread) and dir^n.
- The largest eigen ~~value~~ vector of covariance matrix points to the largest spread of the data.

Q:- Use PCA Table to reduce ~~a~~ dim^n from 2 to 1.

x_1	x_2	target	New ~~value~~ feature
4	11		-4.30535
8	4		3.7361
13	5		5.6920
7	14		-5.1238

→ see sol^n for this value.

$\therefore \bar{x}_1 = 8 \quad \bar{x}_2 = 8.5$

~~Q~~ ~~@~~ Covariance matrix :-

	x_1	x_2
x_1		
x_2		

$$cov(x_j, x_j) = \frac{\sum_{j=1}^{m} (x_j - \bar{x}_1)^2}{m-1}$$

$$cov(x_1, x_1) = \frac{(4-8)^2 + (8-8)^2 + (13-8)^2 + (7-8)^2}{3}$$

$$= 14$$

~~cov(x1, x2) = (4-8)(11-8.5)+~~

$$cov(x_1, x_2) = \frac{\sum_{i=1}^{m} (x_1 - \bar{x}_1)(x_2 - \bar{x}_2)}{m-1}$$

$$= \frac{(4-8)(11-8.5) + (8-8)(4-8.5) + (13-8)(5-8.5) + (7-8)(14-8.5)}{3}$$

$$= -11$$

$$cov(x_2, x_2) = \frac{(11-8.5)^2 + (4-8.5)^2 + (5-8.5)^2 + (14-8.5)^2}{3}$$

$$= 23$$

Covariance Matrix :-

$$cov(x_1, x_2) = \begin{array}{c|cc} & x_1 & x_2 \\ \hline x_1 & 14 & -11 \\ x_2 & -11 & 23 \end{array}$$

weighted Impact of $\lambda_1 = \dfrac{\lambda_1}{\lambda_1+\lambda_2}$

" " $\lambda_2 = \dfrac{\lambda_2}{\lambda_1+\lambda_2}$

Dt.:
Pg.:

2) $\det(S - \lambda I) = 0$

$$\left| \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

$$\left| \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} \right| = 0$$

$$\begin{vmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{vmatrix} = 0$$

$$(14-\lambda)(23-\lambda) - (-11)(-11) = 0$$

$$(14-\lambda)(23-\lambda) - 11^2 = 0$$

$$\boxed{\begin{aligned} \lambda_1 &= 30.38 \\ \lambda_2 &= 6.61 \end{aligned}}$$ → take larger value.

$\Rightarrow (S - \lambda_1 I)X = 0$ → eigen vector

$$\begin{bmatrix} 14-\lambda_1 & -11 \\ -11 & 23-\lambda_1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$e_1 = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}.$$

$$U = \begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.3849 \end{bmatrix}$$

$\Rightarrow \|U\| = \sqrt{11^2 + (14-\lambda)^2}$

$= 19.7348$

$$e_1 = \frac{U_1}{\|U\|} = \begin{bmatrix} 11/19.73 \\ -16.3045/19.7348 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

$\boxed{PC_1 = v^T \cdot x}$

$\Rightarrow \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} x_{11} - \bar{x}_1 \\ x_{21} - \bar{x}_2 \end{bmatrix}$

$= \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix}$

$= \begin{bmatrix} -4.30535 \end{bmatrix}.$

Dt.: 13/03/23
Pg.:

⇒ Linear Discriminant Analysis (LDA) :-

- LDA does not change the location but tries to maximize the separativity of classes.
-

⇒ Fisher Dicriminant Ratio :-

$$J(m) = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$ → maximize this (numerator) → minimize this (denominator)

↳ Fisher ratio.

S_W → scatter matrix within class (Inter-class scatter matrix).

S_B → class scatter matrix

$$\boxed{S_W = \sum_{i=1}^{c} S_i}$$ (c = no. of classes)

$$\boxed{J(m) = \frac{W^T S_B W}{W^T S_W W}}$$

$y = w^T x$

M1

Dt.: 14/03/23

$$\frac{dJ(m)}{dW} = 0$$

$$S_W^{-1} S_B W - JW = 0$$

$$S_W^{-1} S_B W = JW$$

X_1	X_2	class
4	1	1
9	10	2
2	4	1
2	3	1
6	8	2
3	6	1
9	5	2
8	7	2
10	8	2
4	4	1

$c_1 = (x_1, x_2)\ \{(4,1), (2,4), (2,3), (3,6), (4,4)\}$

$c_2 = (x_1, x_2)\ \{(9,10), (6,8), (9,5), (8,7), (10,8)\}$

(I) Compute mean:-

$$\mu_1 = \left\{ \frac{4+2+2+3+4}{5}, \frac{1+4+3+6+4}{5} \right\}$$

$$\mu_1 = (3, 3.6)$$

$$\mu_2 = \left(\frac{9+6+9+8+10}{5}, \frac{10+8+5+7+8}{5} \right)$$

$$\mu_2 = (8.4, 7.6)$$

(II) Compute scatter matrix:- (S_w):-

$$S_1 = \sum_{x \in w_1} (x-\mu_1)(x-\mu_1)^T$$

$$= \underbrace{\begin{bmatrix} 1 & -1 & -1 & 0 & 1 \\ -2.6 & 0.4 & -0.6 & 2.4 & 0.4 \end{bmatrix}}_{(x-\mu_1)} \underbrace{\begin{bmatrix} 1 & -2.6 \\ -1 & 0.4 \\ -1 & -0.6 \\ 0 & 2.4 \\ 1 & 0.4 \end{bmatrix}}_{(x-\mu_1)^T}$$

or

$$\begin{bmatrix} 1 \\ -2.6 \end{bmatrix} \begin{bmatrix} 1 & -2.6 \end{bmatrix} + \begin{bmatrix} -1 \\ 0.4 \end{bmatrix} \begin{bmatrix} -1 & 0.4 \end{bmatrix} + \cdots$$

Dt.:
Pg.:

i) $\begin{bmatrix} 1 \\ -2.6 \end{bmatrix} \begin{bmatrix} 1 & -2.6 \end{bmatrix} = \begin{bmatrix} 1 & -2.6 \\ -2.6 & 6.76 \end{bmatrix}$

ii) $\begin{bmatrix} -1 \\ 0.4 \end{bmatrix} \begin{bmatrix} -1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & -0.4 \\ -0.4 & 0.16 \end{bmatrix}$

iii) $\begin{bmatrix} -1 \\ -0.6 \end{bmatrix} \begin{bmatrix} -1 & -0.6 \end{bmatrix} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 0.36 \end{bmatrix}$

iv) $\begin{bmatrix} 0 \\ 2.4 \end{bmatrix} \begin{bmatrix} 0 & 2.4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 5.76 \end{bmatrix}$

v) $\begin{bmatrix} 1 \\ 0.4 \end{bmatrix} \begin{bmatrix} 1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 0.16 \end{bmatrix}$

$S_1 = \frac{1}{5} \begin{bmatrix} 4 & -2 \\ -2 & 13.2 \end{bmatrix}$

$= \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.64 \end{bmatrix}$

for class 2 :-

i) $\begin{bmatrix} 0.6 \\ 2.4 \end{bmatrix} \begin{bmatrix} 0.6 & 2.4 \end{bmatrix} = \begin{bmatrix} 0.36 & 1.44 \\ 1.44 & 5.76 \end{bmatrix}$

ii) $\begin{bmatrix} -2.4 \\ 0.4 \end{bmatrix} \begin{bmatrix} -2.4 & 0.4 \end{bmatrix} = \begin{bmatrix} 5.76 & -0.96 \\ -0.96 & 0.16 \end{bmatrix}$

iii) $\begin{bmatrix} 0.6 \\ -2.6 \end{bmatrix} \begin{bmatrix} 0.6 & -2.6 \end{bmatrix} = \begin{bmatrix} 0.36 & -1.56 \\ -1.56 & 6.76 \end{bmatrix}$

iv) $\begin{bmatrix} -0.4 \\ -0.6 \end{bmatrix} \begin{bmatrix} -0.4 & -0.6 \end{bmatrix} = \begin{bmatrix} 0.16 & 0.24 \\ 0.24 & 0.36 \end{bmatrix}$

v) $\begin{bmatrix} 1.6 \\ 0.4 \end{bmatrix} \begin{bmatrix} 1.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 2.56 & 0.64 \\ 0.64 & 0.16 \end{bmatrix}$

$S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$

$S_W = S_1 + S_2 = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$

MINHAAZ

(II) Compute S_B (b/w class):-

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

$$= \begin{bmatrix} 3-8.4 \\ 3.6-7.6 \end{bmatrix} \begin{bmatrix} 3-8.4 & 3.6-7.6 \end{bmatrix}$$

$$= \begin{bmatrix} -5.4 \\ -4 \end{bmatrix} \begin{bmatrix} -5.4 & -4 \end{bmatrix}$$

$$S_B = \begin{bmatrix} 29.16 & 21.6 \\ 21.6 & 16 \end{bmatrix}.$$

(III) Find best LDA Projection Vector:-

$$S_W^{-1} S_B W = \lambda W$$

$$\Rightarrow |S_W^{-1} S_B - \lambda I| = 0$$

$$\Rightarrow \lambda = 15.65$$

Dt.: 27/03/23
Pg.:

⇒ **loss fn. for logistics regression:-**

$$\boxed{L = -\frac{1}{m}\sum_{j=1}^{m} y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)}$$

↳ cost fn. (for full table).

$$\boxed{\sigma(z) = \frac{1}{1+e^{z}}}$$

$$\boxed{\sigma'(z) = \sigma(z)\left[1-\sigma(z)\right]}$$

$$\boxed{L = -y_i \log \hat{y}_i - (1-y_i)\log(1-\hat{y}_i)}$$

↳ loss fn. (only for single row).

$$z = \sum_{j=1}^{m} w_j x_j$$

$$\hat{y} = \sigma(z)$$

$$y = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1m} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2m} \\ \vdots & & & & \\ X_{m1} & X_{m2} & X_{m3} & \cdots & X_{mm} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}$$

Dt:
Pg.:

$$\hat{y}_1 = \sigma(x_{11}w_1 + x_{12}w_2 + \dots + x_{1m}w_m)$$

$$\hat{y}_2 = \sigma(x_{21}w_1 + x_{22}w_2 + \dots + x_{2m}w_m)$$

$$\hat{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \sigma \underbrace{\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & & & \\ x_{m1} & x_{m2} & \dots & x_{mm} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}}_{W}$$

$$\boxed{\hat{Y} = \sigma(XW)}$$

Taking $\Rightarrow$ $\sum_{i=1}^{m} y_i \log \hat{y}_i = y_1 \log \hat{y}_1 + y_2 \log \hat{y}_2 + \dots + y_m \log \hat{y}_m$

$$= [y_1 \; y_2 \; \dots \; y_m] \begin{bmatrix} \log \hat{y}_1 \\ \log \hat{y}_2 \\ \vdots \\ \log \hat{y}_m \end{bmatrix}$$

$$= [y_1 \; y_2 \dots y_m] \log \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_m \end{bmatrix}$$

$$= Y \log \hat{Y}$$

Taking, $\sum_{i=1}^{m} (1-\hat{y}_i)\log(1-\hat{y}_i)$

$$(1-Y)\log[1-\sigma(XW)]$$

$$\boxed{L = \frac{-1}{m}\left[Y\log\hat{Y} + (1-Y)\log[1-\sigma(XW)]\right]}$$

Differentiating wrt ~~1st term~~ W.

. Taking 1st term:-

$$\frac{\partial L}{\partial W} = \frac{\partial(Y\log\hat{Y})}{\partial W}$$

$$= \frac{Y\,\partial\log\hat{Y}}{\partial W}$$

$$= \frac{Y}{\hat{Y}}\frac{\partial(\sigma(XW))}{\partial W}$$

$$= \frac{Y}{\hat{Y}}\sigma(XW)[1-\sigma(XW)]\frac{\partial XW}{\partial W}$$

$$= \frac{Y}{\cancel{\hat{Y}}} * \cancel{\hat{Y}}[1-\hat{Y}]X$$

$$\boxed{\frac{\partial L}{\partial W} = Y(1-\hat{Y})X}$$

· Taking 2nd term:-

$$\frac{\partial L}{\partial W} = \frac{\partial}{\partial W}\left[(1-Y)\log[1-\sigma(XW)]\right]$$

$$= (1-Y)\frac{\partial}{\partial W}\log[1-\sigma(XW)]$$

$$= \frac{(1-Y)}{1-\sigma(XW)}\frac{\partial}{\partial W}(1-\sigma(XW))$$

$$= \frac{(1-Y)}{1-\hat{Y}}\left(-\frac{\partial}{\partial W}\sigma(XW)\right)$$

$$= -\frac{(1-Y)}{(1-\hat{Y})}\sigma(XW)[1-\sigma(XW)]X$$

$$= -\frac{(1-Y)}{\cancel{(1-\hat{Y})}}(\hat{Y})\cancel{(1-\hat{Y})}X$$

$$\boxed{\frac{\partial L}{\partial W} = -\hat{Y}(1-Y)X}$$

old multiply

Dt.:
Pg.:

- Combining both:-

$$\frac{\partial L}{\partial W} = \frac{-1}{m}\left[Y(1-\hat{Y})X - \hat{Y}(1-Y)X \right]$$

$$= \frac{-1}{m}\left[Y - \cancel{Y\hat{Y}} - \hat{Y} + \cancel{\hat{Y}Y} \right] X$$

$$\boxed{\frac{\partial L}{\partial W} = \frac{-1}{m}\left[Y - \hat{Y} \right] X}$$

- Gradient descent:-

$$W_{new} = W_{old} + \eta \frac{\partial L}{\partial W_{old}}$$

$$\boxed{W_{new} = W_{old} - \eta \frac{(Y-\hat{Y})}{m} X}$$