Q.1:- (a) Define the objective function $J(\underline{\beta})$ as

$$J(\underline{\beta}) = \|\underline{y} - \bar{X}\underline{\beta}\|^2 + \lambda\|\underline{\beta}\|^2$$

$$= (\underline{y} - \bar{X}\underline{\beta})^T(\underline{y} - \bar{X}\underline{\beta}) + \lambda\,\underline{\beta}^T\underline{\beta} \quad \text{— ①}$$

To find "$\underline{\beta}$" that minimizes above eq-①; i.e,

$$\left.\frac{d\,J(\underline{\beta})}{d\underline{\beta}}\right|_{\underline{\beta}=\hat{\underline{\beta}}_\lambda} = 0 \Rightarrow \frac{d}{d\underline{\beta}}\left[(\underline{y} - \bar{X}\underline{\beta})^T(\underline{y} - \bar{X}\underline{\beta}) + \lambda\underline{\beta}^T\underline{\beta}\right] = 0$$

$$= -2\bar{X}^T\underline{y} + 2\bar{X}^T\bar{X}\underline{\beta} + 2\lambda\underline{\beta} = 0$$

By simplifying above equation leads to

$$\hat{\underline{\beta}}_\lambda = [\bar{X}^T\bar{X} + \lambda\mathbb{I}]^{-1}\bar{X}^T\underline{y}. \quad \text{— ②} \ast\ast$$

Consider SVD of $\bar{X}$ as ; $\bar{X} = UDV^T$ (D is diagonal matrix)

$$\therefore \hat{\underline{\beta}}_\lambda = [VD^TU^TUDV^T + \lambda\mathbb{I}]^{-1}VD^TU^T\underline{y}$$

$$= [VD^TDV^T + \lambda\mathbb{I}]^{-1}VD^TU^T\underline{y} \quad (\because U^TU = \mathbb{I})$$

$$= [VD^TDV^T + \lambda VV^T]^{-1}VD^TU^T\underline{y} \quad (\because VV^T = \mathbb{I})$$

$$= V[D^TD + \lambda\mathbb{I}]^{-1}DU^T\underline{y} \quad \text{— ③}$$

now Consider $\|\hat{\underline{\beta}}_\lambda\|^2 = \hat{\underline{\beta}}_\lambda^T\hat{\underline{\beta}}_\lambda = \left[\tilde{\underline{y}}^TD[D^TD+\lambda\mathbb{I}]^{-1}V^TV[D^TD+\lambda\mathbb{I}]^{-1}D\tilde{\underline{y}}\right]$

(where $\tilde{\underline{y}} = U^T\underline{y}$ and

$\tilde{\underline{y}}^T\tilde{\underline{y}} = \underline{y}^TUU^T\underline{y} = \underline{y}^T\underline{y}$)

Notice that $D[D^TD+\lambda\mathbb{I}]^{-1}[D^TD+\lambda\mathbb{I}]^{-1}D = \begin{cases} \dfrac{d_i^2}{(d_i^2+\lambda)^2} & ; i=j \\ 0 & ; i\neq j \end{cases}$

$$\therefore \|\hat{\beta}_\lambda\|^2 = \sum_{i=1}^{p} \frac{d_i}{d_i^2+\lambda} y_i^2 \text{ and } \|\hat{\beta}_{\lambda=0}\|^2 = \sum_{i=1}^{p} \frac{d_i}{d_i^2} y$$

as $\left(\frac{d_i}{d_i+\lambda}\right)^2$ decreased as $\lambda \uparrow$

hence $\|\hat{\beta}_\lambda\| \Big|_{\lambda>0} \le \|\beta_\lambda\|_{\lambda=0}$

1. (b)

initialize : $\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\hat{f}_j \equiv 0, \forall i, j$

for $j=1$:

$$\hat{f}_1 \leftarrow S_1\left[\{y_i - \hat{\beta}_0\}_{i=1}^{n}\right]$$

$\because$ linear regression

$\therefore \hat{f}_1(x_{i1}) = \beta_{01} + \beta_{11} x_{i1} = \underline{x}_i^T \underline{\beta}_1$, where $\underline{x}_i = [1 \;\; x_{i1}]^T$

$\underline{\beta}_1 = [\beta_{01} \;\; \beta_{11}]^T$

let $\underset{\sim}{X}_1 = [\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n]^T$

$$\underline{\beta}_1 = (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1^T (\underline{y} - \beta_0)$$

$$\Rightarrow \hat{f}_1(\underset{\sim}{X}_1) = \underset{\sim}{X}_1 \underline{\beta}_1 = \underset{\sim}{X}_1 (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1^T (\underline{y} - \beta_0) \;\#$$

for $j=2$:

$$\hat{f}_2 \leftarrow S_2\left[\{y_i - \hat{\beta}_0 - \hat{f}_1(x_{i1})\}_{i=1}^{n}\right]$$

$\because$ linear regression

$\hat{f}_2(x_{i2}) = \beta_{02} + \beta_{12} x_{i2} = \underline{x}_i^T \underline{\beta}_2$, where $\underline{x}_i = [1 \;\; x_{i2}]^T$

$\underline{\beta}_2 = [\beta_{02} \;\; \beta_{12}]^T$

let $\underset{\sim}{X}_2 = [\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n]^T$

$$\underline{\beta}_2 = (\underset{\sim}{X}_2^T \underset{\sim}{X}_2)^{-1} \underset{\sim}{X}_2^T (\underline{y} - \beta_0 - \underset{\sim}{X}_1 \underline{\beta}_1)$$

$$\Rightarrow \hat{f}_2(\underset{\sim}{X}_2) = \underset{\sim}{X}_2 \underline{\beta}_2 = \underset{\sim}{X}_2 (\underset{\sim}{X}_2^T \underset{\sim}{X}_2)^{-1} \underset{\sim}{X}_2^T (\underline{y} - \beta_0 - \underset{\sim}{X}_1 \underline{\beta}_1) \;\#$$

Q 2

(a) Basis representation of cubic splines with K-Knots is

$$f(x) = \sum_{j=0}^{3} \beta_j x^j + \sum_{k=1}^{K} \beta_{k+3} (x - \xi_k)_+^3$$

for $x < \xi_1$ (first spline)

$(x - \xi_k)_+ = 0 \ \forall k$, i.e.

$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

because of linearity condition

$f'(x) =$ Constant

$\Rightarrow f'(x) = 0 + \beta_1 + 2\beta_2 x + 3\beta_3 x^3$

$\therefore \beta_2 = 0 ; \beta_3 = 0$

for $x > \xi_K$ (last spline)

$(x - \xi_k)_+ = (x - \xi_k)^3 \ \forall k$

$\therefore f(x) = \sum_{j=0}^{3} \beta_j x^j + \sum_{k=1}^{K} \beta_{k+3} (x - \xi_k)^3$

$= \beta_0 + \beta_1 x + \sum_{k=1}^{K} \beta_{k+3} (x - \xi_k)^3$

$(\because \beta_2 = \beta_3 = 0)$

$= \beta_0 + \beta_1 x + \sum_{k=1}^{K} \beta_{k+3} (x^3 - 3x^2 \xi_k + 3x \xi_k^2 - \xi_k^3)$

$f'(x) = \beta_1 + 3x^2 \sum_{k=1}^{K} \beta_{k+3} - 6x \sum_{k=1}^{K} \xi_k \beta_{k+3} + 6 \sum_{k=1}^{K} \xi_k \beta_{k+3}$

(Coefficients of x^2 and $x = 0$)

$\therefore \sum_{k=1}^{K} \beta_{k+3} = 0 ; \sum_{k=1}^{K} \xi_k \beta_{k+3} = 0$

2(b) degree-d spline has degree of freedom of "K+4", where 'K' is knots of cubic spline

Q3:- (a) Let P_1, P_2, P_3 are three predictors, then given

	p_1	p_2	p_3	y
x_1	0	4	1	1
x_2	1	1	-1	-1
x_3	2	4	-1	0
x_4	3	1	1	2

To select a predictor P_j and cut point 's' such that split $R_1(j,s) \triangleq \{p \mid p_j < s\}$ and $R_2(j,s) \triangleq \{p \mid p_j \geq s\}$ yeilds the greatest Rss reduction

i.e find j and s that minimizes

$$\sum_{i: p_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i: p_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 ; \quad \hat{y}_{R_j} = \mathrm{avg}(y_i \mid p_i \in R_j)$$

Consider P_1 ; $s = 3, 2, 1$

for $s=3$, $\underbrace{R_1 = \{x_1, x_2, x_3\}}_{\hat{y}_{R_1} = 0}$; $\underbrace{R_2 = \{x_4\}}_{\hat{y}_{R_2} = 2}$

$Rss = (1-0)^2 + (-1-0)^2 + (0-0)^2 + (0-0)^2$

$= 2$ *

for $s=2$ $\underbrace{R_1 = \{x_1, x_2\}}_{\hat{y}_{R_1} = 0}$; $\underbrace{R_2 = \{x_3, x_4\}}_{\hat{y}_{R_2} = 1}$

$Rss = (1-0)^2 + (1-0)^2 + (0-1)^2 + (2-1)^2$

$= 4$ *

for $s=1$; $R_1 = \{x_1\}$, $R_2 = \{x_2\ x_3 x_4\}$

$Rss = 4.66$

P_2 ; $s = 2$

$\underbrace{R_1 = \{x_2, x_4\}}_{\hat{y}_{R_1} = 1/2}$; $\underbrace{R_2 = \{x_1, x_3\}}_{\hat{y}_{R_2} = 1/2}$

$Rss = (-1-0.5)^2 + (2-0.5)^2 + (1-0.5)^2 + (0-0.5)^2$

$= 5$ *

P_3 ; $s = 0$

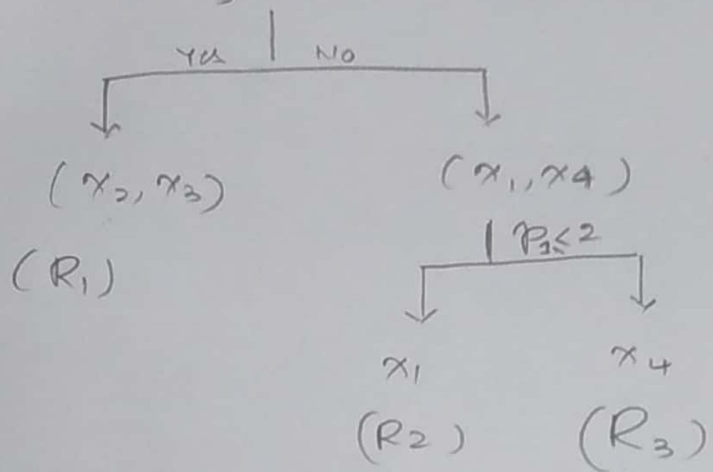
$\underbrace{R_1 = \{x_2\ x_3\}}_{\hat{y}_{R_1} = -1/2}$; $\underbrace{R_2 = \{x_1, x_4\}}_{\hat{y}_{R_2} = 1.5}$

$Rss = (-1+0.5)^2 + (0.5)^2 + (1-1.5)^2 + (2-1.5)^2$

$\Rightarrow (0.5)^2 \times 4 = 1$ *

from above, among all for predictor "P_3"; $s = 0$ has less Res

$\therefore$ first split is $P_3 \le 0$



for second split predictor P_1 and $s = 2$ (Res = 0)

3.(b)

Initialize $\hat{f}(x) = 0$; $r_1 = y_1 = 1$; $r_2 = y_2 = -1$, $r_3 = y_3 = 0$, $r_4 = y_4 = 2$

For $b = 1$ (given $d = 1$; $\lambda = 1$)

(i) Fit a tree $\hat{f}^1(x)$ with $d = 1$ splits to $(\underline{x}, \underline{y} = \underline{r})$

$\therefore$ from previous problem, the required tree is

$x_3 \leq 0$

(x_2, x_3) $\qquad$ (x_1, x_4)

$\hat{y}_{R_1} = y_{avg} = -1/2 = -0.5$ $\qquad$ $\hat{y}_{R_2} = y_{avg} = 3/2 = 1.5$

(ii) $\hat{f}(x) \leftarrow \underbrace{\cancel{\hat{f}(x)}}_{0} + 1 \cdot \hat{f}^1(x) = f^1(x)$

(iii) (Residuals)

$r_1 = r_1 - 1 \cdot \hat{f}(x_1) = 1 - \underbrace{1.5}_{} = -0.5$

$(\because \hat{f}(x_1) = 1.5 \; (\because x_3 > 0))$

$r_2 = r_2 - 1 \cdot \hat{f}(x_2) = -1 + 0.5 = -0.5$

$r_3 = r_3 - 1 \cdot \hat{f}(x_3) = 0 + 0.5 = 0.5$

$r_4 = r_4 - 1 \cdot \hat{f}(x_4) = 2 - 1.5 = 0.5$

For $b = 2$ ($d = 1$, $\lambda = 1$)

(i) fit a tree $\hat{f}^2$ with '1' split to training data (x, r)

previous step

i.e

	x_1	x_2	x_3	r
x_1	0	4	1	-0.5
x_2	1	1	-1	-0.5
x_3	2	4	-1	0.5
x_4	3	1	1	0.5

(ii) by following the similar way as that of previous step, the tree is

$$x_1 \le 2$$

(x_1, x_2) $\qquad$ (x_3, x_4) $\qquad$ (RSS = 0)

$\hat{y}_{R_1} = -0.5$ $\qquad$ $\hat{y}_{R_2} = 0.5$

(iii) Residuals

$$r_1 = -0.5 - \hat{f}^2(x_1) = -0.5 + 0.5 = 0$$

$$r_2 = -0.5 - \hat{f}(x_2) = -0.5 + 0.5 = 0$$

$$r_3 = +0.5 - \hat{f}(x_3) = 0.5 - 0.5 = 0$$

$$r_4 = +0.5 - \hat{f}(x_4) = 0.5 - 0.5 = 0$$

finally $\hat{f}(x) = \hat{f}^1(x) + \hat{f}^2(x)$

$$\therefore f(x = (2,1,-1)) = \hat{f}^1(2,1,-1) + \hat{f}^2(2,1,-1)$$

$$= -0.5 + 0.5 = 0$$

$\therefore \hat{y} = 0$ ※

Q.4:- (a) Desired optimization problem (equivalent form)

$$\max_{\beta_0, \underline{\beta}} \quad \frac{1}{2} \|\underline{\beta}\|^2$$

$$\text{s.t} \quad y_i(\beta_0 + \underline{\beta}^T \underline{x}_i) \geq 1 \quad ; \quad i = 1, 2, \ldots n.$$

the lagrangian function is

$$L(\beta_0, \underline{\beta}, \alpha_i) = \frac{1}{2}\|\underline{\beta}\|^2 - \sum_{i=1}^{n} \left[y_i(\beta_0 + \underline{\beta}^T \underline{x}_i) - 1 \right] \alpha_i \quad \text{--- (1)}$$

the corresponding dual optimization problem is

$$\max_{\alpha_i \geq 0} \underbrace{\min_{\beta_0 \underline{\beta}} L(\beta_0, \underline{\beta}, \alpha_i)}_{L_D(\alpha_i)} \quad \text{--- (2)}$$

notice that

$$\frac{dL(\beta_0, \underline{\beta}, \alpha_i)}{d\beta_0} = -\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \text{--- (3)}$$

and

$$\frac{dL(\beta_0, \underline{\beta}, \alpha_i)}{d\underline{\beta}} = \underline{\beta} - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \quad \text{--- (4)}$$

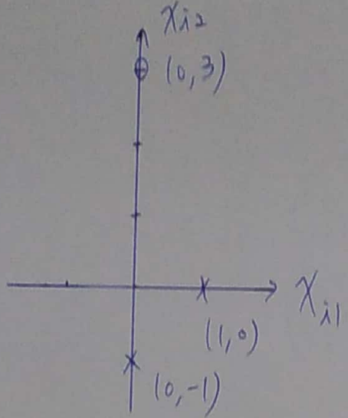
Substitute (2), (3) in (1) leads to

$$L_D(\alpha_i) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \underbrace{x_i^T x_j}_{\langle x_i, x_j \rangle}$$

$\therefore$ from dual problem (2), the solution (ie decision boundary) is

$$f(x) = \beta_0 + \beta^T x = \beta_0 + \sum_{i=1}^{n} \alpha_i y_i \langle x, x_i \rangle$$

4. (b)



$x_1 = (0,-1) = (x_{11}, x_{12})$

$x_2 = (1,0) = (x_{21}, x_{22})$

$x_3 = (0,3) = (x_{31}, x_{32})$

$y_1 = y_2 = 1, \; y_3 = -1$

support vectors: x_2, x_3 #

$w_1 x_{i1} + w_2 x_{i2} + c = 0$

$\Rightarrow (1-0) x_{i1} + (0-3) x_{i2} + c = 0$

$\Rightarrow x_{i1} - 3 x_{i2} + c = 0$

substitutu $\frac{(1,0)+(0,3)}{2}$ into it

$\Rightarrow x_{i1} - 3 x_{i2} + 4 = 0$ #

margin $= \sqrt{(1-0)^2 + (0-3)^2} / 2 = \frac{\sqrt{10}}{2}$ #

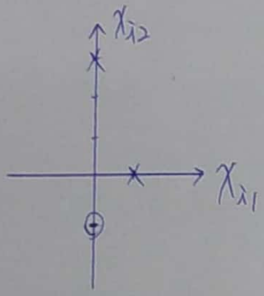
4.(c)

$x_1 = (0, -1) = (x_{11}, x_{12})$, y_1

$x_2 = (1, 0) = (x_{21}, x_{22})$, y_2

$x_3 = (0, 3) = (x_{31}, x_{32})$, y_3

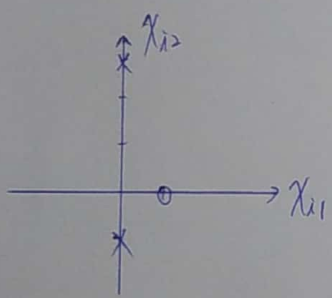
for class y_1



support vectors : x_1, x_2

$$f_1(x) = -x_{i1} - x_{i2}$$

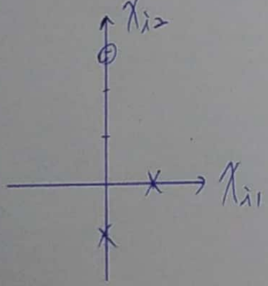
for class y_2



support vectors : x_1, x_2, x_3

$$f_2(x) = x_{i1} - 0.5$$

for class y_3



support vectors : x_2, x_3

$$f_3(x) = -x_{i1} + 3x_{i2} - 4$$

for $x = (4, 2)$

$$k^* = \arg\max_k f_k(x) = \arg\max(-6, 3.5, -2) = 2$$

x belongs to class 2 #

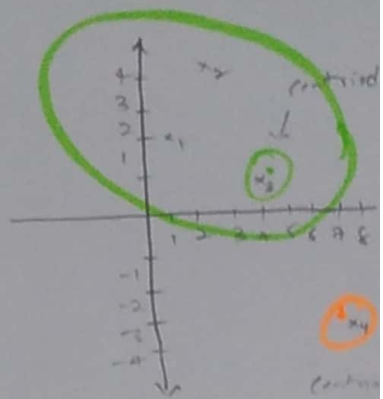
Scanned by CamScanner

Q.5:-(a) Given unsupervised data set $\{x_1\ x_2\ x_3\ x_4\}$, where

$x_1 = (1,2)$; $x_2 = (3,4)$; $x_3 = (4,1)$; $x_4 = (8,-3)$

(a)

datapoint	distance to centroid C_1	C_2
x_1	$\sqrt{10} = 3.16$	$\sqrt{74} = 8.6$
x_2	$\sqrt{10}$	$\sqrt{74}$
x_3	0	$\sqrt{32} = 5.65$
x_4	$\sqrt{32}$	0

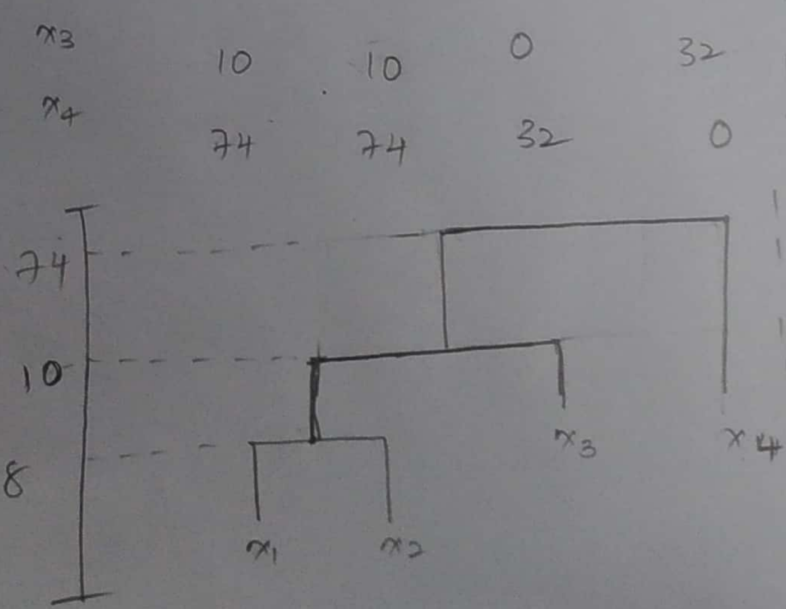


$x_1\ x_2\ x_3$ are close to cluster C_1 and x_4 is close to C_2

$\therefore$ previous clustering remains same

(b) dissimilarity measure (matrix) based on squared point wise distance is as follows

	x_1	x_2	x_3	x_4
x_1	0	8	10	74
x_2	8	0	10	74
x_3	10	10	0	32
x_4	74	74	32	0



Step-1

$d(x_1, x_2) = 8$, $d(x_1, x_3) = 10$;

$d(x_1, x_4) = 74$

Step-2:-

$(x_1, x_2)\ (x_3)\ (x_4)$

$d[(x_1, x_2), x_3] = 10$

$d[(x_1, x_2), x_4] = 32$

Step3

$d[(x_1, x_2, x_3), (x_4)]$

$= 74$

Scanned by CamScanner

5. (c) $X_1 = (X_{11}, X_{12})$, $X_2 = (X_{21}, X_{22})$ and so on.

$$\begin{matrix}(1,2)\\(3,4)\\(4,1)\\(8,-3)\end{matrix} \Rightarrow \begin{cases}\overline{X_{i1}} = 4\\ \overline{X_{i2}} = 1\end{cases}$$

$$X_1 = \begin{bmatrix}1\\3\\4\\8\end{bmatrix} - 4 = \begin{bmatrix}-3\\-1\\0\\4\end{bmatrix}$$

$$X_2 = \begin{bmatrix}2\\4\\1\\-3\end{bmatrix} - 1 = \begin{bmatrix}1\\3\\0\\-4\end{bmatrix}$$

$$\underset{\sim}{X} = [X_1\ X_2] = \begin{bmatrix}-3 & 1\\-1 & 3\\0 & 0\\4 & -4\end{bmatrix}$$

$$\phi_1 = \underset{\|\phi_1'\|^2=1}{\arg\max}\ Var(\phi_{11}' X_1 + \phi_{21}' X_2)$$

= eigenvector corresponding to the max eigenvalue of $\underset{\sim}{X}^T \underset{\sim}{X}$

$$\tilde{X} = (\tilde{X}_1, \tilde{X}_2)^T = X - \phi_1(\phi_1^T X)$$

$$\phi_2 = \underset{\|\phi_2'\|^2=1}{\arg\max}\ Var(\tilde{\phi}_{21}' \tilde{X}_1 + \tilde{\phi}_{22}' \tilde{X}_2)$$

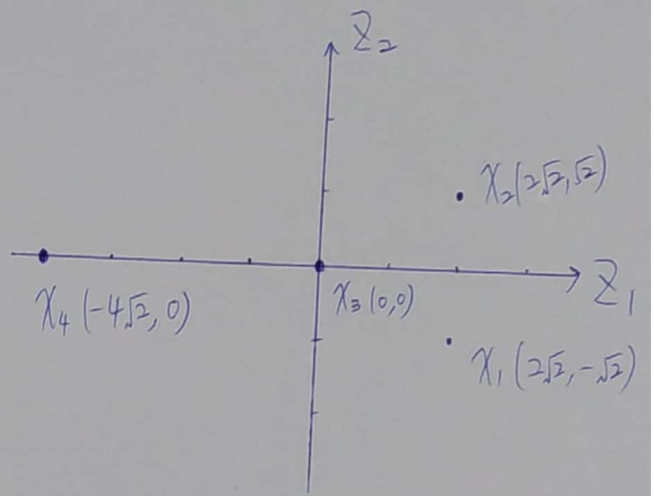
= eigenvector corresponding to the 2^{rd} largest eigenvalue of $\underset{\sim}{X}^T \underset{\sim}{X}$

$$\underset{\sim}{X}^T \underset{\sim}{X} = \begin{bmatrix}26 & -22\\-22 & 26\end{bmatrix}$$

	eigenvalue	eigenvector
$\Rightarrow$	4	$\begin{bmatrix}1\\1\end{bmatrix}$
	28	$\begin{bmatrix}-1\\1\end{bmatrix}$

$$\Rightarrow \begin{cases}Z_1 = -\frac{1}{\sqrt{2}} X_{i1} + \frac{1}{\sqrt{2}} X_{i2}\\ Z_2 = \frac{1}{\sqrt{2}} X_{i1} + \frac{1}{\sqrt{2}} X_{i2}\end{cases}$$

#



#