1) LEARN 'm' and 'c' for least squares fit for data $\{\{x_1 \ldots x_N\}, \{y_1 \ldots y_N\}\}$

$x_i \in R$

Line should pass through (A, B) [2 Marks]

Ans)

$$y = mx + c$$

$$B = mA + c$$

$$\therefore \; y - B = m(x - A) \quad \text{or} \quad y' = mx'$$

Equivalent to transformed $(x_i - A, y_i - B)$ passing through origin $(c = 0)$

$$y' = mx'$$

$$e_i = \left(y'_i - \hat{y}'_i\right)^2$$

$$E = \sum e_i^2 = \sum_i \left(y'_i - \hat{y}'_i\right)^2$$

$$\frac{\partial E}{\partial m} = 0 \Rightarrow \frac{\partial \sum_i \left(y'_i - mx'_i\right)^2}{\partial m} = 0$$

$$\text{or } 2\sum_i \left(y'_i - mx'_i\right)\left(-x'_i\right) = 0$$

$$\text{or } \frac{\sum x'_i y'_i}{\sum x'^2_i} = m$$

Putting

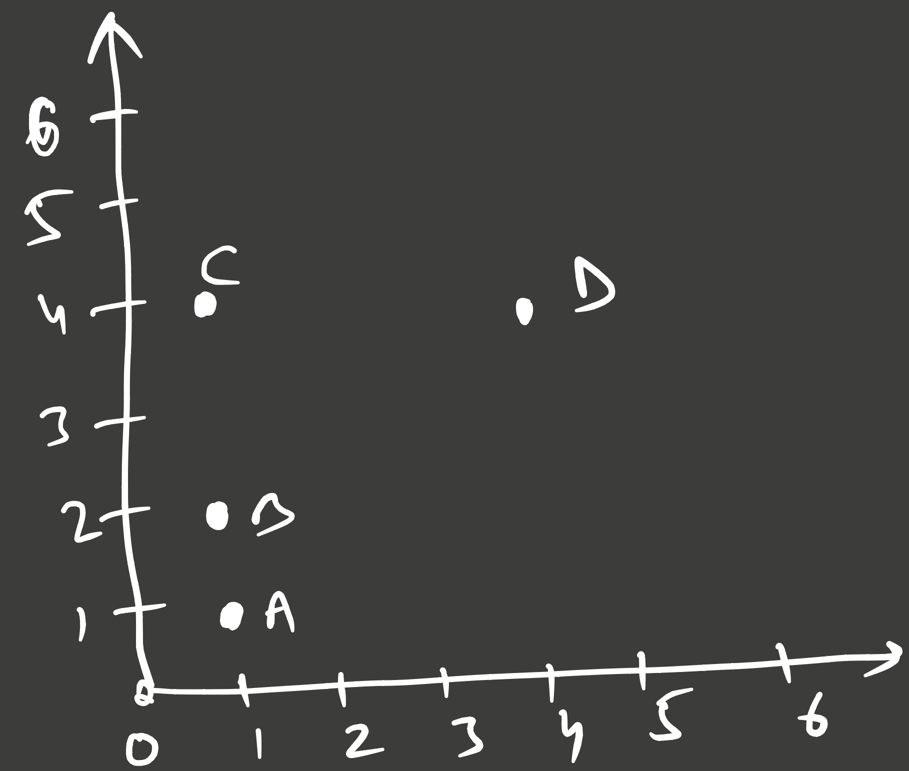
back x', y'

$$m = \frac{\sum_i (x_i - A)(y_i - B)}{\sum_i (x_i - A)^2}$$

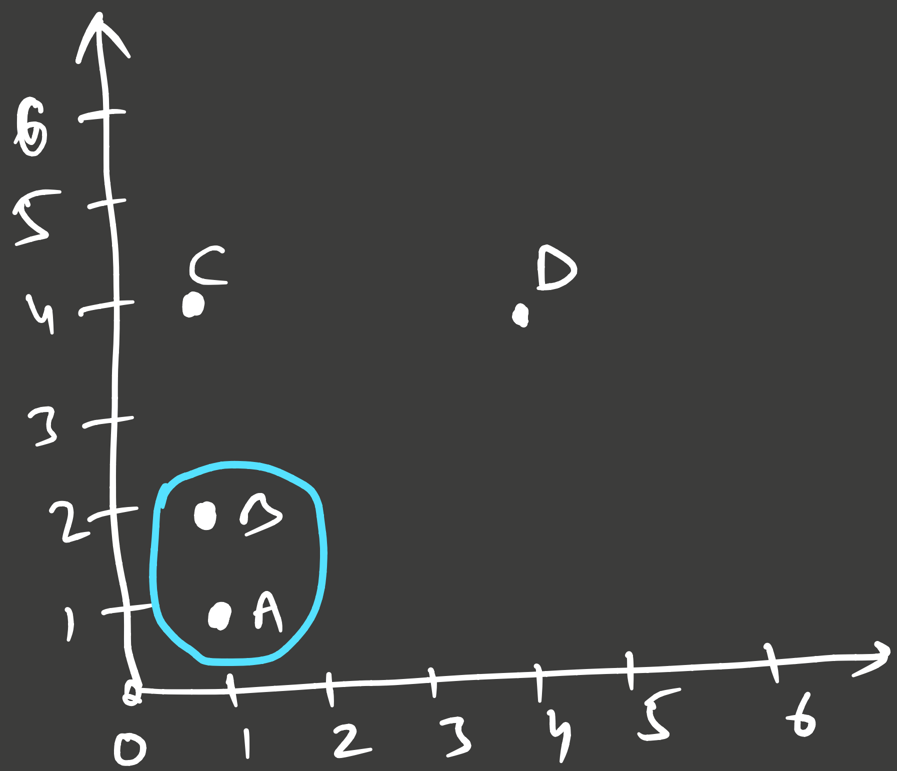
$$c = B - mA$$

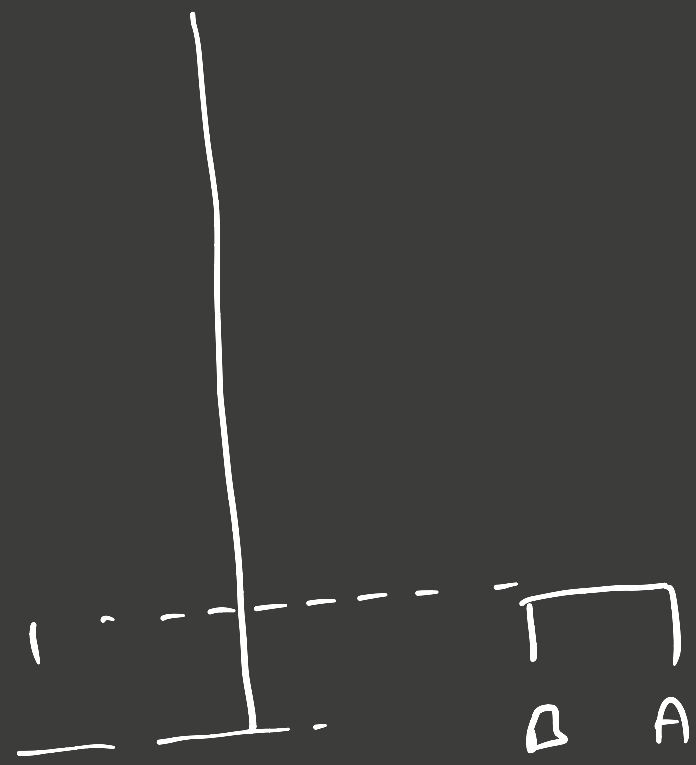
2) Draw a dendogram for following dataset assuming 'single' linkage. Show all steps.

[2 Marks]

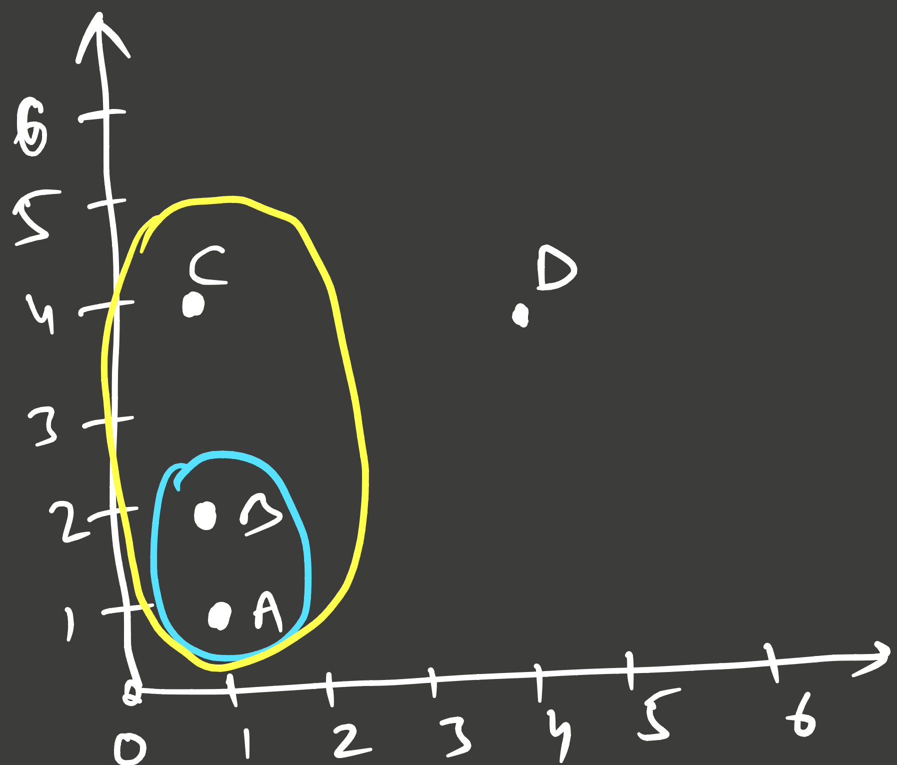


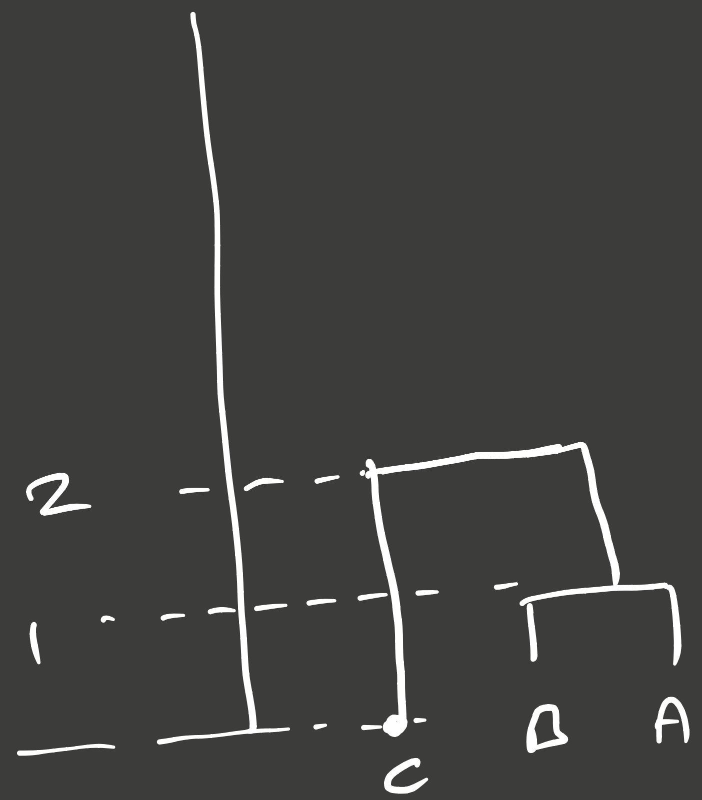
2) Draw a dendogram for following dataset assuming 'single' linkage. Show all steps.





2) Draw a dendogram for following dataset assuming 'single' linkage. Show all steps.

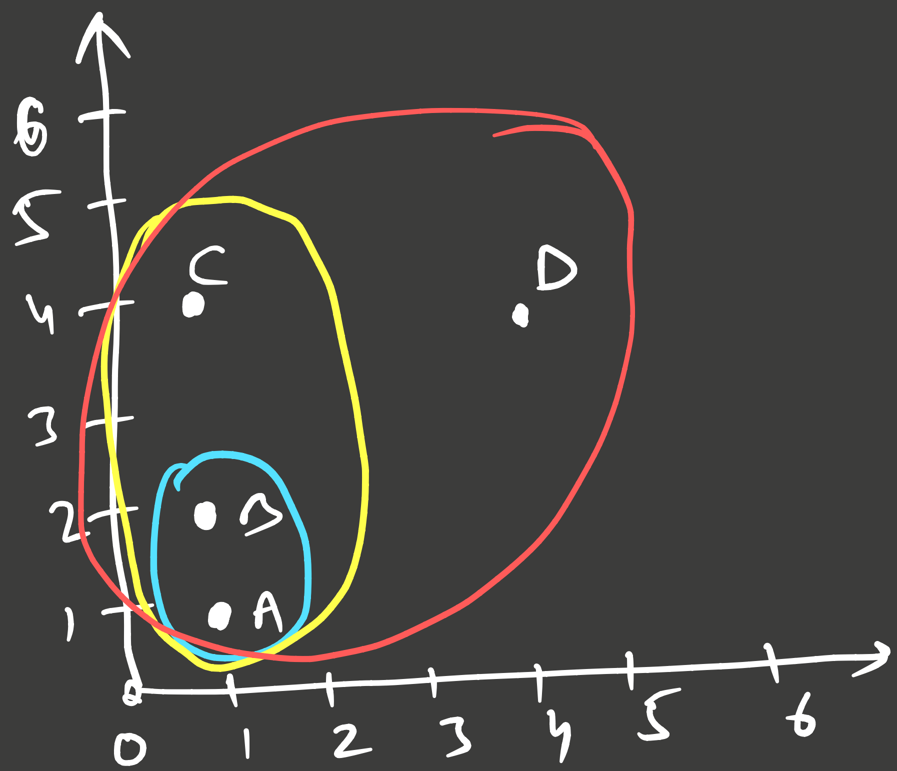


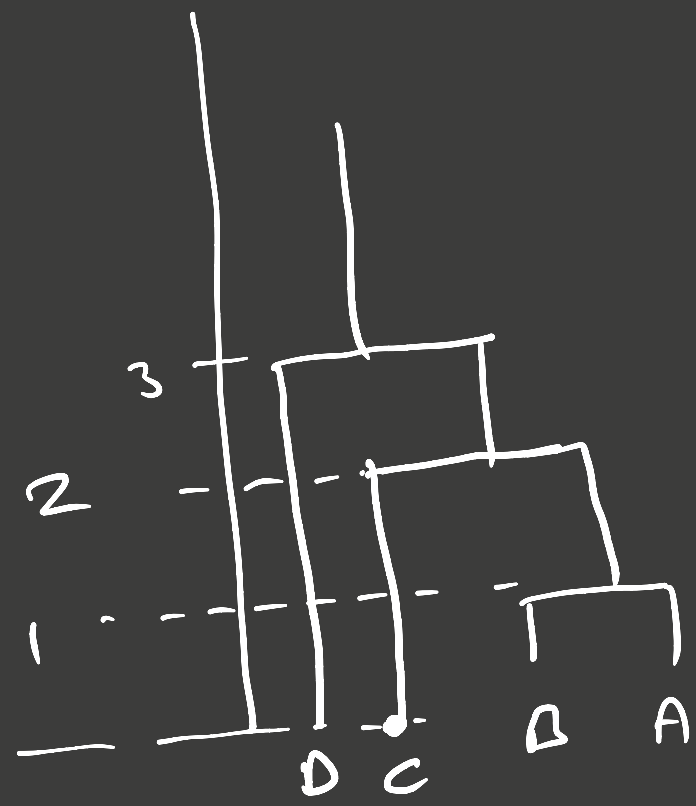


SINGLE LINKAGE LOOKS AT CLOSEST DISTANE

B & C

2) Draw a dendogram for following dataset assuming 'single' linkage. Show all steps.





Q3) Prove that

$$W^{[1]} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad b^{[1]} = \begin{bmatrix} 0 & -1 \end{bmatrix}$$

with ReLU is a solution to XOR problem.

i.e. use forward propagation on

$$X = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

to show the network outputs

$$\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

[2 marks]

MLP FOR XOR (OVER 'm' SAMPLES)

$$X = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} ; \quad y_{TRUE} = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}^T$$

$$w^{[1]} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} ; \quad b^{[1]} = \begin{bmatrix} 0 & -1 \end{bmatrix}$$

$$\therefore Z^{[1]} = \underset{4\times 2}{A^{[0]}} \; \underset{2\times 2}{w^{[1]}} + \underset{1\times 2}{b^{[1]}} \leftarrow \text{Broadcasted}$$

$$Z^{[1]} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 2 & 2 \end{bmatrix} + \begin{bmatrix} 0 & -1 \end{bmatrix}$$

MLP FOR XOR (OVER 'm' SAMPLES)

$$X = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \; ; \; y_{TRUE} = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}^T$$

$$Z^{[1]} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 2 & 2 \end{bmatrix} + \begin{bmatrix} 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -1 \\ 1 & 0 \\ 1 & 0 \\ 2 & 1 \end{bmatrix}$$

$$A^{[1]} = \sigma\left(Z^{[1]}\right) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 2 & 1 \end{bmatrix}$$

MLP FOR XOR (OVER 'm' SAMPLES)

$$X = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \; ; \; y_{TRUE} = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}^T$$

$$A^{[1]} = \sigma\left(Z^{[1]}\right) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 2 & 1 \end{bmatrix}$$

$$W^{[2]} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \; ; \; b^{[2]} = 0$$

$$\therefore Z^{[2]} = A^{[1]} W^{[2]} + b^{[2]} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} + 0 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$$= y_{TRUE}$$

Q4) What is # parameters for a neural net which has image i/ps of size 8X8

&

hidden layer sizes of

100, 20, 1 (o/p layer) (1 Mark)

Ans) $a^{[0]} = [\ldots\ldots]_{1\times 64}$

$W^{[1]} = \begin{bmatrix} & \\ & \end{bmatrix}_{64\times 100}$ $b^{[1]} = 1\times 100$

$$w^{[2]} = 100 * 20 \; ; \quad b^{[2]} = 1 * 20$$

$$w^{[3]} = 1 * 20 \; ; \quad b^{[3]} = 1 * 1$$

$$\therefore \# \text{ Params} = 6400 + 100 + 2000 + 20 + 20 + 1$$

Ans 5)

Rewrite dataset

$$\begin{array}{ccc} a & a & b \\ -a & -a & -b \end{array}$$

We can see that $\theta_0 = 0$ ∴

$$\therefore X = \begin{bmatrix} a & a \\ -a & -a \end{bmatrix} \qquad y = \begin{bmatrix} b \\ -b \end{bmatrix}$$

$$X^T X = \begin{bmatrix} a & -a \\ a & -a \end{bmatrix} \begin{bmatrix} a & a \\ -a & -a \end{bmatrix} = \begin{bmatrix} 2a^2 & 2a^2 \\ 2a^2 & 2a^2 \end{bmatrix}$$

$$X^T X + \delta^2 I = \begin{bmatrix} 2a^2+\delta^2 & 2a^2 \\ 2a^2 & 2a^2+\delta^2 \end{bmatrix}$$

$$\det\left(X^T X + \delta^2 I\right) = \left(2a^2+\delta^2\right)^2 - \left(2a^2\right)\left(2a^2\right)$$

$$= 4a^4 + \delta^4 + 4a^2\delta^2 - 4a^4$$

$$= \delta^4 + 4a^2\delta^2$$

$$\left(X^T X + \delta^2 I\right)^{-1} = \frac{1}{\delta^4 + 4a^2\delta^2} \underset{2\times 2}{\begin{bmatrix} 2a^2+\delta^2 & -2a^2 \\ -2a^2 & 2a^2+\delta^2 \end{bmatrix}}$$

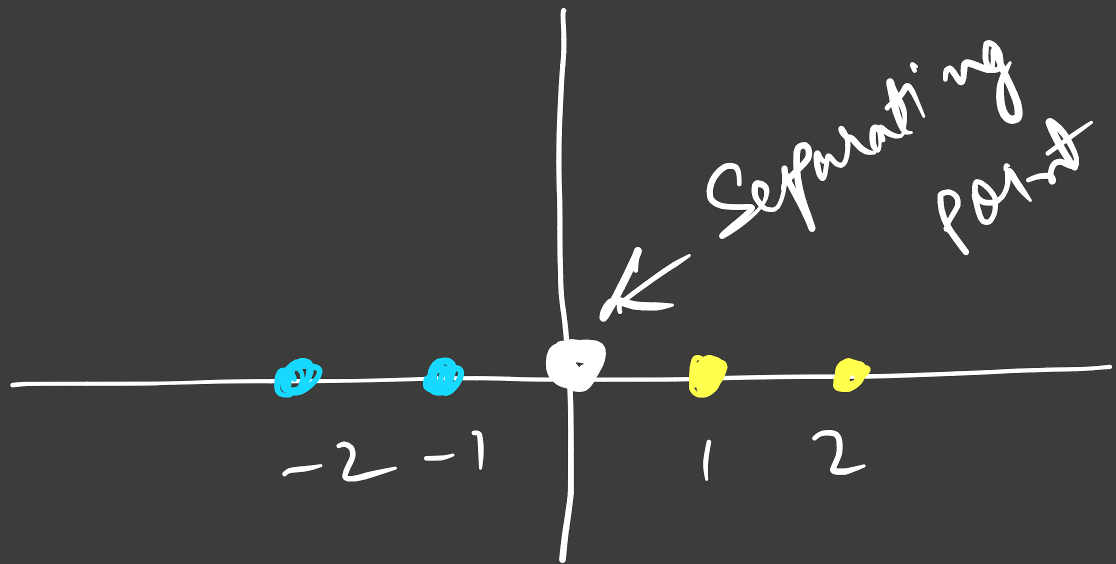
$$X^T y = \underset{2\times 2}{\begin{bmatrix} a & -a \\ a & -a \end{bmatrix}} \underset{2\times 1}{\begin{bmatrix} b \\ -b \end{bmatrix}} = \underset{2\times 1}{\begin{bmatrix} 2ab \\ 2ab \end{bmatrix}} = k\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$(X^\top X + \delta^2 I)^{-1}(X^\top y) = k_1 \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix} k_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= k_1 k_2 \begin{bmatrix} \alpha + \beta \\ \alpha + \beta \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

$$\therefore \quad \theta_1 = \theta_2$$

Answer 6)



4 points

x_1	y
1	1
2	1
-1	-1
-2	-1

Separating hyperplane: $w_1 x_1 + b = 0$

$$y_i(w_1 x_1 + b) \geq 1$$

$$\Rightarrow 1(w_1 + b) \geq 1 \quad \ldots (1)$$

$$1(2w_1 + b) \geq 1 \quad \ldots (2)$$

$$-1(-w_1 + b) \geq 1 \quad \ldots (3)$$

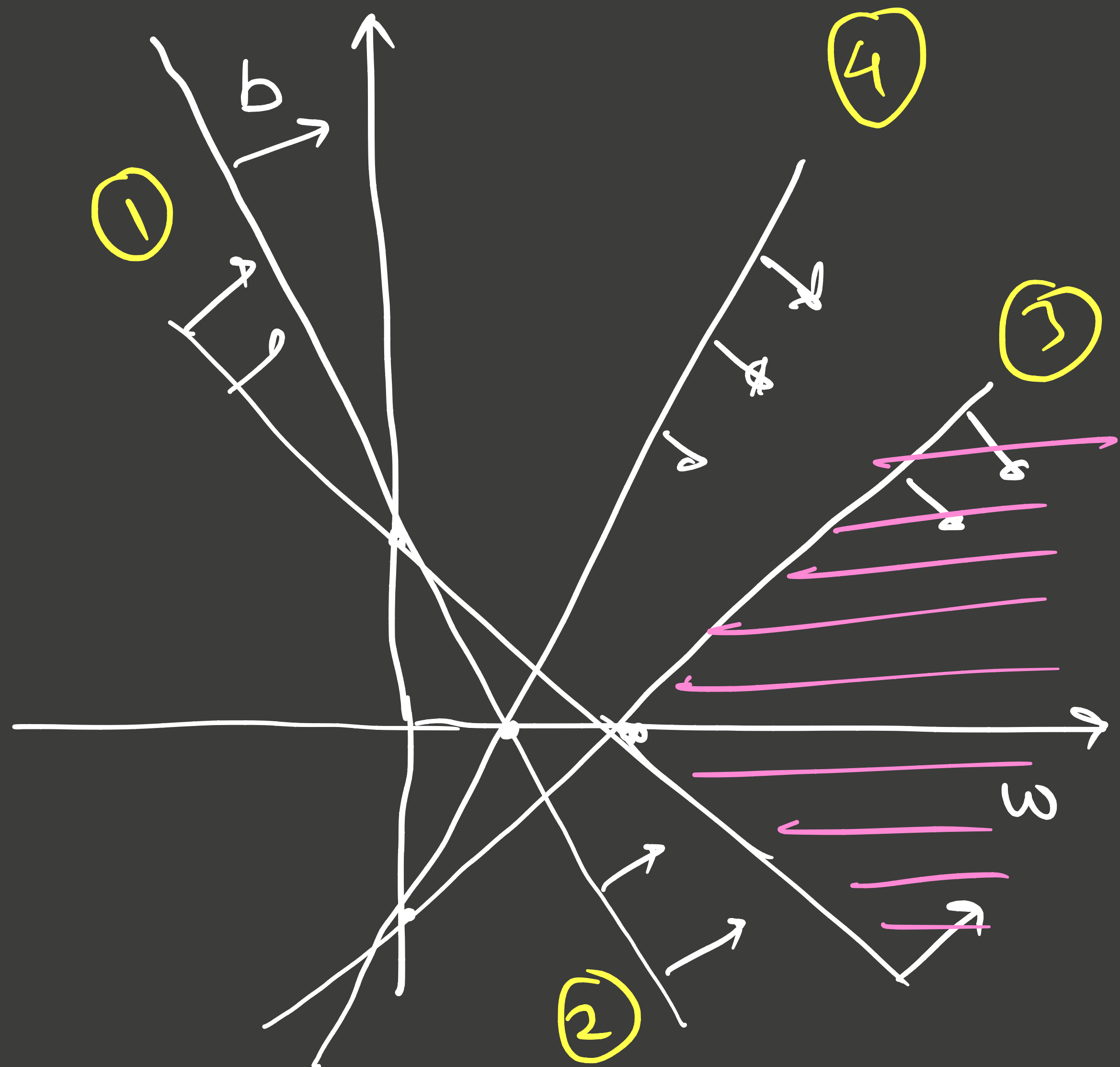
$$-1(-2w_1 + b) \geq 1 \quad \ldots (4)$$

$$1\,(w_1 + b) \geq 1 \quad \cdots (1)$$

$$1\,(2w_1 + b) \geq 1 \quad \cdots (2)$$

$$-1\,(-w_1 + b) \geq 1 \quad \cdots (3) \Rightarrow w_1 - b \geq 1$$

$$-1\,(-2w_1 + b) \geq 1 \quad \cdots (4) \Rightarrow 2w_1 - b \geq 1$$



$$w_{min} = 1$$

$$b = 0$$

$$\vec{w} \cdot \bar{x} + b = 0$$

or $\quad x = 0$

Minimum w_1 satisfying constraints is

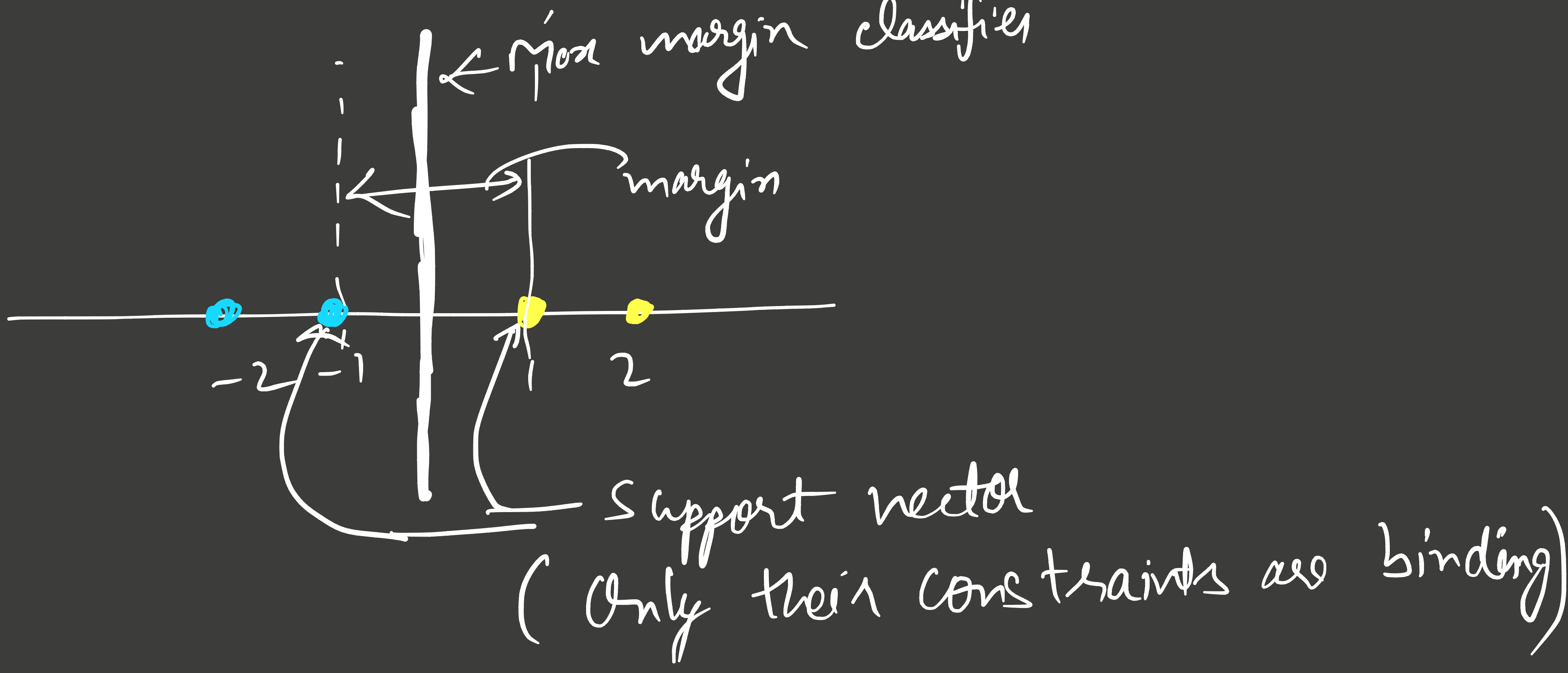
$w_1 = 1$

&

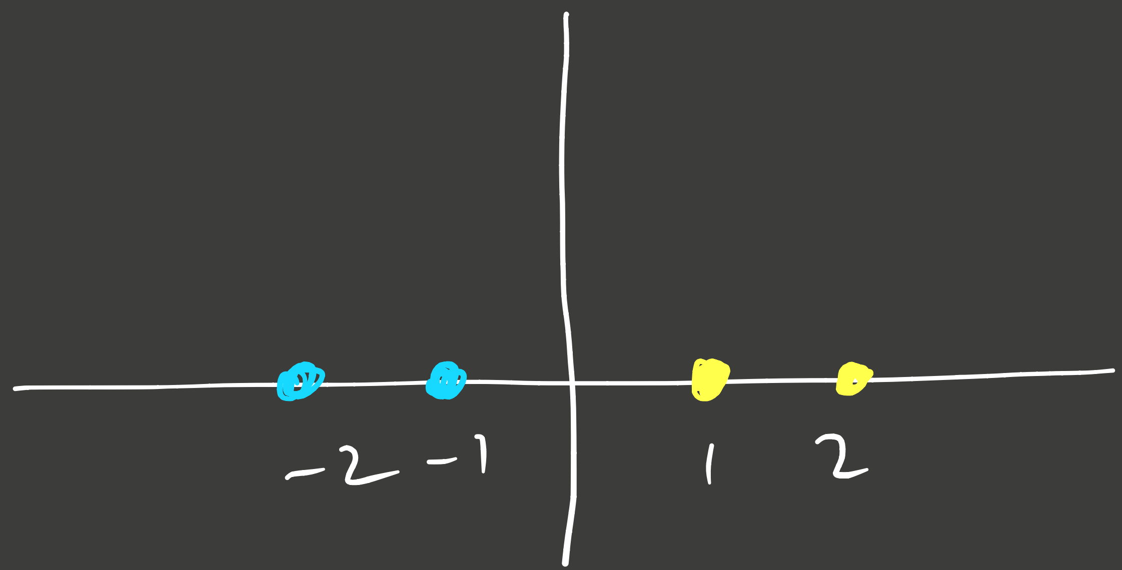
correspondingly $b = 0$

$\therefore$ Max. margin classifier is $1 \cdot x + 0 = 0$

or $x = 0$



Revisiting the simple example (1-D) in dual.



4 points

x_i	y	
1	1	α_1
2	1	α_2
-1	-1	α_3
-2	-1	α_4

$$L(\alpha) = \sum_{i=1}^{4} \alpha_i - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j \vec{x_i} \vec{x_j}$$

s.t. $\alpha_i \geq 0$

$$\sum \alpha_i y_i = 0$$

$$\alpha_i \left(y_i (\vec{w} \cdot \vec{x_i} + b) - 1 \right) = 0$$

Revisiting the simple example (1-D) in dual.

$$L(\alpha) = \sum_{i=1}^{4} \alpha_i - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j \vec{x_i} \vec{x_j}$$

s.t. $\alpha_i \geq 0$

$\sum \alpha_i y_i = 0$

$\alpha_i (y_i (\vec{w} \cdot \vec{x_i} + b) - 1) = 0$

$\alpha_2 = \alpha_4 = 0 \because$ pt 2 and 4 are NOT S.V.

$\alpha_1 y_1 + \alpha_3 y_3 = 0$

$\alpha_1 (+1) + \alpha_3 (-1) = 0 \Rightarrow \alpha_1 = \alpha_3 = \alpha \text{ (say)}$

$$L(\alpha) = (\alpha_1 + \alpha_3) - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} (\alpha_i \alpha_j y_i y_j \bar{x}_i \cdot \bar{x}_j)$$

$$= 2\alpha - \frac{1}{2} \Big\{ \alpha_1 \alpha_3 y_1 y_3 x_1 x_3$$

$$+ \alpha_3 \alpha_1 y_3 y_1 x_3 x_1$$

$$+ \alpha_1^2 y_1^2 x_1^2$$

$$+ \alpha_3^2 y_3^2 x_3^2 \Big\}$$

$$L(\alpha) = 2\alpha - \frac{1}{2}\Big[\alpha^2(1)(-1)(1)(-1) + \alpha^2(-1)(1)(-1)(1)$$

$$+ \alpha^2(1)(1) + \alpha^2(1)(1)\Big\}$$

$$= 2\alpha - 2\alpha^2$$

$$\frac{\partial L(\alpha)}{\partial \alpha} = 0 \quad \text{to maximize}$$

$$\text{or} \quad 2 - 4\alpha = 0 \Rightarrow \alpha = \frac{1}{2}$$

$$\alpha_1 = \alpha_3 = \frac{1}{2}; \; \alpha_2 = \alpha_4 = 0$$

$$\vec{w} = \sum_{i=1}^{4} \alpha_i \, y_i \, \vec{x_i}$$

$$= \alpha_1 * 1 * 1 + \alpha_3 * (-1)(-1)$$

$$= \frac{1+1}{2} = 1$$

FOR SUPPORT VECTORS WE HAVE

$$\alpha_i = 0$$

&

$$y_i (\vec{w} \cdot \vec{x}_i + b) - 1 = 0 \qquad i \in S.V.$$

or $y_i (\bar{w} \cdot \bar{x}_i + b) = 1$

or $y_i^2 (\bar{w} \cdot \bar{x}_i + b) = y_i$

or $\bar{w} \cdot \bar{x}_i + b = y_i \qquad (\because y_i^2 = 1)$

or $b = y_i - \bar{w} \, \bar{x}_i$

In practice $b = \dfrac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (y_i - \bar{w} \cdot \vec{x}_i)$

So.,

$$b = \frac{1}{2}\left\{\left(1 - (1)(1)\right) + \left(-1 - (1)(-1)\right)\right.$$

$$= \frac{1}{2}\left\{0 + 0\right\} = 0$$

$\therefore \quad w = 1$

&

$b = 0$

$$L(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$$

$$-\frac{1}{2}\Big\{ \alpha_1 \alpha_1 * (1*1) * (1*1)$$

$$+$$

$$\alpha_1 \alpha_2 * (1*1) * (1*2)$$

$$+$$

$$\alpha_1 \alpha_3 * (1*-1) * (1*-1)$$

$$+$$

$$\cdots$$

$$\cdots$$

$$+$$

$$\cdots$$

$$\alpha_4 \alpha_4 * (-1*-1) * (-2*-2)$$

$$\Big\}$$

How to solve? QP solver

FOR TRIVIAL EXAMPLE,

BY SYMMETRY, $\alpha_1 = \alpha_3 = \alpha$ (say)

& (ALSO $\sum y_i \alpha_i = 0$)

$\alpha_2 = \alpha_4 = 0$ (NON S.V.)

MAXIMIZE $(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)$

$$-\frac{1}{2}\Big[\alpha_1\alpha_3 (1)(-1)(1)(-1)$$

$$+ \alpha_3\alpha_1 (-1)(-1)(1)(-1)\Big]$$

MAXIMIZE $2\alpha - \frac{1}{2}(\alpha^2 + \alpha^2)$
α

$$\frac{\partial}{\partial a}\left(2\alpha - \alpha^2\right) = 0 \Rightarrow 2 - 2\alpha = 0$$

$$\Rightarrow \alpha = 1$$

$$\frac{\partial}{\partial \alpha^2}\left(2\alpha - \alpha^2\right) = 1$$

$\therefore$ Maxima.

$\therefore \alpha_1 = \alpha_3 = 1;\ \alpha_2 = \alpha_4 = 1$

$$\vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x}_i = \alpha_1 y_1 x_1 + \alpha_3 y_3 x_3$$

$$= 1*$$

FINDING 'b'

FOR SUPPORT VECTORS WE HAVE

$\alpha_i = 0$

&

$y_i(\vec{w} \cdot \vec{x}_i + b) - 1 = 0 \qquad i \in S.V.$

or $y_i(\bar{w} \cdot \bar{x}_i + b) = 1$

or $y_i^2(\bar{w} \cdot \bar{x}_i + b) = y_i$

or $\bar{w} \cdot \bar{x}_i + b = y_i \qquad (\because y_i^2 = 1)$

or $b = y_i - \bar{w}\,\bar{x}_i$

In practice $b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (y_i - \bar{w} \cdot \vec{x}_i)$

Q6)