CS 551 - Deep Learning
Mid Sem Assignment

Name : P. V. Sriram
Roll : 1801CS37

1) Differentiating w.r.t w_0 and w_1 will give us the conditions

$$\frac{\partial (J(w))}{\partial w_0} = \frac{\partial}{\partial w_0}\left(\frac{1}{2n}\sum_{i=1}^{n}(y_i - w_0 - w_1 x_i)^2\right)$$

$$\Rightarrow \frac{1}{2n}\left(2\sum_{i=1}^{n}(y_i - w_0 - w_1 x_i)\right) = 0$$

$$= \frac{1}{n}\sum_{i=1}^{n}(y_i - w_0 - w_1 x_i) = 0$$

III^{ly},

$$\frac{\partial}{\partial w_1}(J(w)) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w_0 - w_1 x_i)\, x_i = 0$$

This means that the prediction error $(y_i - w_0 - w_1 x_i)$ does not vary with any linear function of inputs.

So from the above conditions, we get

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - w_0^* - w_1^* x_i)(x_i - \bar{x}) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - w_0^* - w_1^* x_i)(w_0^* + w_1^* x_i) = 0$$

as

$(x_i - \bar{x})$ and $(w_0^* + w_1^* x_i)$ are both

linear functions of inputs.

2) We know that,

Likelihood under the iid assumption:

$$L(D, \theta) = \prod_{i=1}^{m} P_\theta(x_i)$$

Applying ln on both sides

$$\log(L(D, \theta)) = \sum_{i=1}^{m} \log(P_\theta(x_i))$$

here, $P_\theta(x) = 2\theta e^{-\theta x^2}$

$$\Rightarrow \log(L(D, \theta)) = \sum_{i=1}^{m} \log(2\theta e^{-\theta x^2})$$

$$= \sum_{i=1}^{m} (\log 2 + \log \theta + \log x_i - \theta x_i^2)$$

Taking the derivative w.r.t θ,

$$\frac{\partial(\operatorname{Log}(L(D, \theta)))}{\partial \theta} = \sum_{i=1}^{m} \left(\frac{1}{\theta} - x_i^2\right)$$

$$= \frac{m}{\theta} - \sum_{i=1}^{m} x_i^2$$

For MLE, $\frac{\partial(\operatorname{Log}(L(D, \theta)))}{\partial \theta} = 0$

$$\Rightarrow \frac{m}{\theta} - \sum_{i=1}^{m} x_i^2 = 0$$

$$\therefore \boxed{\theta = \frac{m}{\sum_{i=1}^{m} x_i^2}}$$

3) Given,

Dataset with 2 inputs x_1, x_2

and output y

a)

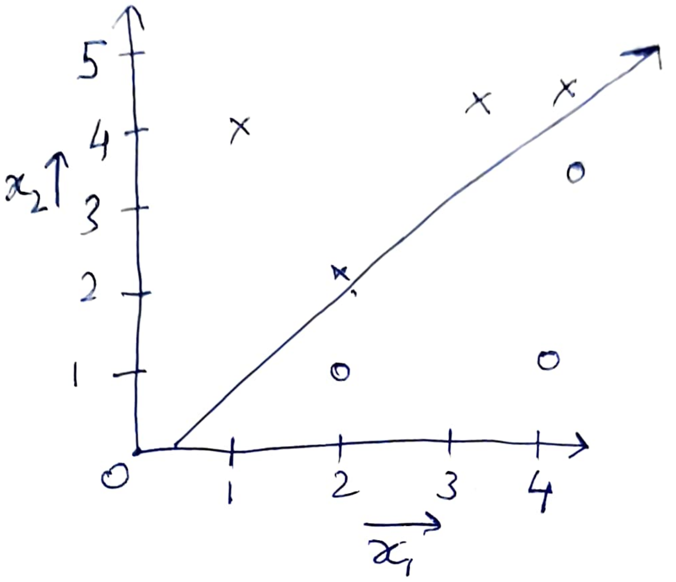
X_1	X_2	Y
3	4	R
2	2	R
4	4	R
1	4	R
2	1	B
4	3	B
4	1	B

$\Rightarrow$ The optimal separating hyperplane has to be between the observations $\{(2,1)$ and $(2,2)\}$ and between the observations $\{(4,3)$ and $(4,4)\}$

So it is a line that passes through the points $(2, 1.5)$ and $(4, 3.5)$

$\Rightarrow x_1 - x_2 = 0.5$

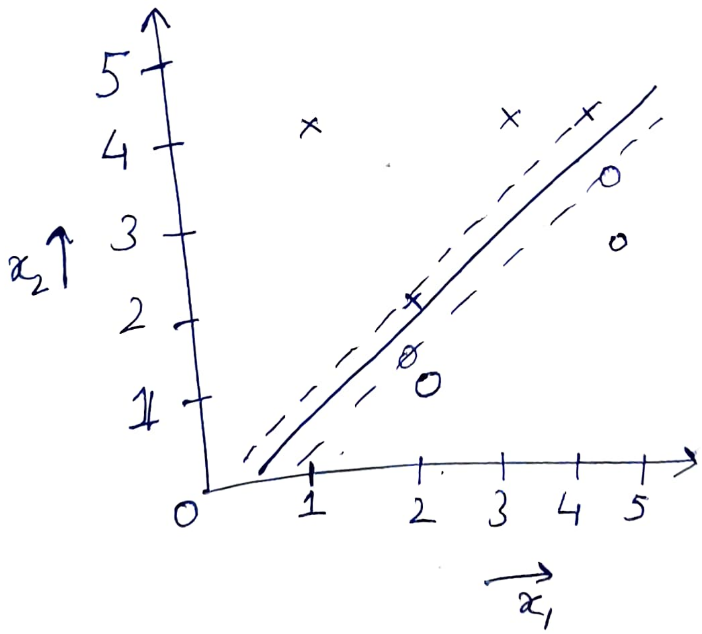
Plot



Optimal separating hyperplane

$\Rightarrow \boxed{x_1 - x_2 = 0.5}$

b) Margin for maximal margin hyperplane



$\boxed{\text{Margin} = \frac{1}{2\sqrt{2}}}$

~~All the (2,1) from~~ All the support vectors are equi distant from the plane, i.e) $\frac{|2-1-0.5|}{\sqrt{2}} = \frac{1}{2\sqrt{2}}$

c) Support vectors are the data points closer to hyperplane which influences the position and orientation of the hyperplane

Here, The support vectors are

$\Rightarrow (2,1), (2,2), (4,3), (4,4)$

4) We need to represent the given points in a lower dimensional space using PCA.

I :

We subtract the mean from the data to make it ready for further computations and to give equal importance to all features

	a	b	c	d		
x	1	3	2	4	$\Sigma x = 10$	$\bar{x} = 2.5$
y	2	4	1	3	$\Sigma y = 10$	$\bar{y} = 2.5$

$\vec{\mu} = (2.5, 2.5)$

Now, new data $X = \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \mu, \ \forall i \in \{0,3\}$

New points,

x	-1.5	0.5	-0.5	1.5
y	-0.5	1.5	-1.5	0.5

II :

Covariance matrix (C)

$$\Rightarrow \frac{1}{N-1}\left(\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T\right)$$

$$= \frac{1}{3}\left(\begin{bmatrix}-1.5\\-0.5\end{bmatrix}\begin{bmatrix}-1.5 & -0.5\end{bmatrix} + \begin{bmatrix}0.5\\1.5\end{bmatrix}\begin{bmatrix}0.5 & 1.5\end{bmatrix}\right.$$

$$\left. + \begin{bmatrix}-0.5\\-1.5\end{bmatrix}\begin{bmatrix}-0.5 & -1.5\end{bmatrix} + \begin{bmatrix}1.5\\0.5\end{bmatrix}\begin{bmatrix}1.5 & 0.5\end{bmatrix}\right)$$

$$= \frac{1}{3}\begin{bmatrix}5 & 3\\3 & 5\end{bmatrix} \Rightarrow C = \begin{bmatrix}1.\overline{66} & 1\\1 & 1.\overline{66}\end{bmatrix}$$

III

Finding the principal components of the above data.

→ We find the eigen values and the eigen vectors of the covariance matrix C.

The formula is given by $CE = \lambda E$, where E is the eigen vector and λ is the eigen value.

To prove this, we have

$$|C - \lambda I| = 0 \Rightarrow \begin{vmatrix} 1.67-\lambda & 1 \\ 1 & 1.67-\lambda \end{vmatrix} = 0$$

$$(1.67-\lambda)^2 - 1 = 0 \Rightarrow \lambda = 1.67 \pm 1$$

$$\underline{\underline{\lambda_1 = 2.67}}, \ \underline{\underline{\lambda_2 = 0.67}}$$

Now, we see that $\lambda_1 > \lambda_2$. We consider λ_1 for obtaining the principle component as it has higher value

Now, we have $CE = 2.67E$

$$\Rightarrow \begin{bmatrix} 1.67 & 1 \\ 1 & 1.67 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} 2.67 e_1 \\ 2.67 e_2 \end{bmatrix}$$

$$\Rightarrow 1.67 e_1 + e_2 = 2.67 e_1$$

and

$$e_1 + 1.67 e_2 = 2.67 e_2$$

$$\Rightarrow e_1 = e_2 = 1$$

And making this a unit vector,

$$E = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

IV

Projecting the original data points along the principle component (Eigen value E)

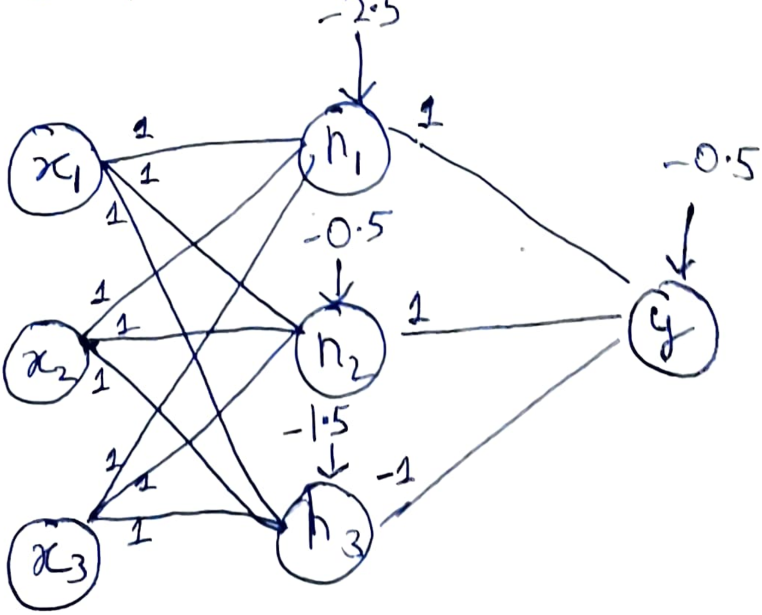
$$\begin{bmatrix} -1.5 & 0.5 \\ 0.5 & 1.5 \\ -0.5 & -1.5 \\ 1.5 & 0.5 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} -1.414 & -1.414 & -1.414 \\ & & -1.414 \end{bmatrix}$$

5) Required to replicate 3 input XOR gate using multilayer Perceptron.

i.e) 3 input XOR :-

x_1	x_2	x_3	y
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

Proposed MLP :-



A single hidden layered perceptron model is proposed, with the given weights and biases.

Important to note that the activation function used at each step is a hard sigmoid function

i.e) $\sigma(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$

$$\text{pre-}h_1 = x_1 + x_2 + x_3 - 2.5$$
$$h_1 = \sigma(\text{pre-}h_1)$$

$$\text{pre-}h_2 = x_1 + x_2 + x_3 - 0.5$$
$$h_2 = \sigma(\text{pre-}h_2)$$

$$\text{pre-}h_3 = x_1 + x_2 + x_3 - 1.5$$
$$h_3 = \sigma(\text{pre-}h_3)$$

$$\text{pre-}Y = h_1 + h_2 + h_3 - 0.5$$
$$Y = \sigma(\text{pre-}Y)$$

Inputs			Internal Nodes						Output	
x_1	x_2	x_3	Pre-h_1	h_1	Pre-h_2	h_2	Pre-h_3	h_3	Pre-y	y
0	0	0	−2.5	0	−0.5	0	−1.5	0	−0.5	0
0	0	1	−1.5	0	0.5	1	−0.5	0	0.5	1
0	1	0	−1.5	0	0.5	1	−0.5	0	0.5	1
0	1	1	−0.5	0	1.5	1	0.5	1	−0.5	0
1	0	0	−1.5	0	0.5	1	−0.5	0	0.5	1
1	0	1	−0.5	0	1.5	1	0.5	1	−0.5	0
1	1	0	−0.5	0	1.5	1	0.5	1	−0.5	0
1	1	1	0.5	1	2.5	1	1.5	1	0.5	1