①

Name : Ruchit Bhardwaj
USC ID : 1111-4177-99

(Q1) a) layer 1 = 3 neurons
layer 2 = 3 neurons
~~layer 3 = 3 neurons~~.

layer ①
$$n(1) = \underline{w}(1)^T \cdot \underline{x} + \underline{b}(1) \Rightarrow n^{(1)}_{3\times 1}.$$
$$a(1) = F(1)(n(1)) = a(1)_{3\times 1} = 3 \text{ neurons}.$$

layer ②
$$n(2) = \underline{w}(2)^T \cdot \underline{a}(1) + \underline{b}(2)$$
$$= \underset{3\times3}{w(2)^T}\ \underset{3\times1}{a(1)} + b(2)_{3\times 1} = n(2)_{3\times 1}$$
$$a(2) = F(2)[n(2)]$$
$$= a(2)_{3\times 1} \rightarrow 3 \text{ neurons}$$

~~layer 3~~

(b) 3×1 goes to softmax which finds prob of each class when a(2) is 3×1 & softmax also gives 3×1 output

so $\boxed{K = 3}$

(c) ②

(c) $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$n(1) = a(1) \rightarrow$ because $f^{(1)}$ is linear.

$$n(1) = a(1) = \underline{W}^{(1)T}\underline{X} + b(1) = \begin{bmatrix} 0 & -1 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

$$a(1) = \begin{bmatrix} -x_2 \\ x_1 + 2x_2 - 1 \\ x_2 \end{bmatrix}$$

$n(2) = a(2) \rightarrow$ because $f^{(1)}$ is linear.

$$= \begin{bmatrix} -1 & 1 & 0 \\ 2 & 2 & 0 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} -x_2 \\ x_1 + 2x_2 - 1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} x_2 + x_1 + 2x_2 - 1 + 1 \\ -2x_2 + 2x_1 + 4x_2 - 2 \\ x_2 + 1 \end{bmatrix} = \begin{bmatrix} x_1 + 3x_2 \\ 2x_1 + 2x_2 - 2 \\ x_2 + 1 \end{bmatrix}$$

$\downarrow$

all are multinomial regressions

③

This is passed to logit (softmax)

$$P_1(x) = \frac{e^{x_1+3x_2}}{e^{x_1+3x_2} + e^{2x_1+2x_2-2} + e^{x_2+1}}$$

$$Den = \text{Denominator} = e^{x_1+3x_2} + e^{2x_1+2x_2-2} + e^{x_2+1}$$

$$P_2(x) = \frac{e^{2x_1+2x_2-2}}{Den} = \frac{e^{2x_1+2x_2-2}}{e^{x_1+3x_2} + e^{2x_1+2x_2-2} + e^{x_2+1}}$$

$$P_3(x) = \frac{e^{x_2+1}}{Den} = \frac{e^{x_2+1}}{e^{x_1+3x_2} + e^{2x_1+2x_2-2} + e^{x_2+1}}$$

Q2 $\pi = [0.3 \quad 0.7]$

$O = \{VL, S\}$

a

	C	H
C	0.7	0.3
H	0.1	0.9

	VS	S	M	L	VL
C	0.2	0.3	0.3	0.1	0.1
H	0.2	0.2	0.2	0.3	0.1

VL S

states		P(O | states)	
C	C	$0.3(0.1)(0.7)(0.3)$	$= 63\times10^{-4}$
C	H	$0.3(0.1)(0.3)(0.2)$	$= 18\times10^{-4}$
H	C	$0.7(0.1)(0.1)(0.3)$	$= 21\times10^{-4}$
H	H	$0.7(0.1)(0.9)(0.2)$	$= 126\times10^{-4}$

$$P(O) = \sum_{states} P(O|states) = 228\times10^{-4} = \boxed{0.0228}$$

b

	C	H
C	0.6	0.4
H	0.4	0.6

	VS	S	M	L	VL
C	0.1	0.2	0.2	0.1	0.4
H	0.5	0.2	0.1	0.1	0.1

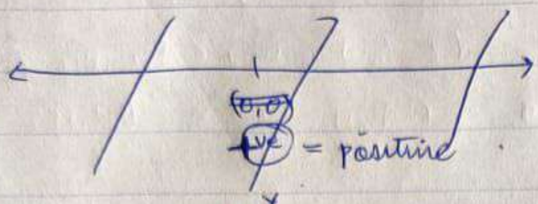
VL S.

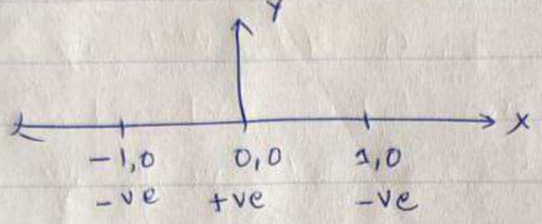
S		P(O|S)	
C	C	$0.3(0.4)(0.6)(0.2)$	$= 144\times10^{-4}$
C	H	$0.3(0.4)(0.4)(0.2)$	$= 96\times10^{-4}$
H	C	$0.7(0.1)(0.4)(0.2)$	$= 56\times10^{-4}$
H	H	$0.7(0.1)(0.6)(0.2)$	$= 84\times10^{-4}$

$$P(O) = \sum_x P(O|X) = 380\times10^{-4}$$

$$\boxed{0.038}$$

Q3) a)





⊕ = (0,0)

− = (+1,0)

− = (−1,0)

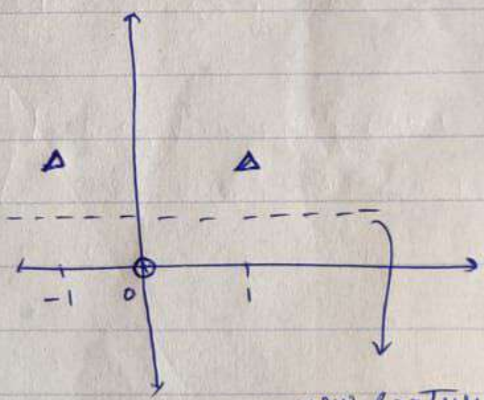
Hence Datapoints are <u>not linearly separable</u>.

(b) $\phi(x) = [u_1(x) \;\; u_2(x)]^T$

let $u_1(x) = x$ &

$u_2(x) = x^2$

$$\boxed{\phi(x) = [x \;\; x^2]}$$



new feature space (2-D) is linearly separable

Q3) ① given $w_1 u_1(x) + w_2 u_2(x) + b = 0$

$u_1(x) = x$ & $u_2(x) = x^2$

$\rightarrow$

$w_1 x + w_2 x^2 + b = 0$.

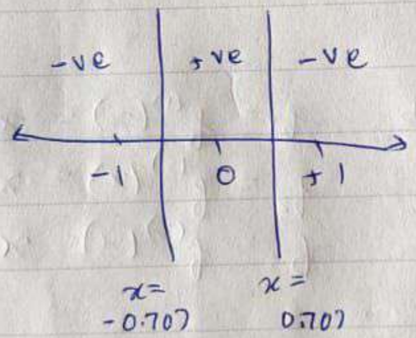
decision boundary $u_2(x) = 0.5$.

$w_1 = 0,\ w_2 = 1 \rightarrow b = -0.5$.

$W = [0\ 1]$ & $b = -0.5$

$u_2(x) = x^2 = 0.5 =$

$x = \underline{\pm 0.707}$ for 1-Dimensional $= \pm\sqrt{0.5}$



so, $\begin{cases} [-\sqrt{0.5}, \sqrt{0.5}, & +\text{ve class} \\ \text{otherwise}, & -\text{ve class} \end{cases}$

Class assigned to $x = 1/3$ will be <u>positive</u>.

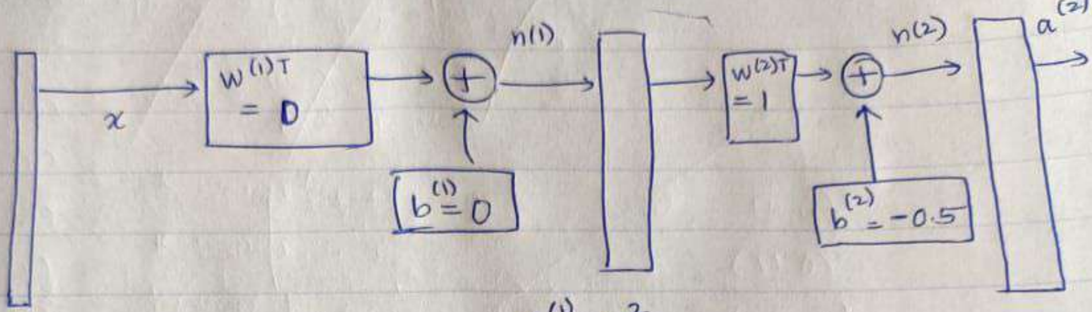
as. $1/3 \in [-\sqrt{0.5}, \sqrt{0.5}]$.

7

consider sign
if −ve, consider −ve
if +ve, consider +ve

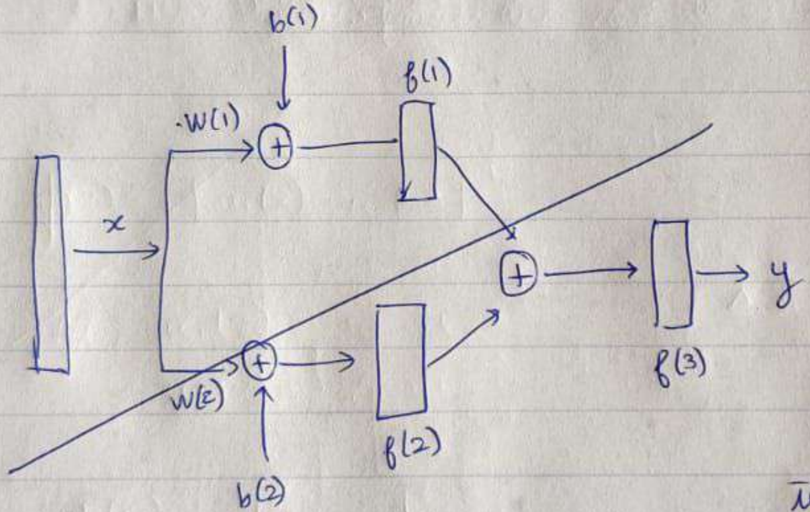
$f^{(2)} = \text{sign}(n)$

~~Q4)~~ -
Q3) d)



$f^{(1)} = n^2$

Q3(d)



w(1) = 0
w(2) = 1
b(1) = 0
b(2) = −0.5
f(1) = n^2
f(2) = sign(n)

This means

$$\text{if } n = \begin{cases} -ve, & \text{if } n \text{ is neg} \\ +ve, & \text{if } n \text{ is pos} \end{cases}$$

Q4) (a) It seems K-means was applied on $\boxed{A2}$. This is because in A1, some points in outer circle would have been closer to the opposite centroides. This does not seem to be the case here.

(b) Dataset B.

K-means on $\boxed{B2}$.

This is because on B1, points in extreme right of red class are are much closer to remaining centroids & would have been marked with them in K-means.

(c) Dataset C

K-means result is $\boxed{\text{not possible to determine}}$.

This is because C_1 & C_2 both seem to be good contenders as result for K-means. All centroids have their colored classes equidistant [in a reasonable region]. Different initial points can result in both C_1 & C_2.

(d) Dataset D.

Extreme end red points of D2 are closer to other points (centroids) than own. Thus K-means on $\boxed{D1}$

(e) Dataset E.

K-means on $\boxed{E2}$.

There is -ve slope between ~~light & dark blue~~ green blue light

~~blue~~ green centroid is higher than ~~dark~~ blue thus division line seems +ve sloped

(f) Dataset F

K-means on $\boxed{F2}$

Top blue points closer to red centroid & centroids are far from data points.

Summary of Q4.

a) K-means on A2
b) K-means on B2
c) cannot be determined [K-means can be both on C1 & C2]
d) K-means on D1
e) K-means on E2
f) K-means on F2.

Q5

(b) If we split tree on node X, ~~rather than on Z~~, → root of decision tree
subsequent splits on Y & Z.

for $X = 0$,

Y	C_1	C_2
0	5	55
1	55	5

$error_Y = \frac{10}{120}$

Z	C_1	C_2
0	15	45
1	45	15

$error_Z = \frac{30}{120}$.

for $X = 1$

Y	C_1	C_2
0	35	5
1	5	35

$error_Y = \frac{10}{80}$

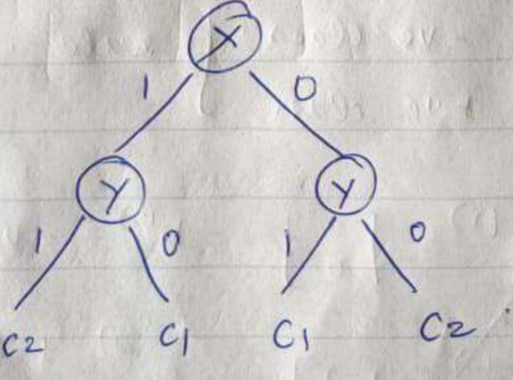
Z	C_1	C_2
0	15	25
1	25	15

$error_Z = \frac{30}{80}$.

~~Prob. of Y is l~~

$error_Y < error_Z$ so Y better split.

This is because Y has less error rate



Overall error rate of tree

$= \frac{10+10}{200} = 0.1$.