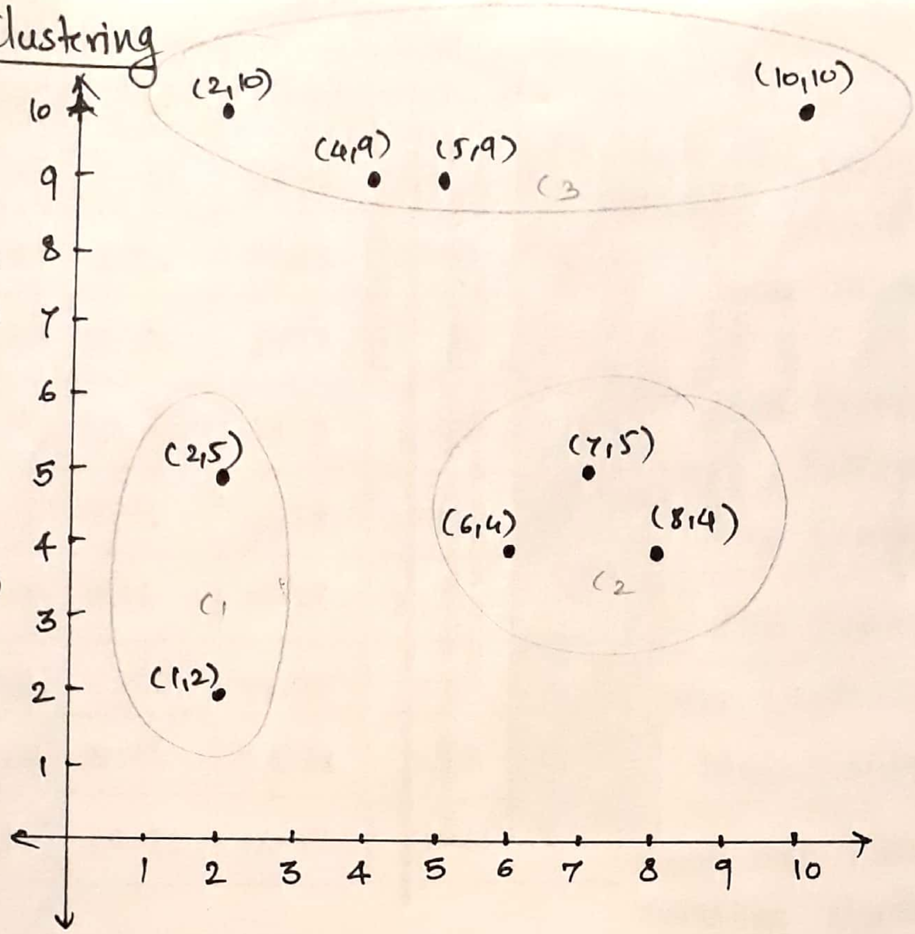
Part 1. Clustering

1 (2,10)
2 (2,5)
3 (8,4)
4 (5,9)
5 (7,5)
6 (6,4)
7 (1,2)
8 (4,9)
9 (10,10)



ii) Begining point. (2,5) (5,8) (4,9) as initial cluster centers

	Points	C1 (2,5)	C2 (5,8)	C3 (4,9)	
1	(2,10)	25	13	5	C3
2	(2,5)	0	18	20	C1
3	(8,4)	37	25	41	C2
4	(5,9)	25	1	1	C3
5	(7,5)	25	13	25	C2
6	(6,4)	17	17	21	C1
7	(1,2)	10	52	58	C1
8	(4,9)	20	2	0	C3
9	(10,10)	89	29	37	C2

(Euclidean distance)^2 is used, between cluster center and point to min distance is cluster belonging

$\boxed{C1}$ Old Centeroid: (2,5)
Points: (2,5), (6,4) (1,2)
New Centeroid: (3, 3.6)

$\boxed{C2}$ Old Centeroid (5,8).
Points: (8,4) (7,5) (10,10)
New Centeroid: (8.3, 6.3)

$\boxed{C3}$: Old Centeroid: (4,9)
Points: (4,9) (2,10) (5,9)
New Centeroid: (3.6, 9.3)

iii)

	Points	C1 (3,3.6)	C2 (8.3,6.3)	C3 (3.6,9.3)	
1	(2,10)	41.96	53.38	3.05	C3
2	(2,5)	2.96	41.38	21.05	C1
3	(8,4)	25.16	5.38	47.45	C2
4	(5,9)	33.16	18.18	2.05	C3
5	(7,5)	17.96	3.38	30.05	C2
6	(6,4)	9.16	10.58	33.85	C1
7	(1,2)	6.56	71.78	60.05	C1
8	(4,9)	30.16	25.78	0.25	C3
9	(10,10)	89.96	16.58	41.45	C2

$\boxed{C1}$ Old Centeroid : (3,3.6)

Points: (2,5) (6,4) (1,2)

New Centeroid: (3,3.6)

$\boxed{C2}$ Old centeroid (8.3,6.3)

Points: (8,4) (7,5) (10,10)

New Centeroid: (8.3,6.3)

$\boxed{C3}$ Old Centeroid: (3.6,9.3)

Points: (2,10) (5,9) (4,9)

New Centeroid: (3.6, 9.3)

Above are resulting centers and final resulting clusters

Question B

	P1	P2	P3	P4	P5
P1	1				
P2	0.10	1			
P3	0.41	0.64	1		
P4	0.55	0.47	0.44	1	
P5	0.35	0.98	0.85	0.76	1

→ (P2, P5)

Single (max) (~~minimum~~)

	P1	P3	P4	(P2, P5)
P1	1			
P3	0.41	1		
P4	0.55	0.44	1	
(P2, P5)	0.35	0.85	0.76	1

((P2, P5), P3)

	P1	P4	((P2, P5), P3)
P1	1		
P4	0.55	1	
((P2, P5), P3)	0.41	0.76	1

(((P2, P5), P3), P4)

	P1	(((P2, P5), P3), P4)
P1	1	
(((P2, P5), P3), P4)	0.41	1

Complete (minimum)

	P1	P3	P4	(P2, P5)
P1	1			
P3	0.41	1		
P4	0.55	0.44	1	
(P2, P5)	0.35	0.64	0.44	1

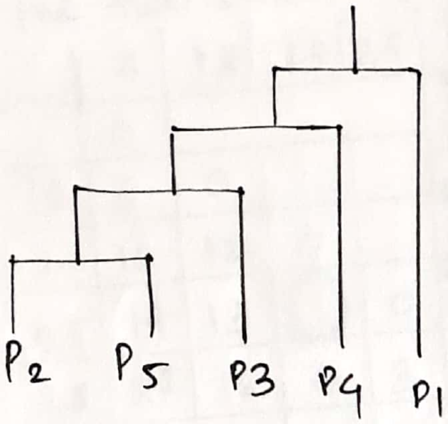
((P2, P5), P3)

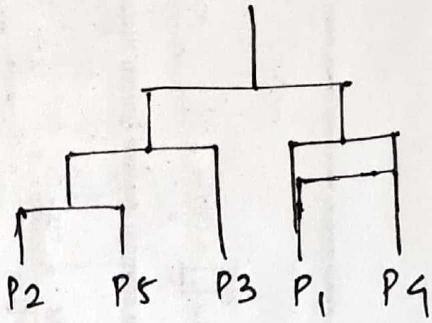
	P1	P4	((P2, P5), P3)
P1	1		
P4	0.55	1	
((P2, P5), P3)	0.35	0.44	1

((P2, P5), P3) (P1, P4)

	(P1, P4)	((P2, P5), P3)
(P1, P4)	1	
((P2, P5), P3)	0.35	1

1)





Questanc

one-dimensional: {6, 12, 18, 24, 25, 28, 30, 42, 48}

1) Center (5, 7.5)

C1: 5	C2: 7.5
6	- 12
	- 18
	- 24
	- 25
	- 28
	- 30
	- 42
	- 48

2) Center (15, 25)

C1: 15	C2: 25
- 6	- 24
- 12	- 25
- 18	- 28
	- 30
	- 42
	- 48

1) $WSS = (6-5)^2 + (12-7.5)^2 + (18-7.5)^2 + (24-7.5)^2 + (25-7.5)^2 + \dots =$
$= 4467$

$BSS = 14.0625$

$\boxed{TSS = 4481.0625}$

1) $WSS = (6-15)^2 + (12-15)^2 + \dots$
$= 943.$

$BSS = 3*(20-15)^2 + 6*(20-25)^2$
$= 225.$

$TSS = 943 + 225 = \boxed{1168.}$

b) Centeroids

C1 (5, 7.5) ⇒ kmeans.

C1: 5 (6)

C2: 7.5 (12, 18, 24, 25, 28, 30, 42, 48)

C1: 6 (6, 12)

C2: 28.375 (18, 24, 25, 28, 30, 42, 48)

C1: 9 (6, 12, 18)

C2: 30.71 (24, 25, 28, 30, 42, 48)

In this way centeroids of the both clusters keep moving until both the clusters are stable.

c) Single link clustering

The two clusters produced are [((28,30),25),((6,12),(18,24))]

	6	12	24	25	28	30	42	48	18
6	0								
12	6	0							
24	18	12	0						
25	19	13	1	0					
28	22	16	4	3	0				
30	24	18	6	5	2	0			
42	36	30	18	17	14	12	0		
48	42	36	24	23	20	18	6	0	
18	12	6	6	7	10	12	24	30	0

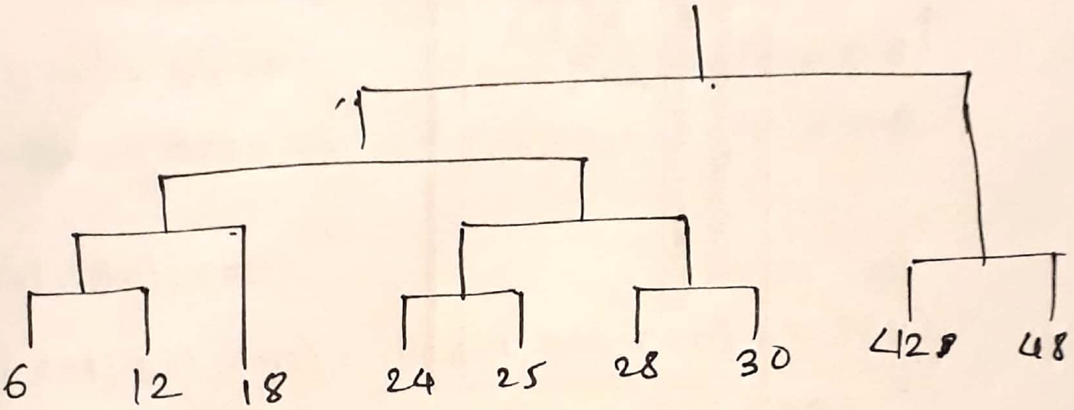
	6	12	18	28	30	42	48	(24,25)
6	0							
12	6	0						
18	12	6	0					
28 ~~(24,25)~~	22	16	10	0				
30	24	18	12	2	0			
42	36	30	24	14	12	0		
48	42	36	30	20	18	6	0	
(24,25)	18	13	6	3	5	17	23	0

	6	12	18	42	48	(24,25)	(28,30)
6	0						
12	6	0					
18	12	6	0				
42	36	30	24	0			
48	42	36	30	6	0		
(24,25)	18	13	6	17	23	0	
(28,30)	22	16	10	12	18	3	0

	6	12	18	42	48	((24,25),(28,30))
6	0					
12	6	0				
18	12	6	0			
42	36	30	24	0		
48	42	36	30	6	0	
((24,25),(28,30))	18	13	6	12	18	0

	((6,12),18)	(42,48)	((24,25)(28,30))
((6,12),18)	0		
(42,48)	24	0	
((24,25),(28,30))	6	12	0

	(42,48)	x
(42,48)	0	
[((6,12),18), ((24,25),(28,30))]	$\boxed{12}$	0



d) kmeans produced more natural clustering in this case

e) k means tryies to stabalize the clusters by constantly shifting centeroid and changing clusters to reduce erros and become Stable.

Part II

Question D

1. Naive Baye's

(a) (1)

$P(A=1|-) = 2/5 = 0.4$ $\quad$ $P(A=0|-) = 3/5 = 0.6$ $\quad$ $P(+) = \frac{5}{10} = 0.5$

$P(B=1|+) = 2/5 = 0.4$ $\quad$ $P(B=0|-) = 3/5 = 0.6$ $\quad$ $P(-) = \frac{5}{10} = 0.5$

$P(C=1|-) = 1$ $\quad$ $P(C=0|-) = 0$

$P(A=1|+) = 3/5 = 0.6$ $\quad$ $P(A=0|+) = 2/5 = 0.4$

$P(B=1|+) = 1/5 = 0.2$ $\quad$ $P(B=0|+) = 4/5 = 0.8$

$P(C=1|+) = 4/5 = 0.8$ $\quad$ $P(C=0|+) = 1/5 = 0.2$

(b) $P(A=1, B=1, C=0)$

(2)

$$P(+|A=1,B=1,C=0) = \frac{P(A=1,B=1,C=0|+) \cdot P(+)}{P(A=1,B=1,C=0)}$$

$$= \frac{P(A=1|+) \cdot P(B=1|+) \cdot P(C=0|+) \cdot P(+)}{P(A=1,B=1,C=0)}$$

$$= \frac{0.6 * 0.2 * 0.2 * 0.5}{P(A=1,B=1,C=0)} = 0.012 / P(A=1,B=1,C=0)$$

$$P(-|A=1,B=1,C=0) = \frac{P(A=1,B=1,C=0|-) \cdot P(-)}{P(A=1,B=1,C=0)}$$

$$= \frac{P(A=1|-) \cdot P(B=1|-) \cdot P(C=0|-) \cdot P(-)}{P(A=1,B=1,C=0)}$$

$$= 0.4 * 0.4 * 0 = 0.$$

We label ~~the~~ it with +ve

(c) $P(A=1|+) = (3+2)/(5+4) = 5/9 = 0.55$

(3) $P(A=1|-) = (2+2)/(5+4) = 4/9 = 0.44$ $\qquad P(A|B) = \dfrac{n_c + mp}{n+c}$

$P(B=1|+) = (\cancel{1}+2)/(5+4) = 3/9 = 0.33$ $\qquad n_c$ = number of times $A \wedge B$ happened

$P(B=1|\cancel{+}) = (2\cancel{+}2)/(5+4) = 4/9 = 0.44$

$P(C=0|+) = (1+2)/(5+4) = 3/9 = 0.33$ $\qquad n$ = number of times B happened

$P(C=0|-) = (0+2)/(5+4) = 2/9 = 0.22$ $\qquad m=4 \quad p=1/2$

(d) $P(A=1, B=1, C=0) = \cancel{K}.$

(4) $$P(+|A=1, B=1, C=0) = \frac{P(A=1, B=1, C=0|+) \cdot P(+)}{P(A=1, B=1, C=0)}$$

$$= \frac{P(A=1|+) \cdot P(B=1|+) \cdot P(C=0|+) \cdot P(+)}{P(A=1, B=1, C=0)}$$

$$= \frac{0.55 * 0.33 * 0.33 * 0.5}{P(A=1, B=1, C=0)} = 0.03/K$$

$$P(-|A=1, B=1, C=0) = \frac{P(A=1, B=1, C=0|-) \cdot P(-)}{P(A=1, B=1, C=0)}$$

$$= \frac{P(A=1|-) \cdot P(B=1|-) \cdot P(C=0|-) \cdot P(-)}{P(A=1, B=1, C=0)}$$

$$= \frac{0.44 * 0.44 * 0.22 * 0.5}{P(A=1, B=1, C=0)} = 0.042592 * 0.5 = 0.021296/K$$

This belongs to class $\boxed{+ve}$

(e) When one of the conditional probabilities is zero, the estimate for
(5) conditional probabilities using m estimate probabilities approach is better, since we dont want the entire expression to become zero.

(ii) AdaBoosting

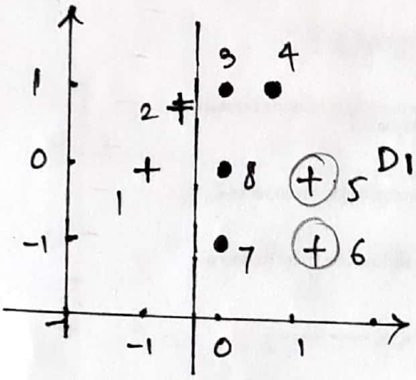
$\epsilon = \underline{(h(x_i) \neq y_i)}$

$\alpha_t = \frac{1}{2} \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$

$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \begin{cases} e^{-\alpha t} & \text{if } h(x_i) = y_i \\ e^{\alpha t} & \text{if } h(x_i) \neq y_i \end{cases}$

$Z_t = 2\sqrt{\epsilon_t(1-\epsilon_t)}$

final classifier $H = \operatorname{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$



	1	2	3	4	5	6	7	8	ϵ_1	α_1	Z_1
D1	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.25	0.55	0.866

$\epsilon_1 = \frac{2}{8} = 0.25 \quad \alpha_1 = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) = 0.55$

$D_2 = \frac{D_1}{Z_1} \begin{cases} e^{-\alpha_1} & \text{if } h(x_i) = y_i \\ e^{\alpha_1} & \text{if } h(x_i) \neq y_i \end{cases}$

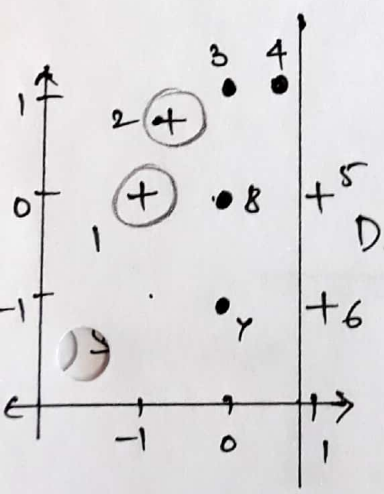
$Z_1 = 2\sqrt{\epsilon_1(1-\epsilon_1)} = 0.866 =$

$D_2 = \frac{D_1}{Z_1} e^{+\alpha} = \frac{0.125}{0.866} e^{0.55} = 0.25$

(invalid classification) (~~1, 2, 3, 4, 7, 8~~) (5, 6)

$D_2' = \frac{D_1}{Z_1} e^{-\alpha} = \frac{0.125}{0.866} e^{-0.55} = 0.083$

(valid classification (5, 2, 3, 4, 7, 8)



	1	2	3	4	5	6	7	8	ϵ_2	α_2	Z_2
D2	0.083	0.083	0.083	0.083	0.25	0.25	0.083	0.083	0.1066	~~1.066~~ 0.807	~~0.6156~~ 0.744

misclassified points are 1, 2.

$\epsilon_2 = \frac{0.083 + 0.083}{1} = \cancel{0.1666}\ 0.1666 \quad \alpha_2 = \frac{1}{2}\ln\left(\frac{1-\epsilon_2}{\epsilon_2}\right) = \cancel{1.0066} = 0.807$

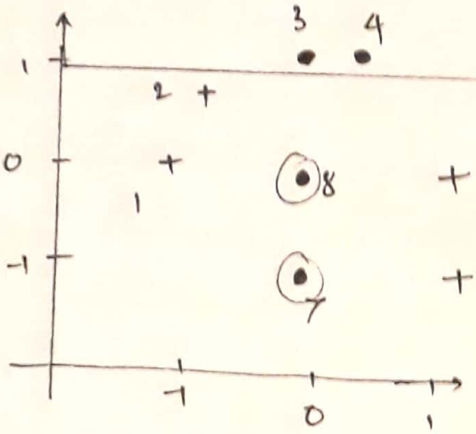
$Z_2 = 2\sqrt{\epsilon_2(1-\epsilon_2)} = \cancel{0.6156}\ 0.744$

Scanned by CamScanner

$$D_3 = \frac{D_2}{Z_2} e^{-\alpha_2} \qquad D_{3,4,7,8} = \frac{0.083}{0.744} \times e^{-0.166} = 0.049$$

$$D_{5,6} = \frac{0.25}{0.744} \times e^{-0.166} = 0.1498$$

$$D_{1,2} = \frac{0.083}{0.744} \times e^{0.166} = 0.25$$

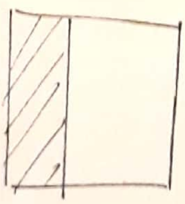
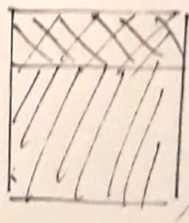
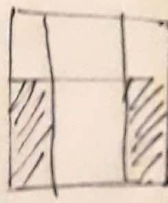


1	2	3	4	5	6	7	8	ϵ	α	Z
0.25	0.25	0.049	0.049	0.1498	0.1498	0.049	0.049	099	1.10	0.5

$$\epsilon_3 = \frac{0.049 + 0.049}{0.994} = 0.099$$

$$\alpha_3 = \frac{1}{2} \ln\left(\frac{1-\epsilon_3}{\epsilon_3}\right) = 1.1019$$

$$Z_3 = 2\sqrt{\epsilon_3(1-\epsilon_3)} = 0.598$$

H = Sign (0.55  + 0.807  + 1.1019 ) = 

ⓑ Training Error :

$$P_{x_i \sim D}[H(x_i) \neq Y] : \prod_{t}^{T} \sqrt{1 - 4\gamma_t^2}$$

$$T.E = \sqrt{1-4(0.25)^2} * \sqrt{1-4(0.334)^2} * \sqrt{1-4(0.426)^2}$$

$$= 0.866 * 0.7512 * 0.5235$$

$$TE = 0.34$$

$$\gamma_t = \frac{1}{2} - \epsilon_t$$

$$\gamma_1 = \frac{1}{2} - 0.25 = 0.25$$

$$\gamma_2 = \frac{1}{2} - \cancel{0.807}\ 0.166 = 0.334$$

$$\gamma_3 = \frac{1}{2} - \cancel{1.1019}\ 0.099 = 0.401.$$

* AdaBoost uses multiple decision stumps instead of single decision stumps and giving each decision stump different weight according to output is better that classifying just by single decision stump
* In AdaBoosting misclassified points are focused more an given more weights

Q5. Decision Tree.

X1	X2	X3	Y
a	c	k	-1
a	w	K	-1
b	w	v	+1
a	c	v	+1
b	w	K	-1
a	c	s	+1
b	w	s	+1
a	u	v	-1
b	c	v	-1
b	c	s	+1
b	g	v	-1

Information Gain Heuristics Using Entropy.

$$E_x[I(x)] = -\sum_{x \in X} p(x) \cdot \log p(x)$$

$$E(\text{Target}) = -\left[P(+1) \cdot \log p(+1) + P(-1) \cdot \log P(-1)\right]$$

$$E(\text{Target}) = -\left[\left(\tfrac{5}{11}\right) \log_2\left(\tfrac{5}{11}\right) + \left(\tfrac{6}{11}\right) \cdot \log_2\left(\tfrac{6}{11}\right)\right]$$

$$E(\text{Target}) = 0.994$$

① Information Gain for X1:

$$IG(X_1) = E(\text{Target}) - E(\text{Target}, X_1)$$

$$E(\text{Target}, X_1) = P(a) \cdot E(a) + P(b) \cdot E(b).$$

$$E(a) = -\left[P(+ve) \log P(+ve) + P(-ve) \cdot \log P(-ve)\right]$$

$$= -\left[(3/5) \cdot \log_2 (3/5) + (2/5) \log_2 (2/5)\right]$$

$$= 0.97$$

$$E(b) = -\left[P(+ve) \log p(+ve) + P(-ve) \cdot \log P(-ve)\right]$$

$$= -\left[(3/6) \log_2 (3/6) + (3/6) \cdot \log_2 (3/6)\right]$$

$$= 1$$

$$E(\text{Target}, X_1) = P(a) \cdot E(a) + P(b) \cdot E(b)$$

$$= \frac{5}{11} * 0.97 + \frac{6}{11} * 1$$

$$= 0.986$$

$$IG(X_1) = E(\text{Target}) - E(\text{Target}, X_1)$$

$$= 0.994 - 0.986 = 8.5 \times 10^{-3} = 0.85 \times 10^{-2}$$

Scanned by CamScanner

② Information Gain for x_2

$IG(x_2) = E(\text{Target}) - E(\text{Target}, x_2)$

$E(\text{Target}, x_2) = P(c) \cdot E(c) + P(g) \cdot E(g) + P(u) \cdot E(u) + P(\omega) \cdot E(\omega)$

$E(c) = -[P_{+ve}(c) * \log_2 P_{+ve}(c) + P_{-ve}(c) * \log_2 P_{-ve}(c)]$

$= -[\frac{3}{5} \cdot \log_2 \frac{3}{5} + \frac{2}{5} \log_2 \frac{2}{5}] = 0.961$

$E(g) = -[P_{+ve}(g) \cdot \log_2 P_{+ve}(g) + P_{-}(g) \cdot \log_2 P_{-}(g)]$

$= -[0 + \frac{1}{1} \log_2 \frac{1}{1}] = 0$

$E(u) = -[P_{+}(u) \cdot \log_2 P_{+}(u) + P_{-}(u) \cdot \log_2 P_{-}(u)]$

$= -[0 + \frac{1}{1} \log_2 \frac{1}{1}] = 0$

$E(\omega) = -[P_{+}(\omega) \cdot \log_2 P_{+}(\omega) + P_{-}(\omega) \cdot \log_2 P_{-}(\omega)]$

$= -[\frac{2}{4} \log_2 \frac{2}{4} + \frac{2}{4} \log_2 \frac{2}{4}] = 1$

$E(\text{Target}, x_2) = \frac{5}{11} \cdot (0.961) + 0 + 0 + \frac{4}{11} \cdot (1). = 0.800 = 80 \times 10^{-2}$

$IG(x_2) = 0.994 - 0.800 = 0.193 = 19.3 \times 10^{-2}$

③ Information Gain for x_3

$IG(x_3) = E(\text{Target}) - E(\text{Target}, x_3).$

$E(\text{Target}, x_3) = P(K) \cdot E(K) + P(S) \cdot E(S) + P(V) \cdot E(V).$

$E(K) = -[P_{+}(K) \cdot \log_2 P_{+}(K) + P_{-}(K) \cdot \log_2 P_{-}(K)] = -[\frac{0}{3} + \frac{3}{3} \log_2 \frac{3}{3}] = 0$

$E(S) = -[P_{+}(S) \cdot \log_2 P_{+}(S) + P_{-}(S) \cdot \log_2 P_{-}(S)] = -[\frac{3}{3} \log_2 \frac{3}{3} + 0] = 0$

$E(V) = -[P_{+}(V) \cdot \log_2 (V) + P_{-}(V) \cdot \log_2 P_{-}(V)] = -[\frac{2}{5} \log_2 \frac{2}{5} + \frac{3}{5} \log_2 \frac{3}{5}] = 0.961$

$E(\text{Target}, x_3) = 0 + 0 + 0.961 * \frac{5}{11} = 0.436$ *

$IG(x_3) = E(\text{Target}) - E(\text{Target}, x_3) = 0.994 - 0.436 = 0.557 = 55.7 \times 10^{-2}$

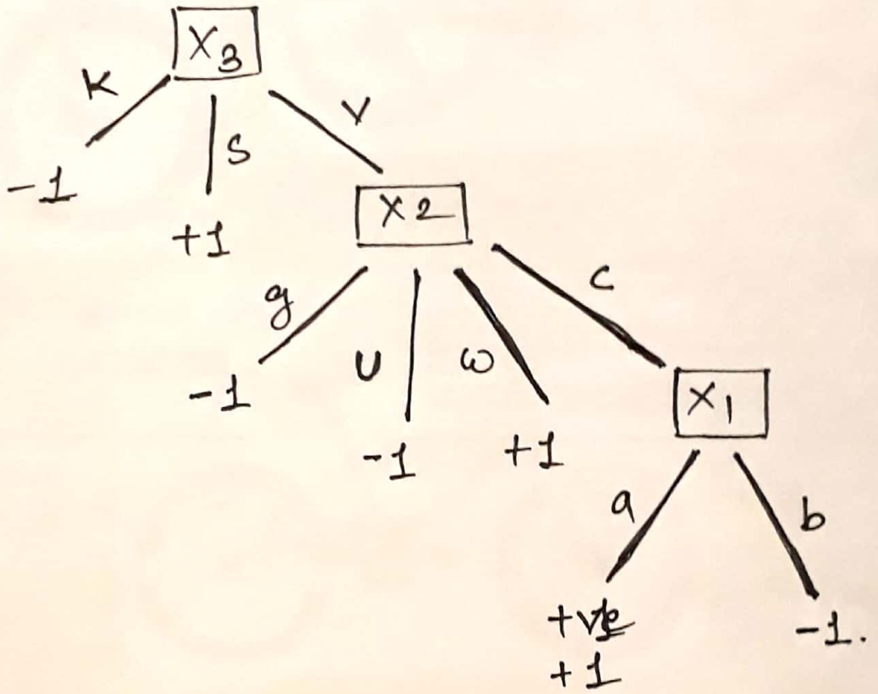
Scanned by CamScanner

$IG(X_1) = 0.85 \times 10^{-2}$

$IG(X_2) = 19.3 \times 10^{-2}$

$IG(X_3) = 55.7 \times 10^{-2}$

X_3 is highest Information Gain. Hence, it will be the head of Decision tree.

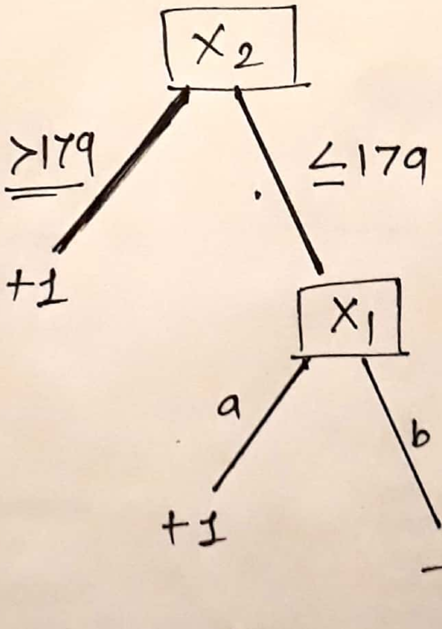


(b) $IG(X_1) = 0.2$

(i) $IG(X_2) = 0.46$

$IG(X_3) = 0.2$

$IG(X_4) = 0$



(ii)

X_1	X_2	X_3	X_4	Y	Pred
b	170	f	d	-1	-1
a	150	f	d	+1	+1
b	60	f	d	+1	-1

Accuracy $= \frac{2}{3} \times 100 =$ ~~0.75~~ 66%

Accuracy is $\boxed{66\%}$