Name : Bilal Malik

Subject : ML

Roll No : 34-BSCS-17

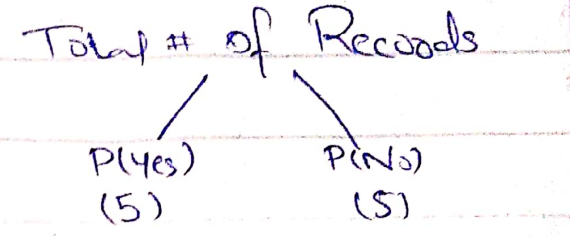
Submitted To : Prof. Ayesha Atta

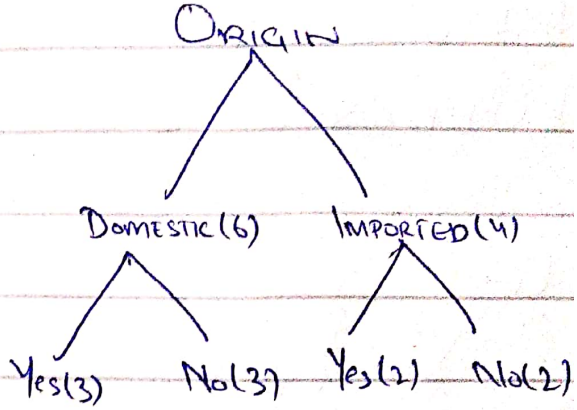
Question No. 1

Solution:

P(Yes) = 1/2

P(No) = 1/2



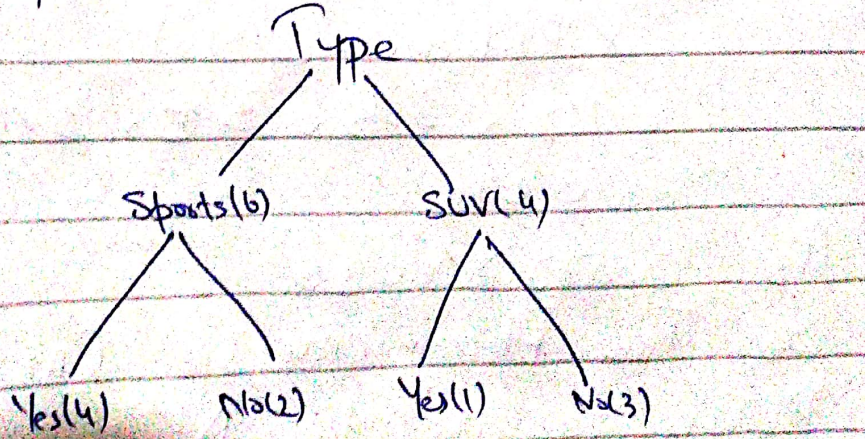


P(domestic/Yes) = 3/5

P(domestic /No) = 3/5

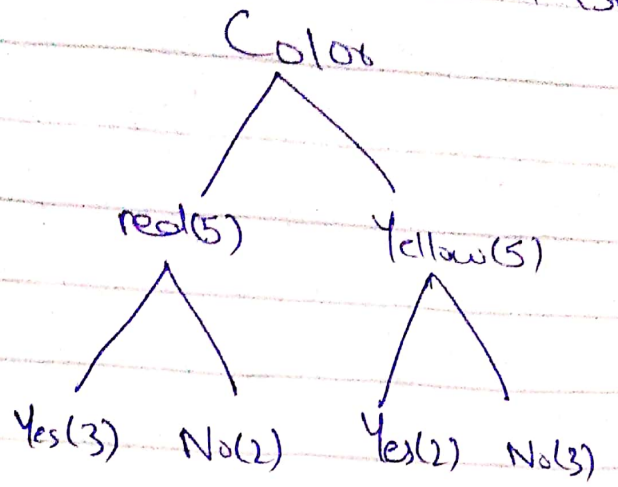
P(Imported /Yes) = 2/5

P(Imported /No) = 2/5



P(Sports/Yes) = 4/5
P(Sports/No) = 2/5

P(SUV/Yes) = 1/5
P(SUV/No) = 3/5



P(RED/Yes) = 3/5
P(RED/No) = 2/5

P(YELLOW/YES) = 2/5
P(YELLOW/No) = 3/5

LET
X = {RED, SUV, DOMESTIC}

P(X/YES) = P(YES) × P(RED/YES) × P(SUV/YES)
× P(DOMESTIC/YES)
= 1/2 × 3/5 × 1/5 × 3/5
= 0.036

P(X/No) = P(No) × P(RED/No) × P(SUV/No)
× P(DOMESTIC/No)
= 1/2 × 2/5 × 3/5 × 3/5
= 0.072

As 0.072 > 0.036
So THE PROBABILITY OF CAR
RAN(L STOLEN IS No.

Question No# 2.

Decision Tree

P(Yes) = 9 P(No) 5

Entropy(S) = $\frac{-9}{9+5}\log_2\left[\frac{9}{9+5}\right] = \frac{5}{14}\log_2$

$\frac{5}{9+5}$ = .940

Calculating entropy for each attribute

Outlook	yes	No	Entropy
Sunny	2	3	.971
Rain	3	2	.971
Overcast	4	0	0

Using formula

$E(\text{Outlook} = \text{Sunny}) = \frac{-3}{5}\log\left(\frac{2}{5}\right) = \frac{3}{5}\log\left(\frac{3}{5}\right) = .971$

Similarly Calculate Others

Information

Calculating Average Entropy

$I(\text{Outlook}) = \frac{3+2}{9+5} \times .971 + \frac{2+3}{9+5} \times .971 + \frac{4}{14}$

= .693

Now calculating gain

Gain = Entropy(S) − I(Attribute)

Entropy(S) = .940

Gain(Outlook) = .940 − .693 = .247

For Temperature

Temp	Yes	No	Entropy
Hot	2	2	1
Mild	4	2	.918
Cool	3	1	.918

Avg Information Entropy :

$$I(\text{Temp}) = \frac{2+2}{9+5} \times 1 + \frac{4+2}{9+5} \times .918 + \frac{3+1}{9+5} \times .811$$

$$= .911$$

Entropy(S) = .940

Gain = .940 − .911 = .029

For Humidity

Humidity	Yes	No	Entropy
High	3	4	.985
Normal	6	1	.591

Avg Information Entropy

I(HUMIDITY) = 3+4/9+5 × .985 + 6+1/9+5 × .591

ENTROPY(S) = .940

GAIN(HUMIDITY) = .940 - .788 = .152

FOR WIND

WIND	YES	NO	ENTROPY
STRONG	3	3	1
WEAK	6	2	.811

AVERAGE INFO ENTROPY

I(WINDY) = 3+3/9+5 × 1 + 6+2/9+5 × .811 = .8912

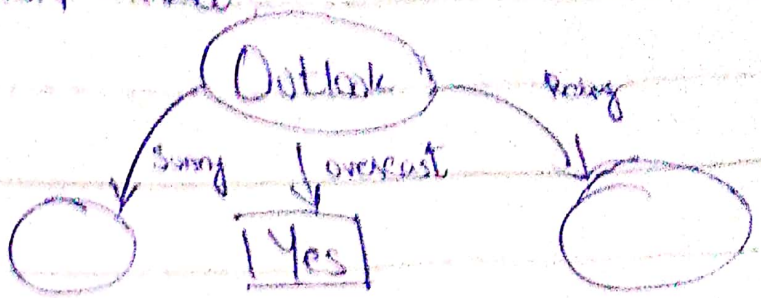
E(S) = .940

GAIN(WIND) = .940 - .892 = .048

SINCE outlook has biggest gain. It will be our root node

Attributes	Gain
Outlook	.247
Temp	.029
Humidity	.152
WIND	.048

Decision Tree:



Here further splitting is needed

Calculate entropy for Sunny & Rainy or Overcast

For these 2 values.

P(Yes) = 2 , P(No) = 3

Total = 5

$$\text{Entropy}(S) = -\frac{2}{2+3}\log_2\frac{2}{2+3} - \frac{3}{2+3}\log_2\frac{3}{2+3}$$

E(Sunny) = .971

For Humidity.

Humidity	Yes	No	Entropy
High	0	3	0
Normal	2	0	0

Avg Info Entropy

I(Humidity) = 0

Gain = .971

For Wind

Wind	Yes	No	Entropy
Strong	1	1	1
Weak	1	2	.98

I(Windy) = .951 Gain = .020

For Temperature

Temp	Yes	No	Entropy
Cool	1	0	0
Hot	0	2	0
Mild	1	1	1

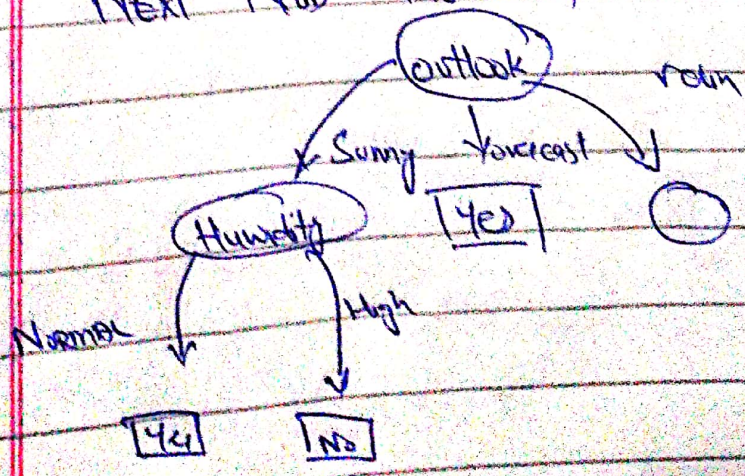
I(Temp) = .4

Gain = .571

Now

Attributes	Gain
Temperature	.571
Humidity	.971
Windy	.02

Next Node Humidity



Now for rain, further splitting is necessary

P=3, N=2

Total=5

Entropy(S)$_{rain}$ $= -\frac{3}{3+2}\log_2\frac{3}{3+2} - \frac{2}{3+2}\log_2\frac{2}{2+3}$

$= .971$

Calculating Entropy For Each Attribute

For Humidity

Avg Info Entropy.

I(Humidity) = .951

Gain = .020

For Wind

Avg Info Entropy

I(Wind) = 0

Gain = .971

For Temp,

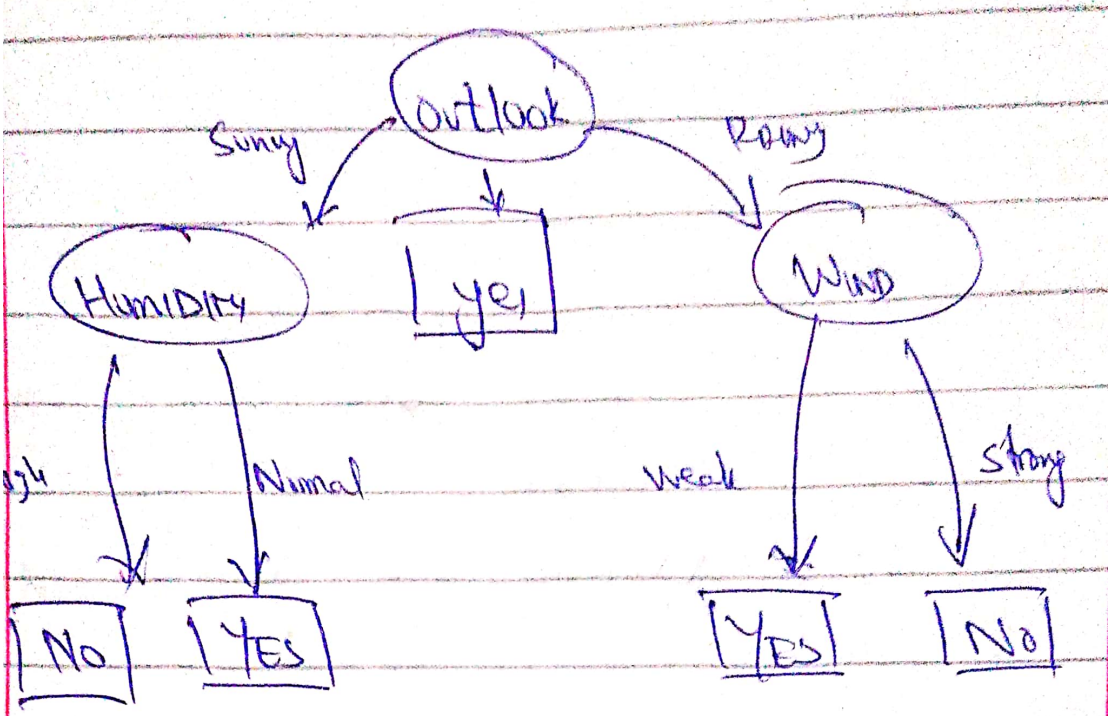
I(Temp) = .951

Gain = .020

Attributes	Gain
Humidity	.02
Wind	.971
Temperature	.02

Next Node Wind

FINAL DECISION TREE:



QUESTION: 3 (a)

K-MEAN ALGORITHM:

(2, 3, 4, 10, 11, 12, 20, 25 & 30) K=2

$C_1 = 4$ & $C_2 = 12$

By finding the distance of dataset

$K_1 = \{2, 3, 4\}$

$K_2 = \{10, 11, 12, 20, 25, 30\}$

$C_1 = 2+3+4/3 = 5$

$C_2 = 10+11+12+20+25+30/6 = 18$

Now for Iteration #2

$C_1 = 3$ & $C_2 = 18$

Now As for dataset

$K_1 = \{2, 3, 4, 10\}$

$K_2 = \{11, 12, 20, 25, 30\}$

$C_1 = 2+3+4+10/4 = 4.75 \approx 5$

$C_2 = 11+12+20+25+30/5 = 19.6 \approx 20$

For ITERATION #3

$C_1 = 5$ & $C_2 = 20$

$K_1 = \{2, 3, 4, 10, 11, 12\}$

$K_2 = \{20, 25, 30\}$

$C_1 = 2+3+4+10+11+12/6 = 7$

$20+25+30/3 = 25$

For ITERATION # 4

$c_1 = 7$ & $c_2 = 25$

$K_1 = \{2, 3, 4, 10, 11, 12\}$

$K_2 = \{20, 25, 30\}$

As K_1 & K_2 of this iteration are same as iteration #4

$c_1 = 7$ & $c_2 = 25$

So the cluster is

$K_1 = \{2, 3, 4, 10, 11, 12\}$

$K_2 = \{20, 25, 30\}$

Q No. 3 (B)

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

Sol:- Take any 2 points

① c_1 (2,3) c_2 (5,6)

② Assign values in 2 clusters

distance formula = $\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$

x	y	distance c_1(2,3)	distance c_2(5,6)	Select cluster
2	3	0	4.24	c_1
5	6	4.24	0	c_2
8	7	7.21	3.16	c_2
1	4	1.41	4.47	c_1
2	2	1	5	c_1
6	7	5.65	1.41	c_2
3	4	1.41	2.82	c_1
8	6	6.70	3	c_2

③ Check which one in c_1 & c_2.

c_1 = (2,3) (1,4) (2,2) (3,4)

Mean (x,y) = $(x_1+x_2+x_3+x_4)/4$, $(y_1+y_2+y_3+y_3)/4$

= (2+1+2+3)/4, (3+4+2+4)/4

= (2, 3.25)

$C_2 = (5,6)(8,7)(6,7)(8,6)$

Mean $(x,y) = \frac{5+8+6+8}{4}, \frac{6+7+7+6}{4}$

$c_2 = (6.75, 6.5)$

Mean is different, repeating steps

X	Y	distance $c_1(2,3.25)$	distance $c_2(6.75,6.5)$	select Cluster
2	3	.25	5.9	c_1
5	6	4.069	1.820	c_2
8	7	7.075	1.346	c_2
1	4	1.25	6.269	c_1
2	2	1.25	6.54	c_1
6	7	5.482	.90	c_2
3	4	1.25	4.5	c_1
8	6	6.005	1.346	c_2

$C_1 = (2,3)(1,4)(2,2)(3,4)$

Mean (x,y): $x = \frac{2+1+2+3}{4}$, $y = \frac{3+4+2+4}{4}$

$C_1 = (2, 3.25)$

$C_2 = (5,6)(8,7)(6,7)(8,6)$

Mean (x,y) $x = \frac{5+8+6+8}{4}$, $y = \frac{6+7+7+6}{4}$

$C_2 = (6.75, 6.5)$

Mean is same, so stop it.

Question No 4

	a (B/N)	b (B/N)	c (case)	d (case)	e (case)	f (B/N)	g (B/N)
a	0.00	2.83	4.24	4.12	2.24	6.4	4.12
b	2.83	0	5.10	3.61	3	6.71	6.08
c	4.24	5.10	0	2.24	2.24	2.24	2.24
d	4.12	3.61	2.24	0	2.0	3.16	4.24
e	2.24	3	2.24	2	0	4.24	3.16
f	6.40	6.71	2.24	3.16	4.24	0	4.0
g	4.12	6.08	2.24	4.24	3.16	4	0

So, here we have

Min-points = 3

ε-Neighborhood = 2.25

Core Points: c, d, e

Border Points: a, f, g

Noise Points: b