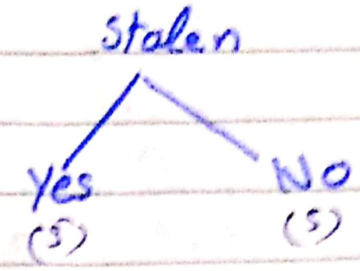
Roll No: 0029-BSCS-17

ML

Syed M. Tafseer ul Hassan

Q#1

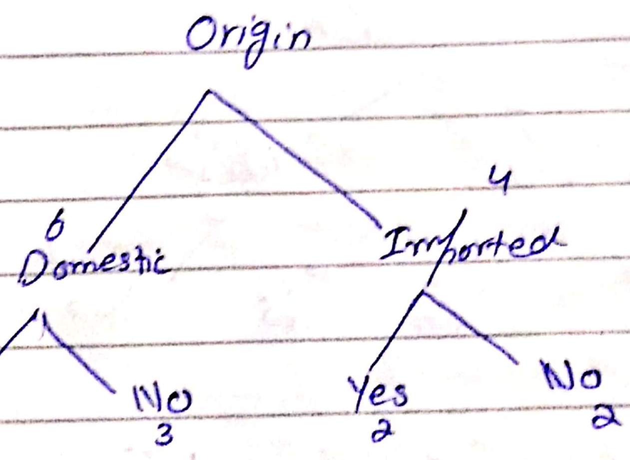
i)



$P(\text{yes}) = \frac{5}{10} = \frac{1}{2}$

$P(\text{No}) = \frac{5}{10} = \frac{1}{2}$

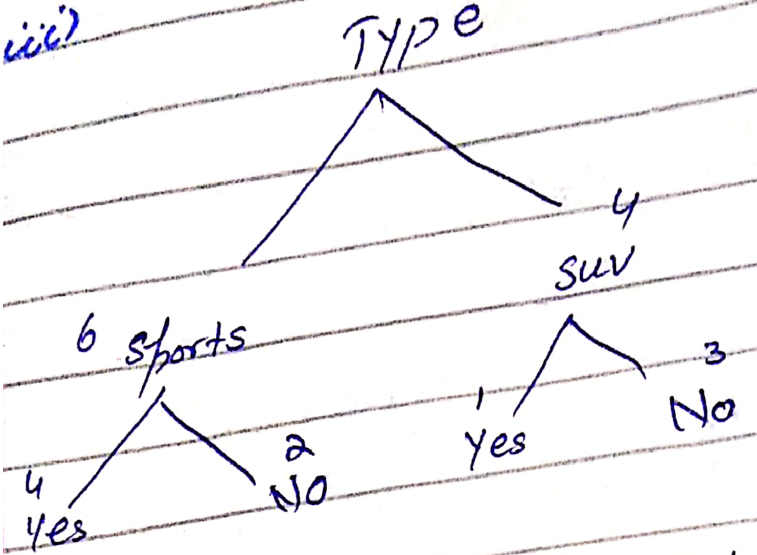
ii)



$P\left(\frac{\text{Domestic}}{\text{Yes}}\right) = \frac{3}{5}$ $P\left(\frac{\text{Imported}}{\text{Yes}}\right) = \frac{2}{5}$

$P\left(\frac{\text{Domestic}}{\text{No}}\right) = \frac{3}{5}$ $P\left(\frac{\text{Imported}}{\text{No}}\right) = \frac{2}{5}$

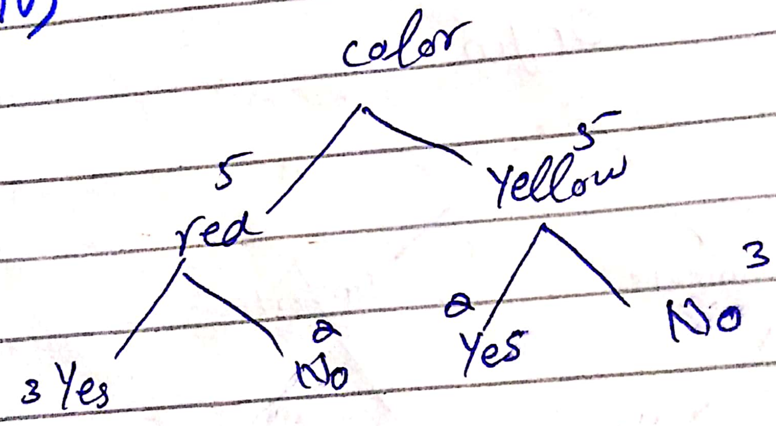
iii)



$P\left(\frac{sports}{Yes}\right) = \frac{4}{5}$ $\qquad$ $P\left(\frac{SUV}{Yes}\right) = \frac{1}{5}$

$P\left(\frac{sports}{No}\right) = \frac{2}{5}$ $\qquad$ $P\left(\frac{SUV}{No}\right) = \frac{3}{5}$

iv)



$P\left(\frac{red}{yes}\right) = \frac{3}{5}$ $\qquad$ $P\left(\frac{Yellow}{yes}\right) = \frac{2}{5}$

$P\left(\frac{red}{No}\right) = \frac{2}{5}$ $\qquad$ $P\left(\frac{Yellow}{No}\right) = \frac{3}{5}$

$$P\left(\frac{x}{Yes}\right) = \{P(Yes), P\left(\frac{red}{Yes}\right), P\left(\frac{suv}{Yes}\right), P\left(\frac{dom}{Y}\right.$$

$$= \frac{1}{2} \times \frac{3}{5} \times \frac{1}{5} \times \frac{3}{5}$$

$$= 0.2 \times 0.6 \times 0.2 \times 0.6$$

$$= 0.036$$

$$P\left(\frac{x}{No}\right) = \{P(No), P\left(\frac{red}{No}\right), P\left(\frac{suv}{No}\right), P\left(\frac{domes}{No}\right)$$

$$= \frac{1}{2} \times \frac{2}{5} \times \frac{3}{5} \times \frac{3}{5}$$

$$= 0.072$$

will Not be stolen.

Q#3

a) Draw clusters using given dataset

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

$K = 2$

$c_1 = 4$ & $c_2 = 12$

By finding the distance of dataset with c_1 & c_2:

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

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

$$c_1 = \frac{(2+3+4)}{3} = 3$$

$$c_2 = \frac{(10+11+12+20+25+30)}{6} = 18$$

Now for iteration #2:

$c_1 = 3$ & $c_2 = 18$

By finding distance of dataset with c_1 & c_2:

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

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

$$c_1 = \frac{(2+3+4+10)}{4} = 4.75 \approx 5$$

$$c_2 = \frac{(11+12+20+25+30)}{5} = 19.6 \approx$$

Scanned with CamScanner

Now for iteration #3

$c_1 = 4.75 \approx 5$ & $c_2 = 19.6 \approx 20$

By finding distance from c_1 & c_2:

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

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

$$c_1 = \frac{2+3+4+10+11+12}{6} = 7$$

$$c_2 = \frac{20+25+30}{3} = 25$$

Now 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 iteration #3, so,

$c_1 = 7$ & $c_2 = 25$

so the clusters of this dataset are:

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

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

Scanned with CamScanner

b)

Draw the clusters for given dataset
(2,3) (5,6) (8,7) (1,4) (2,2)
(6,7) (3,4) (8,6)

where $k = 2$

① Assume two points as mean

$c_1 = (2,3)$ $c_2 = (5,6)$

② Assign values in two clusters

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

x	y	distance $c_1(2,3)$	distance $c_2(5,6)$	selec cluste
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

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

$\text{mean}(x,y) = \dfrac{x_1 + x_2 + x_3 + x_4}{4}, \dfrac{y_1 + y_2 + y_3 + y}{4}$

$$= \frac{2+1+2+3}{4}, \frac{3+4+2+4}{4}$$

$c_1 = (2, 3.25)$

$c_2 = (5,6)(8,7)(6,7)(8,6)$

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

$c_2 = (6.75, 6.8)$

mean is not same repeat the steps. with new men.

x	Y	distance $c_1(2,3.25)$	distance $c_2(6.75,6.5)$	sel
2	3	0.25	5.900	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	0.90	c_2
3	4	1.25	4.50	c_1
8	6	6.005	1.346	c_2

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

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

$c_1 = (2, 3.25)$

$c_2 = (5,6)(8,7)(6,7)(8,6)$

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

$c_2 = (6.425, 6.5)$

Scanned with CamScanner

Now mean is same so stop it.

Q #12

$P(yes) = 9$, $P(No) = 5$

total = 14

calculating Entropy

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

$$\frac{5}{9+5} = 0.940$$

Now, calculating entropy for each attribute

outlook	yes	no	Entropy
sunny	2	3	0.971
Rain	3	2	0.971
overcast	4	0	0

using formula

$$E(\text{overlook} = \text{sunny}) = -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\right.$$

$$= 0.971$$

similarly calculate others.

now, calculating average information entropy

$$I(\text{outlook}) = \frac{3+2}{9+5} \times 0.971 + \frac{2+3}{9+5} \times 0.971 + \frac{4+0}{9+5} \times 0$$

$= 0.693$

Hence further splitting is required. we will have to calculate entropy for sunny & Rainy values of overcast

For these two values

p(yes) = 2, p(no) = 3

total = 5

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

$E(Sunny) = 0.971$

For Humidity,

Humidity	yes	no	entropy
high	0	3	0
normal	2	0	0

Avg info Entropy

I (humidity) = 0

Gain (0.971)

For wind,

wind	yes	no	Entropy
strong	1	1	1
weak	1	2	0.918

I (windy) = 0.951

Gain = 0.020

For Temperature

Temp	yes	no	Entropy
Cool	1	0	0
Hot	0	2	0
mild	1	1	1

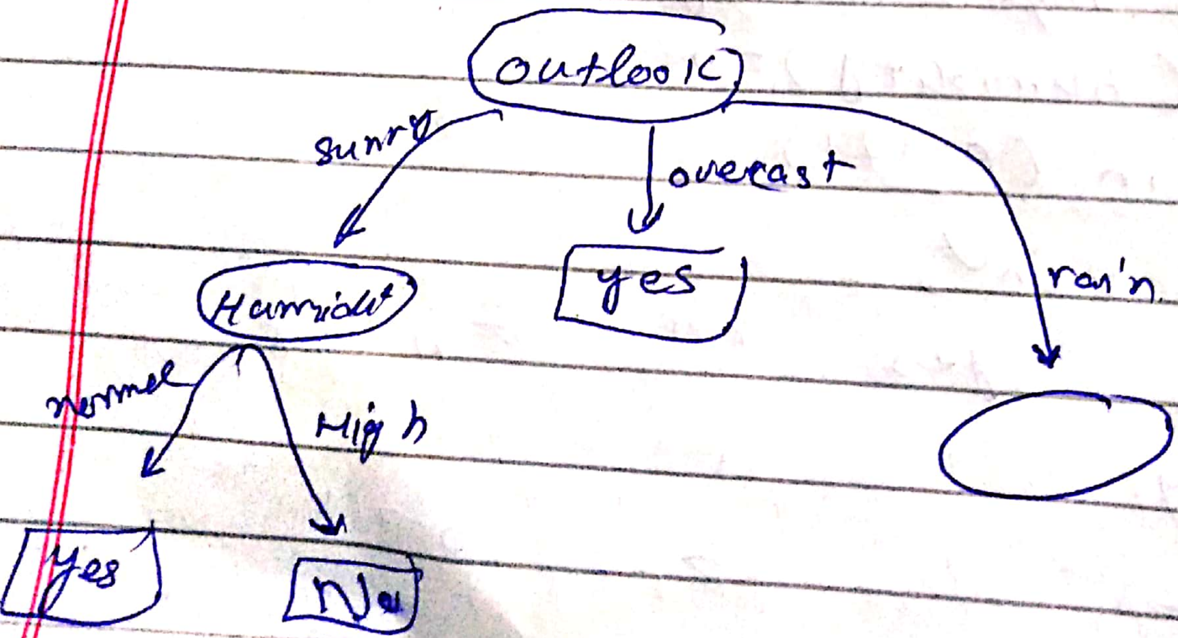
I (Temp) = 0.4

Gain = 0.571

now,

Attributes	gain
Temperature	0.571
Humidity	0.971
windy	0.02

next node humidity.



now. Calculating gain

Gain = Entropy(S) − I(Attribute)

Entropy(S) = 0.940

Gain(outlook) = 0.940 − 0.693 = 0.247

For temperature,

Temp	yes	No	Entropy
Hot	2	2	1
mild	4	2	0.918
cool	3	1	0.811

Avg information Entropy

$$I(\text{Temp}) = \frac{2+2}{9+5} \times 1 + \frac{4+2}{9+5} \times 0.918 + \frac{3+1}{9+5} \times 0.811$$

$$= 0.911$$

Entropy(S) = 0.940

Gain = 0.940 − 0.911 = 0.029

For Humidity

Humidity	yes	No	Entropy
High	3	4	0.985
Normal	6	1	0.591

Avg information Entropy

$$I(\text{Humidity}) = \frac{3+4}{9+5} \times 0.985 + \frac{6+1}{9+5} \times 0.591$$

$$= 0.788$$

E(S) = 0.940

Gain(Humidity) = 0.940 − 0.788 = 0.152

For wind,

wind	yes	No	Entropy
strong	3	3	1
weak	6	2	0.811

Avg Info Entropy

$$I(windy) = \frac{3+3}{9+5} \times 1 + \frac{6+2}{9+5} \times 0.811 = 0.892$$

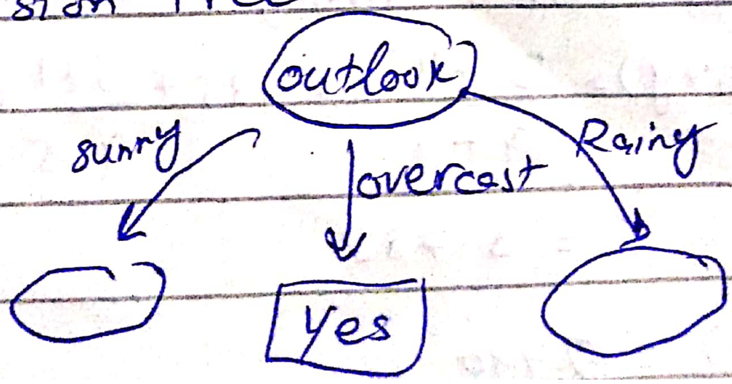
$E(S) = 0.940$

Gain(wind) = 0.940 − 0.892 = 0.048

Since, outlook has biggest gain. it will be ours root node.

Attributes	Gain
outlook	0.247
Temp	0.029
Humidity	0.152
wind	0.048

Decision Tree



now for rain, further splitting is necessary

P = 3 , N = 2

Total = 5

$$Entropy(S_{rainy}) = -\frac{3}{3+2}\log_2\frac{3}{3+2} - \frac{2}{3+2}\log_2\left(\frac{2}{2+3}\right)$$

$= 0.971$

now, calculating entropy for each attribute

For Humidity

Avg infor Entropy

I(Humidity) = 0.951

Gain = 0.020

For wind,

Avg info entropy

I(windy) = 0

Gain = 0.971

For Temp,

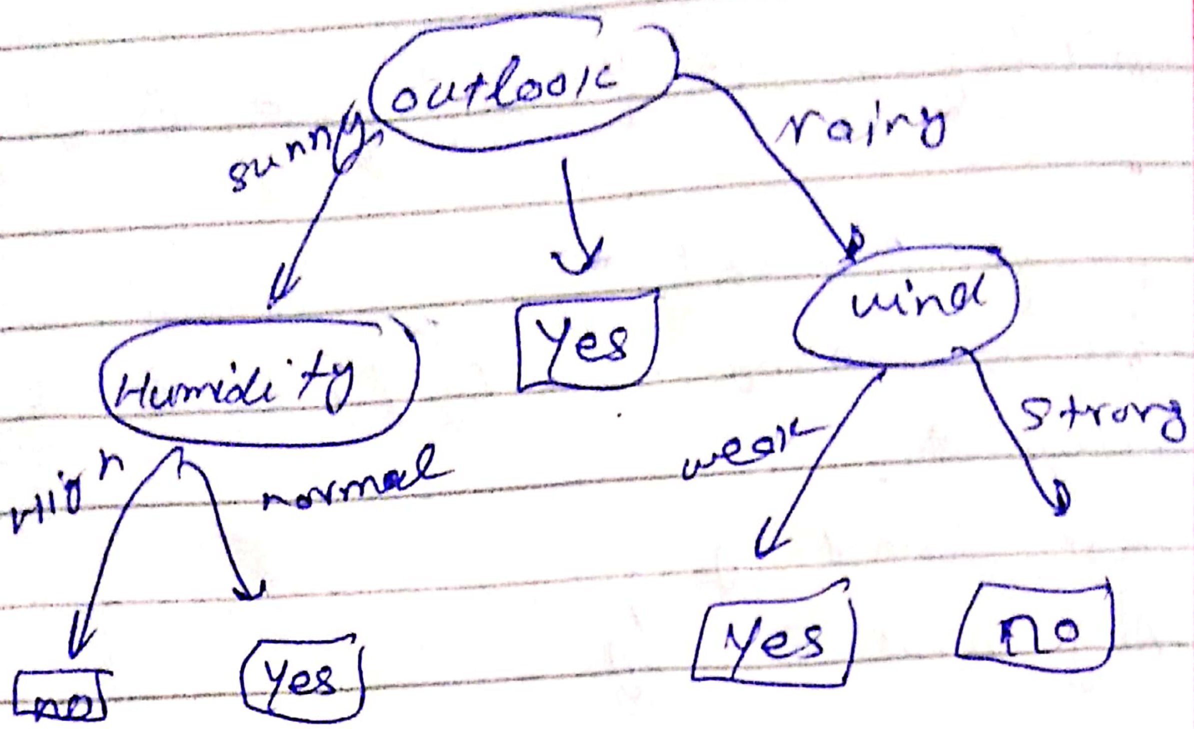
I(Temp) = 0.951

Gain = 0.020

Attributes	Gain
Humidity	0.02
wind	0.971
Temperature	0.02

Next node : wind

Final decision Tree :



Q#4

	B/~~N~~ a	B/N b	core c	Core d	core e	B/~~N~~ f	B/~~N~~ g
a	0.00	2.83	4.24	4.12	2.24	6.40	4.12
b	2.83	0.00	5.10	3.61	3.00	6.71	6.0[illegible]
c	4.24	5.10	0.00	2.24	2.24	2.24	2.2[illegible]
d	4.12	3.61	2.24	0.00	2.00	3.16	[illegible]
e	2.24	3.00	2.24	2.00	0.00	4.24	3.[illegible]
f	6.40	6.71	2.24	3.16	4.24	0.00	4.[illegible]
g	4.12	6.08	2.24	4.24	3.16	4.00	0.[illegible]

ε - neighborhood = 2.25

Core points:- c, d, e

Barder Points: a, f, g
Noise Points: b