Module - 3

(4)

TID	items
T_1	I_1 I_2 I_5
T_2	I_2 I_4
T_3	I_2 I_3
T_4	I_1 I_2 I_4
T_5	I_1 I_3
T_6	I_2 I_3
T_7	I_1 I_3
T_8	I_1 I_2 I_3 I_5
T_9	I_1 I_2 I_3

-> 1 item set

item	frequency	
I_1	6	✓
I_2	7	✓
I_3	5	✓
I_4	2	✓
I_5	2	✓

→ 2 - item set

items	frequency	
I_1 I_2	4	✓
I_1 I_3	4	✓
I_1 I_4	1	
I_1 I_5	2	✓
I_2 I_3	4	
I_2 I_4	2	✓
I_2 I_5	2	✓
I_3 I_4	0	
I_3 I_5	1	
I_4 I_5	0	

→ 3 - item set

item	frequency	
I_1 I_2 I_3	2	✓
I_1 I_2 I_4	1	
I_1 I_2 I_5	2	✓
I_2 I_3 I_4	0	
I_2 I_3 I_5	1	
I_2 I_4 I_5	0	
I_3 I_4 I_5	0	

→ 4 item set

item	frequency
I_1 I_2 I_3 I_5	1

∴ 3 items set is our solution

frequent items are

$\{1, 2, 3\}$ and $\{1, 2, 5\}$

confidence $(x \rightarrow y) = \frac{freq(x, y)}{freq(x)}$

min confidence = 50%.

∴ $\{1\} \rightarrow \{2, 3\}$

confidence $= \frac{2}{6} \Rightarrow 33.33\%$.

$\{2\} \rightarrow \{1, 3\}$

confidence $= \frac{2}{7} \Rightarrow 28.57\%$.

$\{3\} \rightarrow \{1, 2\}$

confidence $= \frac{2}{5} \Rightarrow 40\%$.

$\{1, 2\} \rightarrow \{3\}$

confidence $= \frac{2}{4} \Rightarrow 50\%$.

$\{1, 3\} \rightarrow \{2\}$

confidence $= \frac{2}{4} \Rightarrow 50\%$.

$\{1\} \rightarrow \{2, 5\}$

confidence $= \frac{2}{6} \Rightarrow 33.33\%$

$\{2\} \rightarrow \{1, 5\}$

confidence $= \frac{2}{7} \Rightarrow 28.57\%$

$\{5\} \rightarrow \{1, 2\}$

confidence $= \frac{2}{2} \Rightarrow 100\%$

$\{1, 5\} \rightarrow \{2\}$

confidence $= \frac{2}{2} \Rightarrow 100\%$

$\{2, 5\} \rightarrow \{1\}$

confidence $= \frac{2}{2} \Rightarrow 100\%$

$\therefore$ strong association rules are

$\{1, 2\} \rightarrow \{3\}$

$\{2, 3\} \rightarrow \{3\}$

$\{1, 2\} \rightarrow \{5\}$

$\{2, 5\} \rightarrow \{1\}$

$\{1, 3\} \rightarrow \{2\}$

$\{5\} \rightarrow \{1, 2\}$

$\{1, 5\} \rightarrow \{2\}$

(5) Design the multilevel multidimensional pattern mining for flipKart

customer (x, "new") ^ product type (x, 'Emobile") →
buys - products (x, "yes")

customer (x, "old") ^ product type (x, "Laptop")
→ buys - product (x, "No")

S. NO	customer	product - type	Buys - product
1	new	mobile	yes
2	New	laptop	yes
3	old	mobile	yes
4	old	laptop	yes
5	New	mobile	no
6	New	laptop	no
7	old	mobile	no
8	old	laptop	no

Attributes	item
New	I_1
old	I_2
mobile	I_3
laptop	I_4
yes	I_5
no	I_6

min support = 1

L1

item1	frequency
I_1	3
I_2	3
I_3	3
I_4	3
I_5	3
I_6	3

L2

item1	frequency
I_1 I_2	0
I_1 I_3	1
I_1 I_4	2
I_1 I_5	1
I_1 I_6	2
I_2 I_3	2
I_2 I_4	1
I_2 I_5	2
I_2 I_6	1
I_3 I_4	0
I_3 I_5	3
I_3 I_6	0
I_4 I_5	0
I_5 I_6	3
	0
	1

L3

itemset	count
I_1 I_3 I_5	1
I_1 I_4 I_6	2
I_2 I_3 I_5	2
I_2 I_4 I_6	1

4 item set is not possible

we select I_5, I_6 labels as they represent final class labels

associate rule	confidence
$I_1 \wedge I_3 \rightarrow I_5$	100%
I_2 $I_3 \rightarrow I_5$	100%
I_1 $I_4 \rightarrow I_6$	100%
I_2 $I_4 \rightarrow I_6$	100%

customer (x, 'new') ^ product (x, 'mobile') -> buys (x, 'yes')

customer (x, 'old') ^ product (x, 'laptop') -> buys (x, 'No')

customer (x, 'new') ^ product (x, 'laptop') -> buys (x, 'No')

customer (x, 'old') ^ product (x, 'mobile') -> buys (x, 'yes')

Module - 4

(1) Design DT for play tennis example

outlook	Temperature	Humidity	windy	Play tennis
Sunny	hot	high	false	NO
Sunny	hot	high	true	NO
overcast	hot	high	false	yes
Rainy	mild	high	false	yes
Rainy	cool	normal	false	yes
Rainy	cool	normal	true	NO
overcast	cool	normal	true	yes
Sunny	mild	high	false	NO
Sunny	cool	normal	false	yes
Rainy	mild	normal	false	yes
Sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes

Attribute 1 OOK

$S = [9+, 5-]$ => Entropy (S)

$= -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14}$ => 0.94

S_{sunny} [2+, 3-] ⇒ Entropy (S) =

$$-\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} \Rightarrow 0.971$$

S_{rain} [3+ 2-] ⇒ Entropy (S) =

$$-\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.971$$

$S_{overcast}$ [4+ 0-] → 0

$$\text{Gain}(S, \text{outlook}) = 0.94 - \frac{5}{14}(0.971) - \frac{4}{14}(0)$$

$$\frac{5}{14}(0.91) = 0.2464$$

Attribute temperature

$$S = 0.94$$

S_{hot} = [2+ 2-] ⇒ 1

S_{mild} = [4+, 2-] ⇒ Entropy (S)

$$= -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6} \Rightarrow 0.9183$$

S_{cool} = [3+ 1-] ⇒ entropy (S)

$$= -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} \Rightarrow 0.8113$$

Gain (S, temperature)

$$= 0.94 - \frac{4}{14}(1) - \frac{6}{14}(0.9183) - \frac{4}{14}(0.8113)$$

$$= 0.0289$$

Attribute humidity

$S[9+, 5-] = 0.94$

$S_{high}[3+, 4-] = 0.9852$

$S_{normal}[6+, 1-] = 0.5916$

$Gain(S, humidity) = 0.1516$

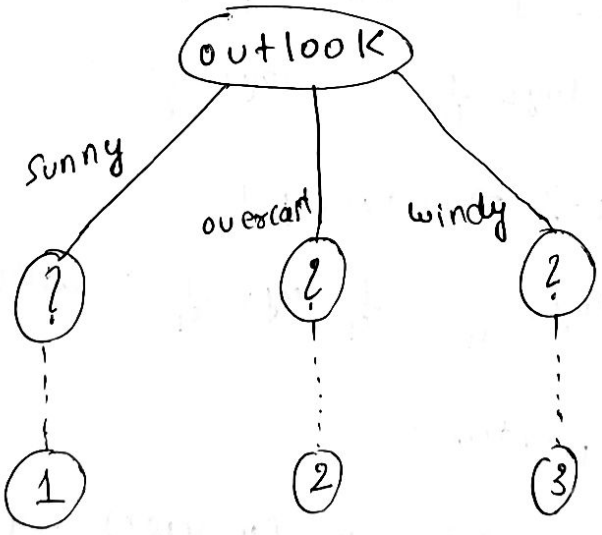
Attribute windy

$S[9+, 5-] = 0.94$

$S_{true}[3+, 3-] = 1$

$S_{false}[6+, 2-] = 0.8113$

$Gain(S, windy) = 0.0478$



as we can see from tables

outlook , owercast ⇒ yes

∴ ② is yes

for sunny, table=

temperature	humidity	windy	playtennis
hot	High	false	No
hot	high	true	No
mild	high	false	No
cool	normal	false	yes
mild	normal	false	yes

$S_{sunny} = [2+ 3-] = 0.97$

S_{hot} $[0+ 2-] = 0$

S_{mild} $[1+ 1-] = 1$

S_{cool} $[1+ 0-] = 0$

∴ Gain (sunny, temperature) = 0.57

S_{high} $[0+ 3-] = 0$

S_{normal} $[2+ 0-] = 0$

∴ Gain (sunny, humidity) = 0.97

for rain, table =

temperature	humidity	windy	play tennis
mild	high	false	yes
cool	normal	false	yes
cool	normal	true	no
mild	normal	false	yes
mild	high	true	no

$S_{rain} = 0.97$

$S_{hot} = [0 + 0-] = 0$

S_{mild} $[2 + 1-] = 0.91$

S_{cool} $[1 + 1-] = 1$

Gain (rain, temperature) $= 0.0192$

S_{high} $[1 + 1-] = 1$

S_{normal} $[2 + 1-] = 0.91$

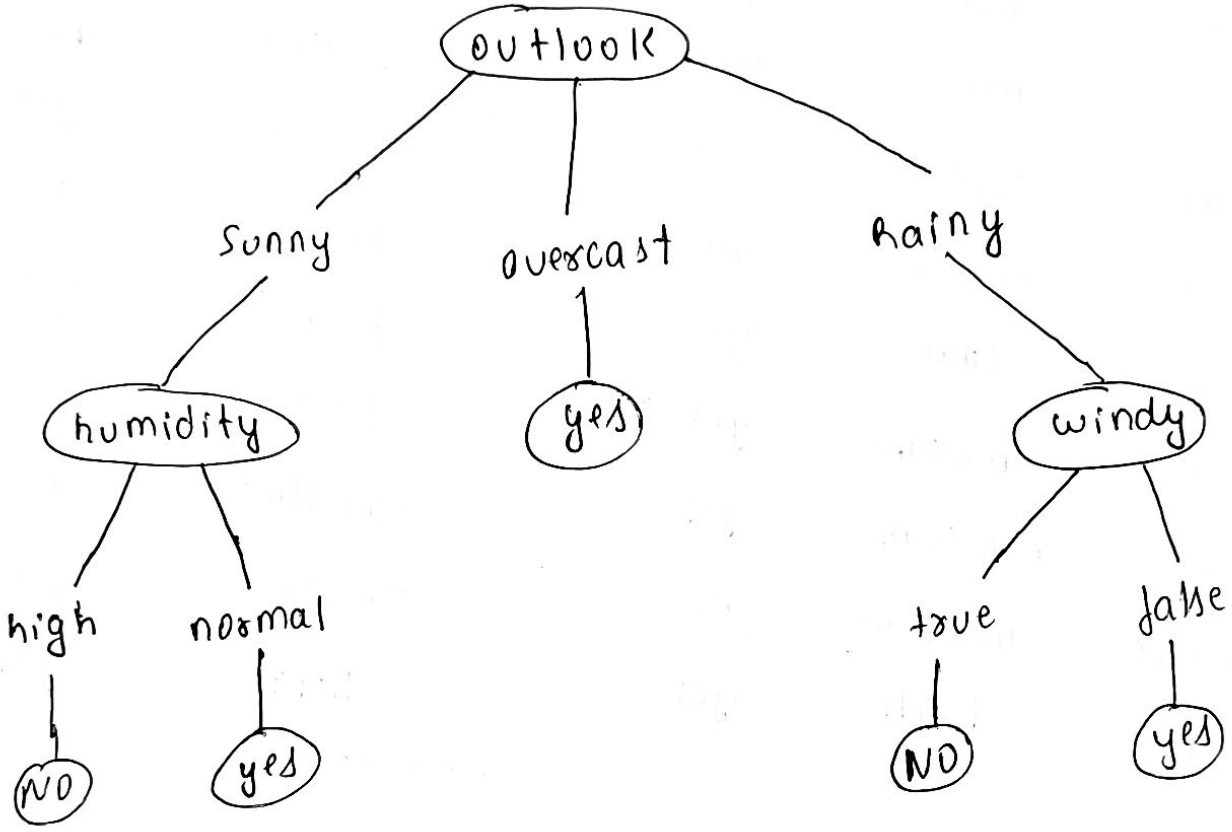
Gain (Rain, humidity) $= 0.0192$

as

Rainy, windy, true ⇒ no

Rainy, windy, false ⇒ yes

we get DT as



---

(2) Design Bayesian classifier for buying computer (table)

age	income	student	credit-rating	computer
<=30	high	no	fair	no
≤30	high	no	excellent	no
31-40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	exceellect	no
31-40	low	yes	excellet	yes
≤30	medium	no	fair	no
≤30	low	yes	fair	yes
>40	medium	yes	fair	yes
≤30	medium	yes	excellent	yes
31-40	medium	no	excellent	yes
>40	high	yes	fair	yes

$$P(\text{yes} \mid \text{age} \leq 30)$$

$$= \frac{P(\text{medium} \mid \text{yes}) + P(\text{yes} \mid \text{yes}) + P(\text{fair} \mid \text{yes})}{P(\text{age} \leq 30)}$$

$$P(\text{yes}) = \frac{9}{14} = 0.64$$

$$P(\text{no}) = \frac{5}{14} = 0.35$$

$P(\text{age} \leq 30 \mid \text{yes}) = \frac{2}{9} \Rightarrow 0.22$

$P(\text{age} \leq 30 \mid \text{NO}) = \frac{3}{5} \Rightarrow 0.6$

$P(\text{income} = \text{medium} \mid \text{yes}) = \frac{4}{9} \Rightarrow 0.44$

$P(\text{income} = \text{medium} \mid \text{NO}) = \frac{2}{5} = 0.4$

$P(\text{student} = \text{yes} \mid \text{yes}) = \frac{6}{9} = 0.66$

$P(\text{student} = \text{yes} \mid \text{no}) = \frac{1}{5} = 0.2$

$P(\text{credit} = \text{fair} \mid \text{NO}) = 6/9 = 0.66$

$P(\text{credit} = \text{fair} \mid \text{NO}) = \frac{2}{5} = 0.4$

$$P(\text{yes} \mid E) = \frac{0.22 \times 0.44 \times 0.66 \times 0.64}{P(E)}$$

$$= \frac{0.028}{P(E)}$$

$$P(\text{NO} \mid E) = \frac{0.6 \times 0.4 \times 0.2 \times 0.4 \times 0.35}{P(E)}$$

$$= \frac{0.007}{P(E)}$$

$\therefore$ The algorithm classifies to yes.