Name:- Sanket Chandrashekhar Harvande (19)

Sign:

Page no.: 01

Date: / /

Assignment No. 02

Q.1 Apply the Naive Baye's classifier algorithm for buys computer classification & classify the tuples:

(X = < age = "young", income = "medium", student = "yes" and credit-rating = "fair")

Y = < Middle, High, Yes, Good >

Z = < old, High, No, fair >

ID	Age	Income	Student	Credit Rating	Buys computer
1	Young	High	No	Fair	No
2	Young	High	No	Good	No
3	Middle	High	No	Fair	Yes
4	old	Medium	No	Fair	Yes
5	old	Low	Yes	Fair	Yes
6	Old	Low	Yes	Good	No
7	Middle	low	Yes	Good	Yes
8	Young	Medium	No	fair	No
9	Young	Low	Yes	fair	Yes
10	old	Medium	Yes	Fair	Yes
11	Young	Medium	Yes	Good	Yes
12	Middle	medium	No	Good	Yes
13	Middle	High	Yes	Fair	Yes
14	old	Medium	No	Good	No.

→ Let C_1 → buys_comp = "yes"

C_2 → buys-comp = "xs"

1. Calculate Class probability

P(1) = 9/14

P(2) = 5/14

Calculate $P(C_1|x) = P(x|C_1) \cdot P(C_1)$

$$P(x|C_1) = \prod_{k=1}^{n} p(x_k|C_1)$$

$\therefore p(x_1|C_1)\, p(\text{age} = \text{"young"}|C_1)$

$= 2/9$

$\therefore p(x_2|C_1) = p(\text{income} = \text{"med"}|C_1)$
$= 4/9$

$\therefore p(x_3|C_1) = p(\text{student} = \text{"yes"}|C_1)$
$= 6/9$

$p(x_4|C_1) = p(\text{c.r.} = \text{"fair"}|C_1)$
$\therefore O(x_1|C_1) = 2/9 \cdot 4/9 \cdot 6/9 \cdot 6/9$

$= 0.0282$

2. Calculate $p(C_2|x) = p(x|C_2) \cdot p(C_1)$
$p(x|C_2) = \prod_{k=1}^{n} p(x_k|C_2)$

$k = 1$

$\therefore p(x_1|C_2) = p(\text{age} = \text{"young"}|C_2)$
$= 3/6$

$p(x_2|C_2) = p(\text{income} = \text{""}/C_2)$
$= 2/6$

$p(x_3/C_3) = p(\text{studmt} = \text{"yes"}|C_2)$
$= 1/6$

$p(x_4|C_2) = p(C_1 = \text{"fair"}/C_2)$
$= 2/6$

$\therefore p(x/C_2) = 3/1 \;\; 2/0 \;\; 1/6 \;\; 2/6$

$\therefore P(C_2|X) = 3/6 \quad 2/6 \quad 1/6 \quad 2/6 \quad 6/4$
$= 0.0036$

$\therefore P(C_1/X) > P(C_2/X)$
$X \in C_1$

$\therefore X \in$ buys-comp = "yes"

Step 2 :- Compute $P(Y|C_i)$ for each class.

P(age = "middle" | buys-computer = "Yes") = 4/9 = 0.444
P(age = "middle" | buys-computer = "No") = 0.

P(income = "high" | buys-computer = "Yes") = 2/9 = 0.222
P(income = "High" | buys-computer = "No") = 2/5 = 0.4

P(student = "Yes" | buys-computer = "Yes") = 6/9 = 0.667
P(student = "Yes | buys-computer = "No") = 1/5 = 0.2

P(credit-rating = "Good" | buys-computer = "Yes") = 3/9 = 0.333
P(credit-rating = "good" | buys-computer = "No") = 3/5 = 0.6

P(Y | Ci) = P(Y | buys-computer = "Yes")
= 0.444 × 0.222 × 0.6667 × 0.333
= 0.021

= P(Y | buys-computer = "No")
= 0. × 0.4 × 0.2 × 0.6 = 0.

P(Y|Ci) * P(Ci) = P(Y | by-computer = "Yes") * P(buys-computer = "Yes")
= 0.021 × 0.043
= 0.0315

= P(Y | buy-computer = "no") * P(buys-computer = "no")
= 0 × 0.357
= 0.

Therefore Y belongs to class ("buy-computer = "Yes")

Step 3 :- compute $P(z/C_i)$ for each class.

P(age = "old" | buys-computer = "Yes") = 3/9 = 0.333
P(age = "old" | buys-computer = "No") = 2/5 = 0.4

P(income = "High" | buys-computer = "Yes") = 2/9 = 0.222
P(income = "High" | buys-computer = "No") = 2/5 = 0.4

P(student = "No" | buys-computer = "Yes") = 3/9 = 0.333
P(student = "No" | buys-computer = "No") = 4/5 = 0.8

P(credit rating = "Fair" | buys-computer = "Yes") = 6/9 = 0.666
P(credit-rating = "Fair" | buys-computer = "no") = 2/5 = 0.4.

$P(z|C_i)$ = P(z | buys-computer = "Yes")
: 0.333 x 0.222 x 0.333 x 0.666
: 0.016.

P(z | buys-computer = "No")
: 0.4 x 0.4 x 0.4 x 0.8
: 0.051

$P(z|C_i) * P(C_i)$: P(z | buys-computer = "Yes") * P(buys_comp = "Yes")
: 0.016 x 0.043
: 0.01028

P(z | buy-computer = "No") * P(buys_computer = "No")
= 0.051 x 0.357 = 0.0182

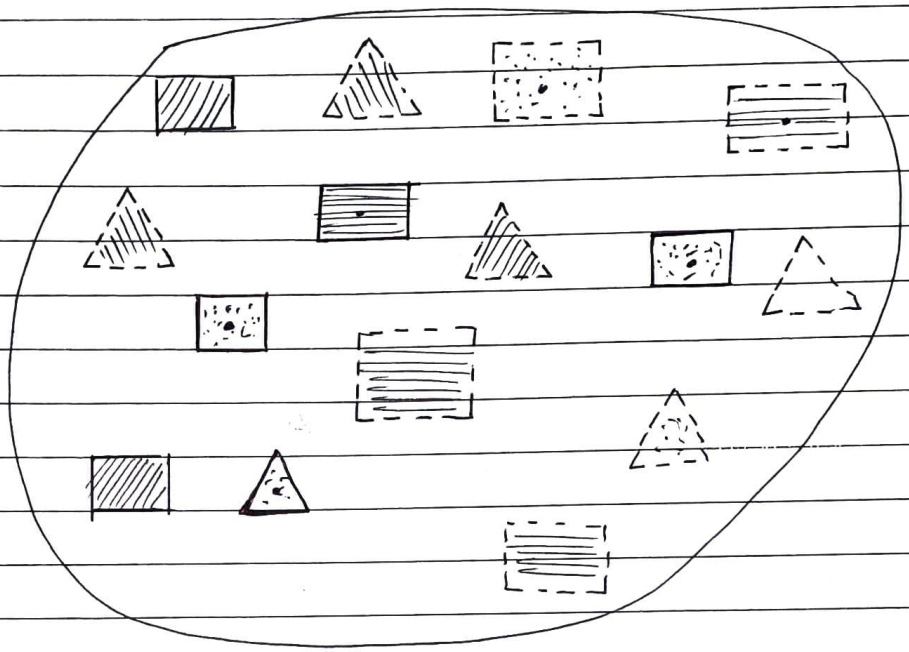
Therefore, z belongs to class ("buys_computer = "no")

Sanket Harvande (19)

Page no.: 5

Date: / /

Q 2. Data set : A set of classified objects is given in the figure below. Apply ID3 to generate tree.
(Attributes : color, outline, Dot, shape).



→

	Colour	Outline	Dot	shape
1	Green	Dashed	No	Triangle
2	Green	Dashed	Yes	Triangle
3	Yellow	Dashed	No	square
4	Red	Dashed	No	Square
5	Red	solid	No	Square
6	Red	solid	Yes	Triangle
7	Green	solid	No	Square
8	Green	Dashed	No	Triangle
9	yellow	Solid	Yes	Square
10	Red	Solid	No	Square
11	Green	Solid	Yes	Square
12	Yellow	Dashed	Yes	square
13	Yellow	Solid	No	square
14	Red.	Dashed.	Yes	Triangle.

Class N : shape = "Triangle"
Class P : shape = "Square"
Total number of records = 14.

P(square) = 9/14 (P)
P(Triangle) = 5/14 (N)

$$I(P,N) = \frac{-P}{P+N}\log_2\frac{P}{P+N} - \frac{N}{P+N}\log_2\frac{N}{P+N}$$

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

$$= 0.940$$

Step 1 :- Compute the entropy for colours.
For color=Red
P_i = with "square" class = 3
n_i = with "Triangle" class = 2

$$I(P_i,N_i) = \frac{-3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5}$$

$$= 0.971$$

Color	P_i	n_i	$I(P_i, n_i)$
Red	3	2	0.971
Green	2	3	0.971
Yellow	4	0	0

$$E(A) = \sum_{i=1}^{v} \frac{P_i + n_i}{P+N} I(P_i, n_i)$$

$$E(color) = \frac{5}{14} \times I(3,2) + \frac{5}{14} \times I(2,3) + \frac{4}{14} \times I(4,0)$$

$$= 0.694$$

Hence G(S, color) = I(P,n) - E(color)
= 0.940 - 0.694
= 0.266.

Compute the entropy for outline.

outline	Pi	ni	I(Pi, ni)
Dashed	3	4	0.985
solid	6	1	0.621

$E(\text{outline}) = \frac{7}{14} \times 0.985 + \frac{7}{14} \times 0.621$

$= 0.803$

G(S, outline) = I(P,n) - E(outline)
= 0.940 - 0.803
= 0.137.

Compute entropy for dot.

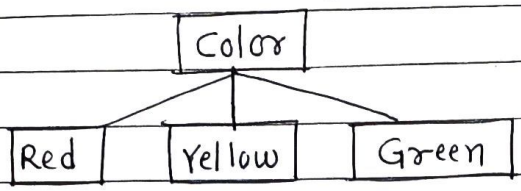
dot	Pi	ni	I(Pi, ni)
No	6	2	0.811
Yes	3	3	1

$E(\text{dot}) = \frac{8}{14} \times 0.811 + \frac{6}{14} \times 1$

$= 0.892$

G(S, dot) = I(P,n) - E(dot)
= 0.940 - 0.892
= 0.048

Therefore
Gain(color) = 0.246
Gain(outline) = 0.137
Gain(dot) = 0.048



Step 2:- As attribute color is at the node

Color	outline	dot	shape
Red	Dashed	No	Square
Red	Solid	No	Square
Red	solid	Yes	~~Square~~ Triangle.
Red	Solid	No	Square
Red	Dashed	Yes	Triangle

$I(P,n) = -3/5 \log_2 3/5 - 2/5 \log_2 2/5 = 0.971$

outline	Pi	ni	I(Pi, ni)
Dashed	1	1	1
Solid	2	1	0.918

$E(\text{outline}) = 2/5 \times 1 + 3/5 \times 0.918$
$= 0.951$

$\text{Gain}(S_{red}, \text{outline}) = I(P,n) - E(\text{outline})$
$= 0.971 - 0.951$
$= 0.02$

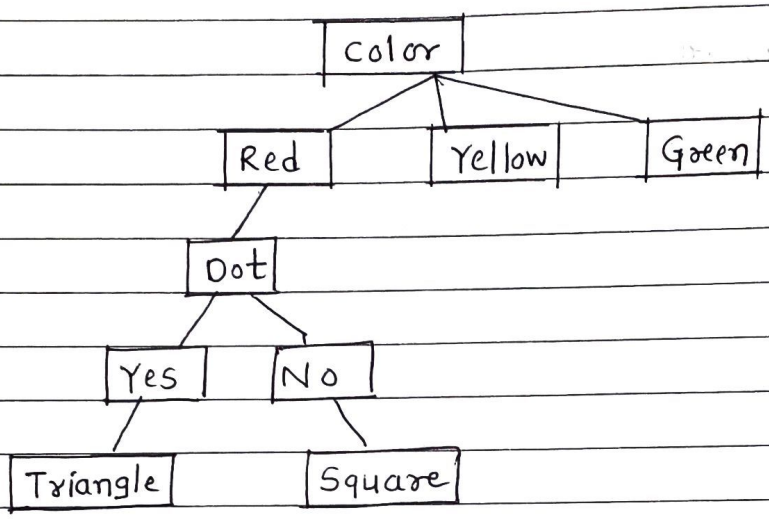
Compute entropy for dot

dot	Pi	ni	I(Pi, ni)
No	3	0	0
Yes	0	2	0

$E(\text{dot}) = 0.$

Gain (S_{red}, Dot) = 0.971 - 0 = 0.971

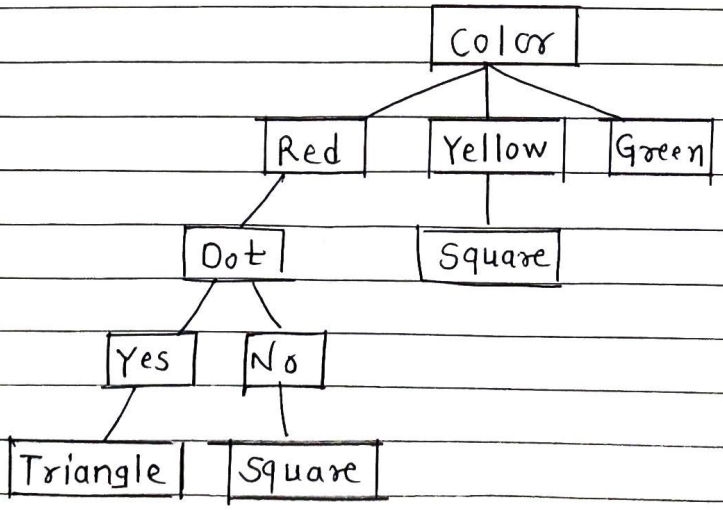
∴ G(outline) = 0.951
G(Dot) = 0.971



Step 3 : Color = Yellow

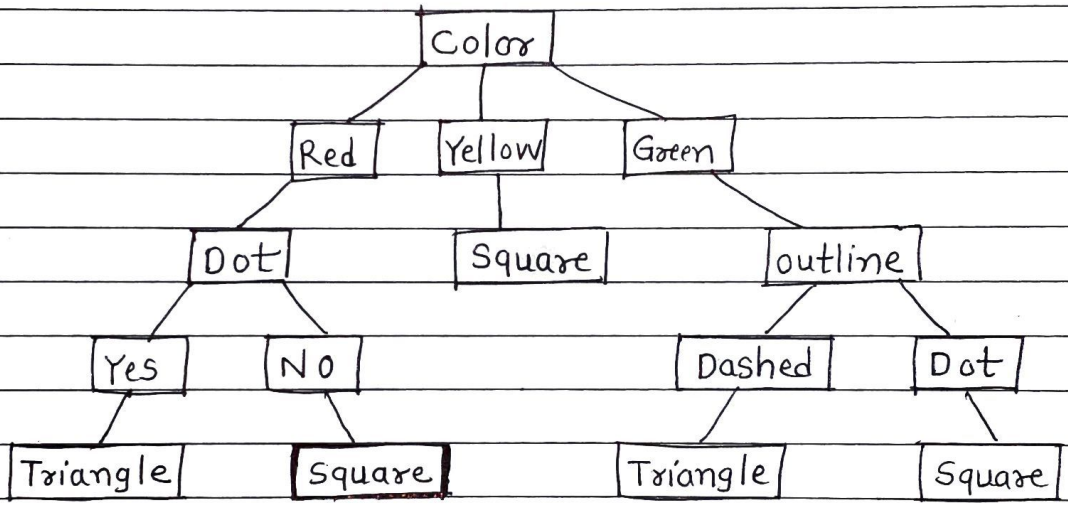
Color	outline	dot	shape
Yellow	Dashed	No	Square
Yellow	solid	Yes	Square
Yellow	Dashed	Yes	square
Yellow	solid	No	square

As all the tuples are label to square so directly assign as Yellow = "Square"



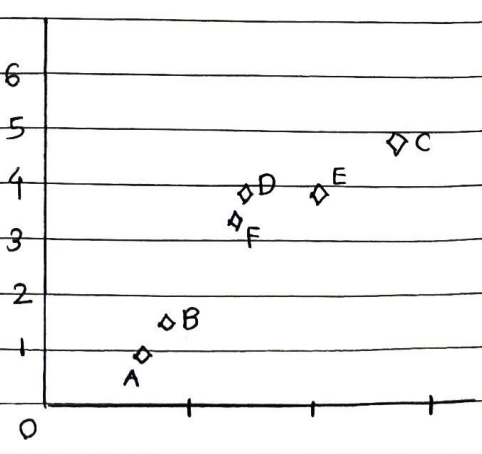
Step 4 :- Color = green

Color	outline	Dot	Shape
Green	dashed	no	triangle
Green	dashed	yes	triangle
Green	Solid	no	square
Green	dashed	no	triangle
Green	solid	yes	Square



Q.3 Suppose we have 6 objects (with name A, B, C, D, E & F) and each object have two measured features (X_1 & X_2) Apply single linkage clustering & draw dendograph.

	X_1	X_2
A	1	1
B	1.5	1.5
C	5	5
D	3	4
E	4	4
F	3	3.5



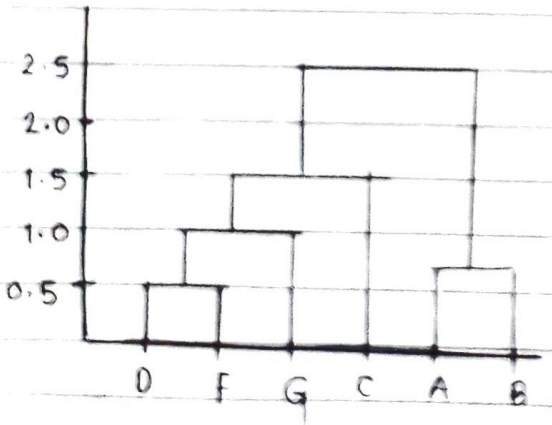
Sanket Harvande (19)

Page no.: 11

Date: / / /

	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

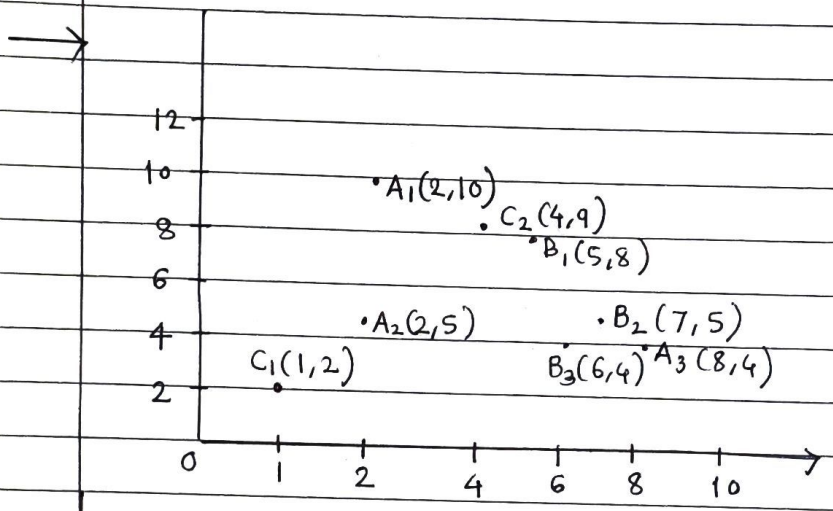
1. In the beginning we have 6 clusters: A, B, C, D, E & F.
2. We merge cluster D & F into cluster C, D, F) at distance 0.50
3. We merge cluster A & B into cluster (A, B) at distance 0.71
4. We merge cluster E & C, D, F) into (C, D, F), E) at dist 1.00.
5. We merge cluster ((D, F), E) & C into (CC, D, F), E), C) at distance 1.41
6. We merge cluster (((D, F), E), C) & (A, B) into (((D, F), E), C), (A, B) at distance 2.50
7. The last cluster contains all the objects. thus conclude the computation.



Sanket Harvande (19)

Page no.: 12
Date: / /

Q.4 Suppose that the data mining tasks to cluster the following eight points (with (x,y) representing location) into three clusters A1 (2,10), A2(2,5), A3(8,4), B_1(5,8) B_2(7,5), B_3(6,4), C_1(1,2), C_2(4,9). Use Euclidean distance function. Suppose initially A1, B1, C1 as the center of each cluster, respectively, use the k-means algorithm to show that ~~th~~ final three clusters.



objects - centroids distance

D^0

A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	
0	5	8.48	3.61	7.07	7.21	8.06	2.24	$X_1 = A_1(2,10)$
3.61	4.24	5	0	3.61	4.12	7.21	1.41	$X_2 = B_1(5,8)$
8.06	3.16	7.28	7.21	6.71	5.39	0	7.62	$X_3 = C_1(1,2)$

object clustering

A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	
1	0	0	0	0	0	0	0	$X_1 = A_1(2,10)$
0	0	1	1	1	1	0	1	$X_2 = B_1(5,8)$
0	1	0	0	0	0	1	0	$X_3 = C_1(1,2)$

Iteration - 1

$X_1 = (2,10)$

$X_2 = \left(\frac{8+5+7+6+4}{5}, \frac{4+8+5+4+9}{5}\right) = (6,6)$

$X_3 = \left(\frac{2+1}{2}, \frac{5+2}{2}\right) = (1.5, 3.5)$

D' - Iteration - 1, objects - centroids

A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	
0	5	8.48	3.61	7.07	7.21	8.06	2.24	$x_1 = (2, 10)$
5.66	4.12	2.83	2.24	1.41	2	6.40	3.61	$x_2 (6, 6)$
6.52	1.58	6.52	5.70	5.70	4.52	1.58	6.04	$x_3 (1.5, 3.5)$

G'

A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	
1	0	0	0	0	0	0	0	$x_1 (2, 10)$
0	0	1	1	1	1	0	0	$x_2 (2, 6)$
0	1	0	0	0	0	1	0	$x_3 (1.5, 3.5)$

Iteration - 2

$x_1 = ((2+4)/2), (10+9)/2) = (3, 9.5)$

$x_2 = \left\{ \frac{8+5+7+6}{4}, \frac{4+8+5+4}{4} \right\} = (6.5, 5.25)$

$x_3 = \left\{ \frac{2+1}{2}, \frac{5+2}{2} \right\} = (1.5, 3.5)$

D^2

A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	
1.12	2.35	7.43	2.5	6.02	6.26	7.76	1.12	$x_1 (3, 9.5)$
6.54	4.51	1.95	3.13	0.56	1.35	6.38	7.68	$x_2 (6.5, 2.5)$
6.52	1.58	6.52	5.70	5.70	4.52	[illegible]	6.04	$x_3 (1.5, 3.5)$

G^2 =

A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	
1	0	0	1	0	0	0	1	x_1
0	0	1	0	1	1	0	0	x_2
0	1	0	0	0	0	1	0	x_3

Iteration - 3

$$X_1 = \left(\frac{2+5+4}{3}, \frac{10+9+8}{3}\right) = (3.67, 9)$$

$$X_2 = \left(\frac{8+7+6}{3}, \frac{4+5+4}{3}\right) = (7, 4.33)$$

$$X_3 = \left(\frac{2+1}{2}, \frac{5+2}{2}\right) = (1.5, 3.5)$$

D^3

A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	
1.95	4.33	6.61	1.66	5.20	5.52	7.49	6.33	X_1
6.01	5.04	1.05	4.17	0.67	1.05	6.44	5.55	X_2
6.52	1.58	6.52	5.70	5.70	4.52	1.58	6.04	X_3

G^3

A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	
1	0	0	1	0	0	0	1	X_1
0	0	1	0	1	1	0	0	X_2
0	1	0	0	0	0	1	0	X_3

So final clusters are

Group 1 = $\{A_1, B_1, C_2\}$

Group 2 = $\{A_3, B_2, C_3\}$

Group 3 = $\{A_2, C_1\}$

Q.5 Consider the following transaction databases.

TID	Items
T01	A, B, C, D
T02	A, B, C, D, E, F, G
T03	A, C, G, H, K
T04	B, C, D, E, K
T05	D, E, F, H, L
T06	A, B, C, D, L
T07	B, I, E, K, L
T08	A, B, D, E, K
T09	A, E, F, H, L
T10	B, C, D, F

Apply Apriori algorithm with min-support of 30 % & min confidence of 70 %. & find all the association rule in the dataset.

→ Step 1 :-

Items	Support
A	6
B	7
C	6
D	7
E	6
F	3
G	2
H	3
I	1
K	4
L	4

Sanket Harvande (19)

Item set above 30% ~~student~~ support.

A	6
B	7
C	6
D	7
E	6
F	3
H	3
K	4
L	4

Step 1 2 Generate 2 items

Item	Support
AB	4
AC	4
AD	4
AE	3
AF	1
AH	2
AK	2
AL	2
BC	5
BD	6
BE	4
BF	1
BH	0
BK	3
BL	2
CD	5
CE	2
CF	1
CH	1
CK	2
CL	1
DE	4
DF	2
DH	1
DK	2
DL	2
EF	2
EH	2
EK	3
EL	3
FH	2
FK	0
FL	2
HK	1
HL	2
KL	1

Item above 30% support

AB	4
AC	4
AD	4
AE	3
BC	5
BD	6
BE	4
BK	3
CD	5
DE	4
EK	3
EL	3

Step 3 : Generating 3 item set

Item	support
ABC	3
ABD	4
ABE	2
ABK	1
ACD	3
ACE	1
ADE	2
AEK	1
AEL	1
BCD	5
BCE	2
BCK	1
BDE	3
BDK	2

Items above 30%

Item	Support
ABC	3
ABD	4
ACD	3
BCD	5
BDE	3

Step 4 :- Generate 4 item

Item set	support
ABCD	3
ABDE	2
BCDE	2

following rules generated.

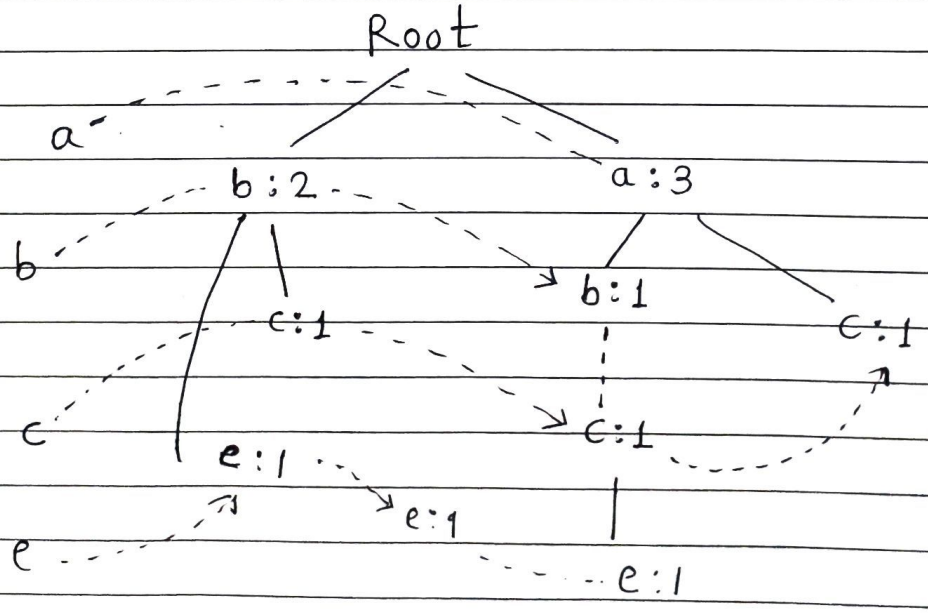
Rule	Confidence	Confidence %
A → BCD	3/6 = 0.5	50
B → ACD	3/7 = 0.43	43
C → ABD	3/6 = 0.5	50
D → ABC	3/7 = 0.43	43
AB → CD	3/4 = 0.75	75
BC → AD	3/5 = 0.6	60
CD → AB	3/5 = 0.6	60
AC → BD	3/4 = 0.75	75
AD → BC	3/4 = 0.75	75
BCD → A	3/5 = 0.6	60
ACD → B	3/3 = 1	100
ABD → C	3/4 = 0.75	75
ABC → D	3/3 = 1	100

above 70% are

$AB \rightarrow CD$
$AC \rightarrow BD$
$AD \rightarrow BC$
$ACD \rightarrow B$
$ABD \rightarrow C$
$ABC \rightarrow D$

Q.6. Transaction item list is given below.
Draw FP tree.
$T_1 = b, e$, $T_2 = a, b, c, e$
$T_3 = b, c, e$ $T_4 = a, c$ $T_5 = a$
Give minimum support = 2.

→



Q.7 What is web Mining? Explain web usage mining with its applications.

→ The process of discovering the useful & previously unknown information from the web data.

Applications :-

1) In case the value of each visitor
2) Collect information in new ways
3) Perform targeted resource management.
4) Test the relevance of content & web site architecture.

Q.8 The training data is supposed to be a part of transportation study regarding mode choice to select Bus, car or Train among commuters along a major route in a city.

Gender	Car Ownership	Travel cost	Income level	Transportation Mode
Male	0	cheap	low	Bus
Male	1	cheap	medium	Bus
Female	1	cheap	medium	Train
Female	0	cheap	low	Bus
Male	1	cheap	medium	Bus
Male	0	standard	medium	Train
Female	1	standard	medium	Train
Female	1	expensive	High	Car
male	2	expensive	medium	Car
female	2	expensive	High	Car

→ Class P :- Transportation mode = Bus
Class Q :- Transportation mode = Train
Class N: Transportation mode = car

No. of records with "Bus" class = 4
No. of records with "Train" class = 3.
No. of records with "Car" class = 3

$I(P,q,n) = (-0.4) \times (-1.322) - (0.3)(-1.737) - (0.3) \times (-1.737)$

$I(4,3,3) = 0.5288 + 0.5211 + 0.5211$

$I(4,3,3) = 1.571$

Step 1:- compute the entropy of gender

for gender = male $P_i = 3, q_i = 1, n_i = 1$

$I(P_i, q_i, n_i) = -(3/5)\log_2(3/5) - (1/5)\log_2(1/5) - (1/5)\log_2(1/5)$

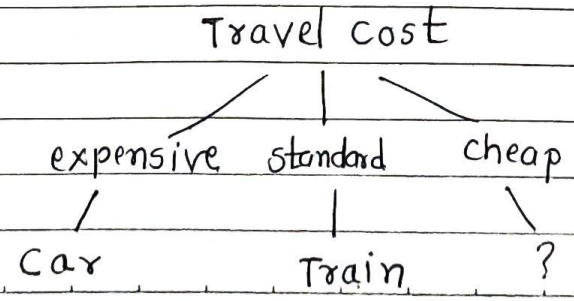
$= 1.371$

Gender	Pi	qi	ni	I(Pi, qi, ni)
Male	3	1	1	1.371
Female	1	2	2	1.522

E(Gender) = 5/10 * 1.371 + 5/10 * 1.522
= 1.447

Gain(Gender) = 1.571 - 1.447 = 0.124

Similarly
G (car ownership) = 0.535
G (Travel cost) = 1.21
G (Income level) = 0.696



Consider travel cost = cheap.

Gender	Car ownership	Travel Cost	Income Level	Transport mode
male	0	Cheap	low	Bus
male	1	Cheap	medium	Bus
female	1	Cheap	medium	Train
female	0	Cheap	low	Bus
male	1	Cheap	medium	Bus.

$I(P, q, n) = (4, 1, 0)$

$= -(4/5)\log_2(4/5) - (1/5)\log_2(1/5) - (0/5)\log_2(0/5)$

$= 0.722$

Compute the entropy for gender

Gender	Pi	qi	ni	I(Pi, qi, ni)
Male	3	0	0	0
Female	1	1	0	1

$E(\text{Gender}) = 3/5 \times 0 + 2/5 \times 1 = 0.4$

$G(\text{Gender}) = 0.722 - 0.4 = 0.322$

compute the entropy for car ownership

Car ownership	Pi	qi	ni	I(Pi, qi, ni)
0	2	0	0	0
1	2	1	0	0.918
2	0	0	0	0

$E(\text{car ownership}) = 0.215 \times 0 + 315 \times 0.918 \times 0.15 \times 0$

$= 0.5551$

$G(\text{car ownership}) = 0.722 - 0.551$

$= 0.171$

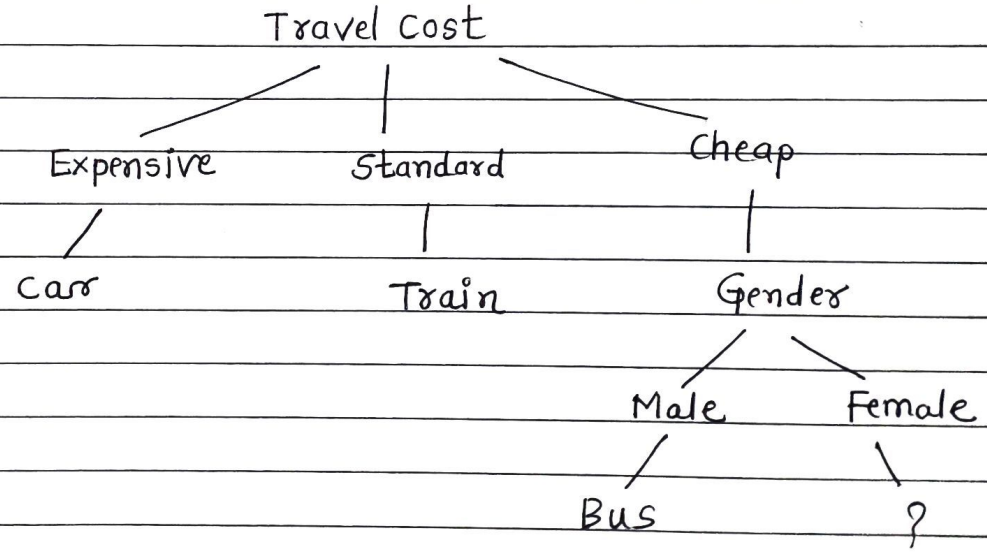
Sanket Harvande(19)
@Harvande

Compute the entropy for income level.

Income level	Pi	qi	ni	I(Pi, qi, ni)
low	2	0	0	0
medium	2	1	0	0.918
High	0	0	0	0

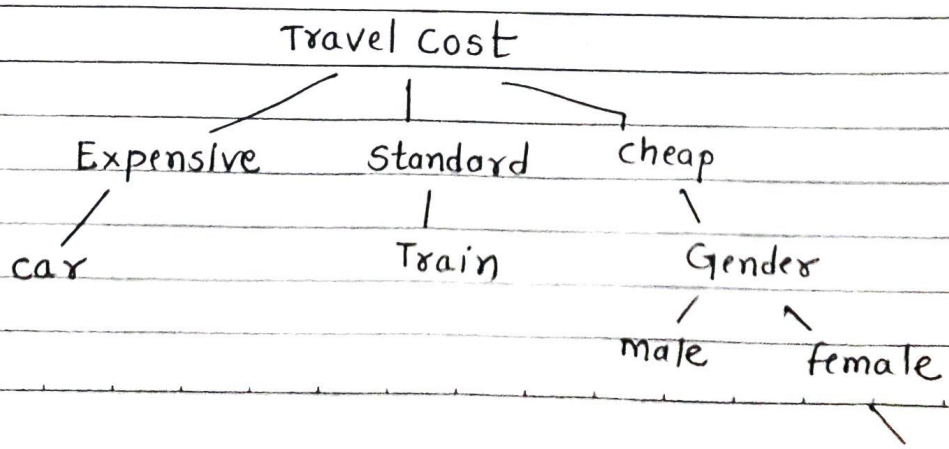
E(income level) = 2/5 * 0 + 3/5 * 0.918 + 0/5 x 0
= 0.551

G(income level) = 0.722 - 0.551 = 0.171

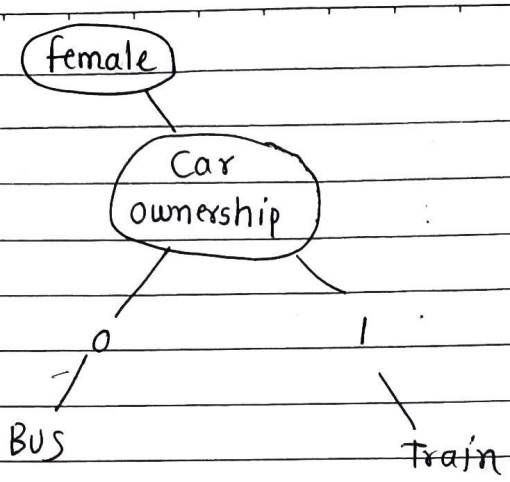


for female

Gender	Car ownership	Income	Transport mode
Female	1	Medium	Train
Female	0	low	Bus.



Sanket C. Harvande



Q.9 Co-ordinates of objects are given below, Apply K-medoids. Number of clusters = 2.

Number	X-co-ordinates	Y-co-ordinates.
1	1	4
2	5	1
3	5	2
4	5	4
5	10	4
6	25	4
7	25	6
8	25	7
9	25	8
10	29	7

→ objects = 1 & 5

Object Number	Dissimilarity from obj 1	Diss. from obj 5	Minimal Dissimilarity	closed representative object.
1	0.00	9.00	0.00	1
2	5.00	5.83	5.00	1
3	4.47	5.39	4.47	1
4	4.00	5.00	4.00	1
5	9.00	0.00	0.00	5
6	24.00	15.00	15.00	5

7	24.08	15.13	15.13	5
8	24.19	15.30	15.30	5
9	24.33	15.52	15.52	5
10	28.16	19.24	19.24	5

Avg = 9.37

Cost = 9.37.

objects 4 & 8

object Number	Dissimilarity from obj 4	Dissimilarity from obj 8	minimal Dissimilarity	Closest representation
1	4	24.19	4.00	4
2	3	20.88	3.00	4
3	2	20.62	2.00	4
4	0	20.22	0.00	4
5	5	15.30	5.00	4
6	20	3	3.00	8
7	20.10	1	1.00	8
8	20.22	0	0.00	8
9	20.40	1	1.00	8
10	24.19	4	4.00	8

Swapping cost = New - old cost
= 2.30 - 9.32
= -7.02

Q.10 Write short Note on

a) Cross-Validation :-

(1) A statistical method or a resampling procedure used to evaluate the skill of machine learning models on a limited data sample.

(2) steps involved in cross-validation

(i). Resolve some portion of sample data set.

(ii) Using the rest data-set train the model
(iii) Test the model using the reserve portion of the data set.

Applications :-
It has great scope in the medical research field.

b) Web - Crawlers :

(1) A web crawler is an automated program that scans or crawls through the internet pages to ~~store~~ create an index of the data.
(2) A web crawler is also known as web spider, web robot, bot, crawler & automatic indexes.
(3) Web crawling is considered to be an important method for collecting data keeping up with the exanding internet.

Different types of crawlers
i) Traditional
ii) Periodic
iii) Incremental
iv) Focused

c) Multilevel & Multidimensional Association Mining

Multivel :-
(1) Items are always in the form of hierachy
(2) Two approaches of multilevel association rule
(3) Using uniform minimum support for all levels
(4) Using reduced minimum support at lower level.

Multidimensional :-

i) Quantitative characteristics are named & consolidated order

ii) Numeric traits should be discretized.

iii) Multidimensional affiliation rule comprises of more than measurement

iv) Three approaches in mining multidimensional affiliation rules can be as follows.

(1) Using static discretization of quantitative qualities.

(2) Using powerful discretization of quantitative traits.

(3) Grid For TUPLES.