SINDH MADRESSATUL ISLAM UNIVERSITY KARACHI

CHARTERED BY GOVERNMENT OF SINDH,RECOGNIZED BY HEC.

NAME: MUSTAFA JAVED

ID: CSC-22S-140

COURSE: ARTIFICAL INTELLIGENCE

DEPARTMENT: COMPUTER SCIENCE

SECTION: 6-C

PROGRAM: BS(CS)

Question No 1

Supervised

Inter Active	Practical Skill	Comm Skills	Job offer
Yes	Very good	Moderate	No
No	Good	Moderate	No
No	Very good	Poor	No
Yes	Good	Moderate	No
No	Average	Poor	Yes
No	Good	Poor	No
Yes	Average	Good	No
No	Very good	Poor	Yes
No	Good	Poor	Yes
No	Very good	Poor	Yes

Entropy

Gain Overall

C1 = yes = 7, C2=No=3

We know that the formula

$$=>-\frac{3}{10}\log log\ 2\left(\frac{3}{10}\right)-\frac{7}{10}\log log\ 2\left(\frac{7}{10}\right)$$

=>0.88

So, It is our overall gain

Date set of Inter Active:Gain

$$\frac{3}{10}\times\left[-\frac{3}{3}\log log\ 2\left(\frac{3}{3}\right)\right]+\frac{7}{10}\times\left[-\frac{3}{7}\log log\ 2\left(\frac{3}{7}\right)-\frac{4}{7}\log log\ 2\left(\frac{4}{7}\right)\right]$$

0+0.69

0.69

0.88-0.69

0.18

Dataset Prac skill

$\frac{4}{10} \times \left[-\frac{2}{4}\log log\ 2\left(\frac{2}{4}\right) - \frac{2}{4}\log log\ 2\left(\frac{2}{4}\right)\right] + \frac{4}{10} \times \left[-\frac{4}{4}\log log\ 2\left(\frac{4}{4}\right)\right] + \frac{2}{10}\left[-\frac{1}{2}\log log\ 2\left(\frac{1}{2}\right) - \frac{1}{2}\log\right.$

0.6

0.88-0.6

0.28

Dataset Comman skill

$\frac{6}{10} \times \left[-\frac{3}{6}\log log\ 2\left(\frac{3}{6}\right) - \frac{3}{6}\log log\ 2\left(\frac{3}{6}\right)\right] + \frac{3}{10} \times \left[-\frac{3}{3}\log log\ 2\left(\frac{3}{3}\right)\right] + \frac{1}{10}\left[-\frac{1}{1}\log log\ 2\left(\frac{1}{1}\right)\right]$

0.6

0.88-0.6

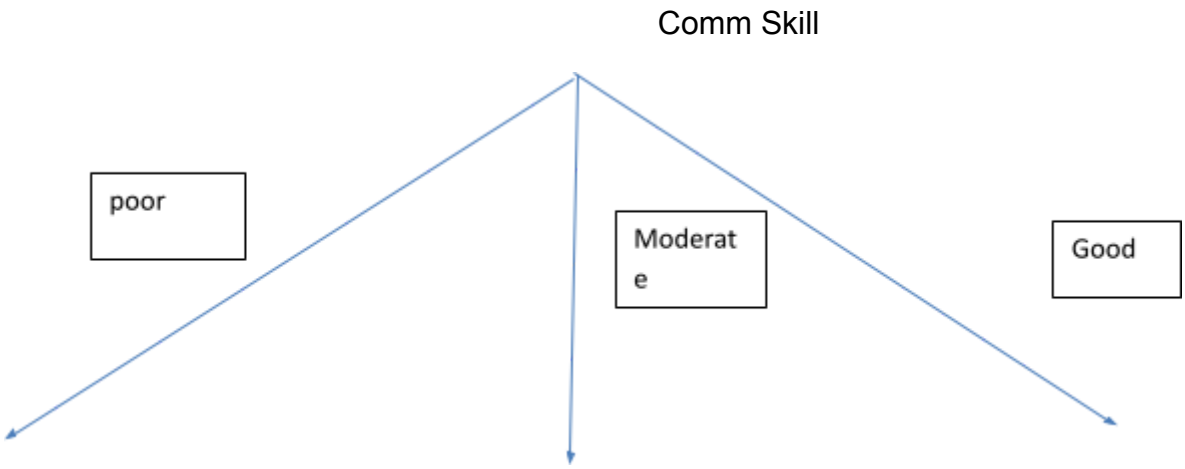
0.28

Gains:

Inter Active:0.18

Prac skill:0.6

Comm skill:0.6



Yes

No

No

Inter Active	Prac Skills	Job offer
No	Very Good	No
No	Average	Yes
No	Good	No

No	Very Good	Yes
No	Good	No
No	Very Good	Yes

Gain overall

$$-\frac{3}{6}\log log\ 2\left(\frac{3}{6}\right)-\frac{3}{6}\log log\ 2\left(\frac{3}{6}\right)$$

1

Gain Inter Active

$$\frac{6}{6}\left[-\frac{3}{6}\log log\ 2\left(\frac{3}{6}\right)-\frac{3}{6}\log log\ 2\left(\frac{3}{6}\right)\right]$$

1

1-1

0

Gain Prac skill

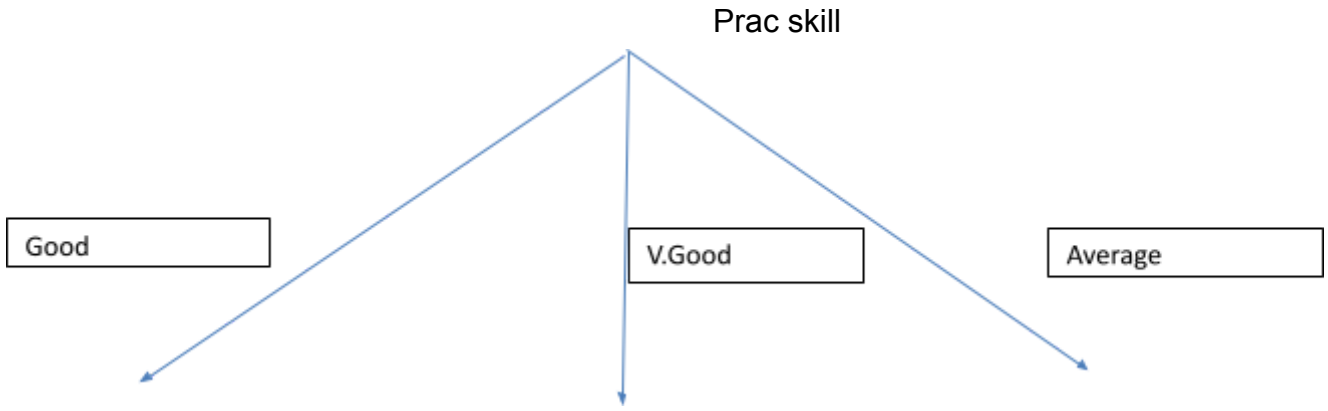
$$\frac{3}{6}\left[-\frac{1}{3}\log log\ 2\left(\frac{1}{3}\right)-\frac{2}{3}\log log\ 2\left(\frac{2}{3}\right)\right]+\frac{2}{6}\times\left[-\frac{2}{2}\log log\ 2\left(\frac{2}{2}\right)\right]+\frac{1}{6}\times\left[-\frac{1}{1}\log log\ 2\left(\frac{1}{1}\right)\right]$$

0.46

1-0.46

0.54

Prac skill has highest gain



Inter Active	Job Offer
No	No
No	Yes
No	Yes

Gain overall

$$\frac{2}{3} \times \left[-\frac{2}{3} \log log\ 2 \left(\frac{2}{3}\right) - \frac{1}{3} \log log\ 2 \left(\frac{1}{3}\right) \right]$$

0.92

Gain Inter Active

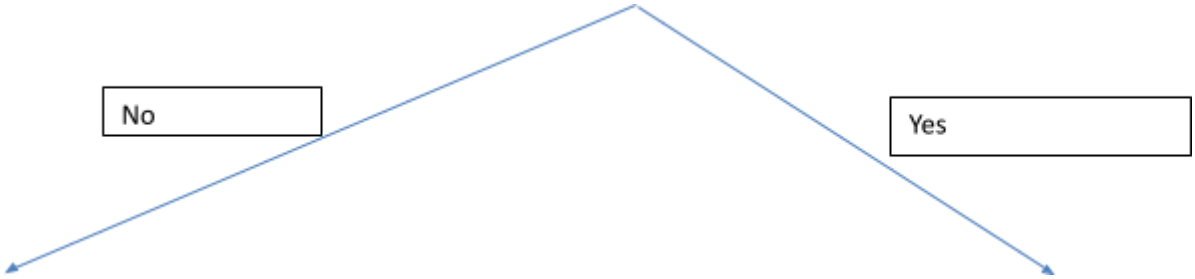
$$\frac{3}{3} \times \left[-\frac{2}{3} \log log\ 2 \left(\frac{2}{3}\right) - \frac{1}{3} \log log\ 2 \left(\frac{1}{3}\right) \right]$$

0.92

0.92-0.92

0

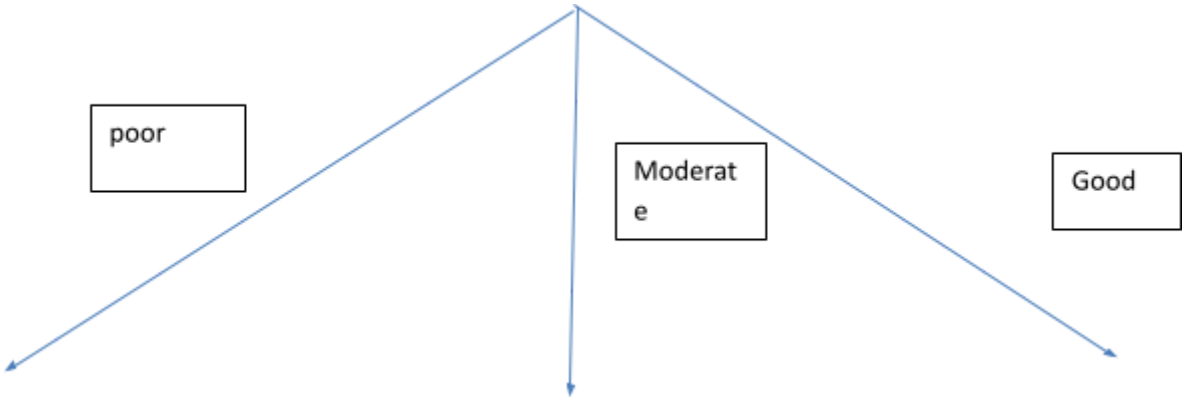
Inter Active



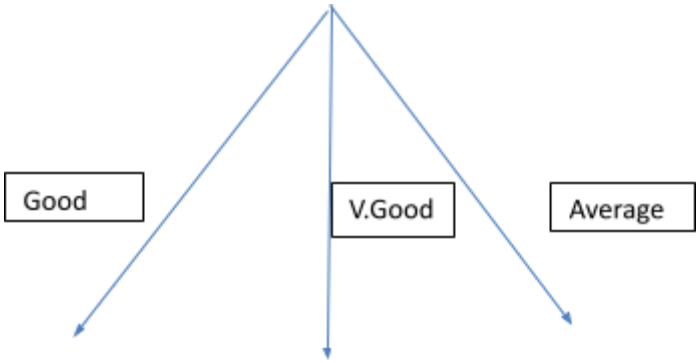
No/Yes

The Final Graph Will be look like this

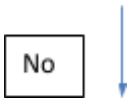
Comm skill



Practical skill



Inter Active



Yes/NO

GINI

Job Offer 2 type
Yes/No

Gini Inter Active

$$\text{Gini (Yes): } 1 - \left(\frac{0}{3}\right)^2 + \left(\frac{3}{3}\right)^2$$

=1-1=0

$$\text{Gini (No):} 1 - \left(\frac{3}{7}\right)^2 + \left(\frac{4}{7}\right)^2$$

=o-4892

$$\text{Gini Weighted Avg: } \left(\frac{3}{10}\right)\times 0 + \left(\frac{7}{10}\right)\times 0.4897$$

=0.34279

Gini Practical Skill

$$\text{Gini (Very Good) = } 1 - \left(\frac{2}{4}\right)^2 \times \left(\frac{2}{4}\right)^2$$

=0.5

$$\text{Gini (Good)=} 1 - \left(\frac{4}{4}\right)^2 \times \left(\frac{0}{4}\right)^2$$

=0

$$\text{Gini (Average)}=1-\left(\frac{1}{2}\right)^2\times\left(\frac{1}{2}\right)^2$$

=0.5

$$\text{Gini (Weighted)}=\left(\frac{1}{2}\right)\times 0.5+\left(\frac{4}{10}\right)\times 0+\left(\frac{2}{10}\right)\times 0.5$$

= 0.3

Gini Comman Skill

$$\text{Gini (Poor)}=1-\left(\frac{3}{6}\right)^2+\left(\frac{3}{6}\right)^2$$

= 0.5

$$\text{Gini(Moderate)}=1-\left(\frac{0}{3}\right)^2+\left(\frac{3}{3}\right)^2$$

= 0

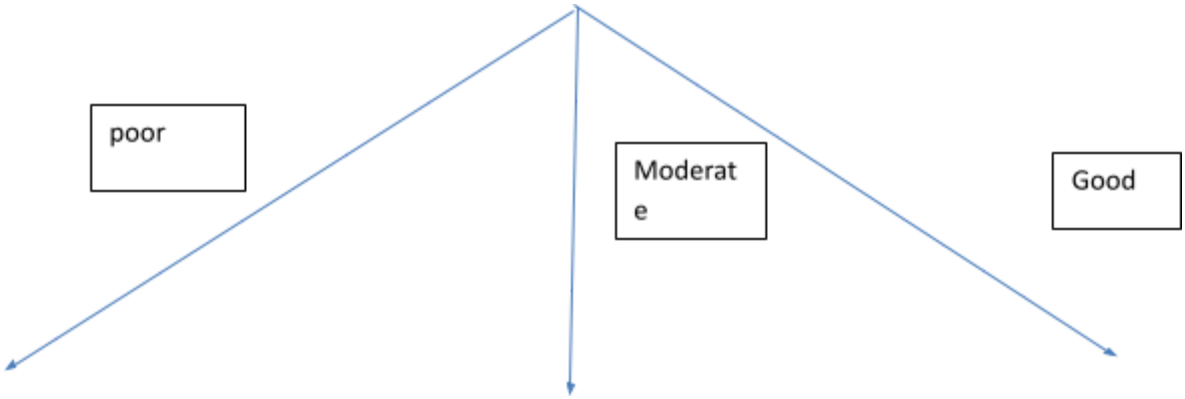
$$\text{Gini(Good)}=1-\left(\frac{0}{1}\right)^2+\left(\frac{1}{1}\right)^2$$

= 0

$$\text{Gini (Weighted)}=\left(\frac{6}{10}\right)\times 0.5+\left(\frac{3}{10}\right)\times 0+\left(\frac{1}{10}\right)\times 0.5$$

= 0.3

Comman Skill



Yes No

No

Inter Active	Prac Skills	Job offer
No	Very Good	No
No	Average	Yes
No	Good	No
No	Very Good	Yes
No	Good	No
No	Very Good	Yes

2nd Iteration

$$\text{Gini (No)} = 1 - \left(\frac{3}{6}\right)^2 + \left(\frac{3}{6}\right)^2$$

= 0.5

$$\text{Gini (Weighted)} = \left(\frac{3}{6}\right)^{\times} 0.5$$

= 0.5

Gini Practical Skill

$$\text{Gini (Very Good)} = 1 - \left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2$$

=0.44

$$\text{Gini (Good)} = 1 - \left(\frac{0}{2}\right)^2 + \left(\frac{2}{2}\right)^2$$

= 0

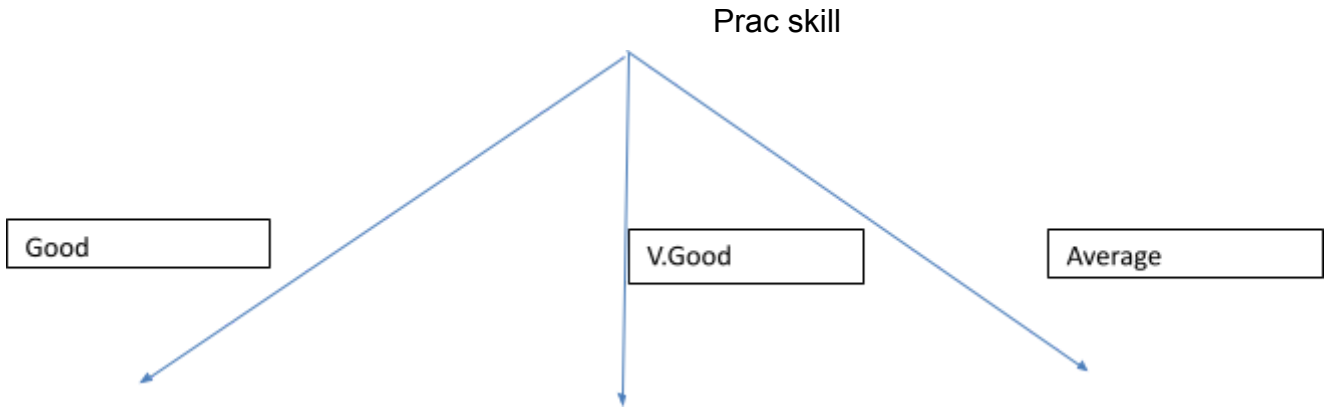
$$\text{Gini (Average)}=1-\left(\frac{1}{1}\right)^2+\left(\frac{0}{1}\right)^2$$

=0

$$\text{Gini (Weighted)}=\left(\frac{1}{6}\right)\times 0.44+\left(\frac{2}{6}\right)\times 0+\left(\frac{1}{6}\right)\times 0.44$$

= 0.22

Prac skill has highest gain



Inter Active	Job Offer
No	No
No	Yes
No	Yes

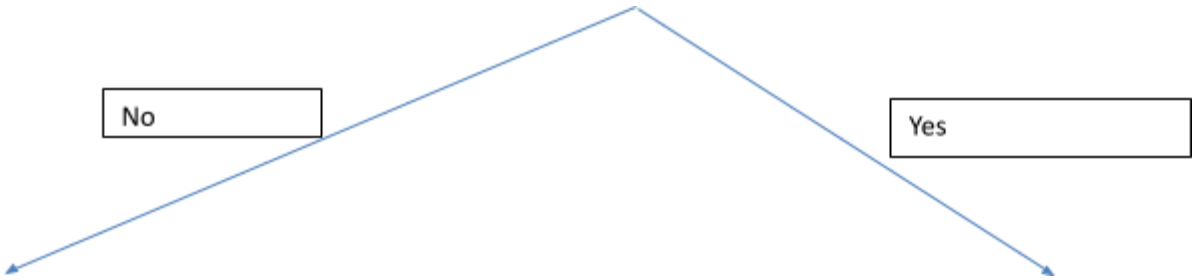
3rd Iteration

Gini Inter Active

$$\text{Gini (No)} = 1-\left(\frac{2}{3}\right)^2+\left(\frac{1}{3}\right)^2$$

0.44

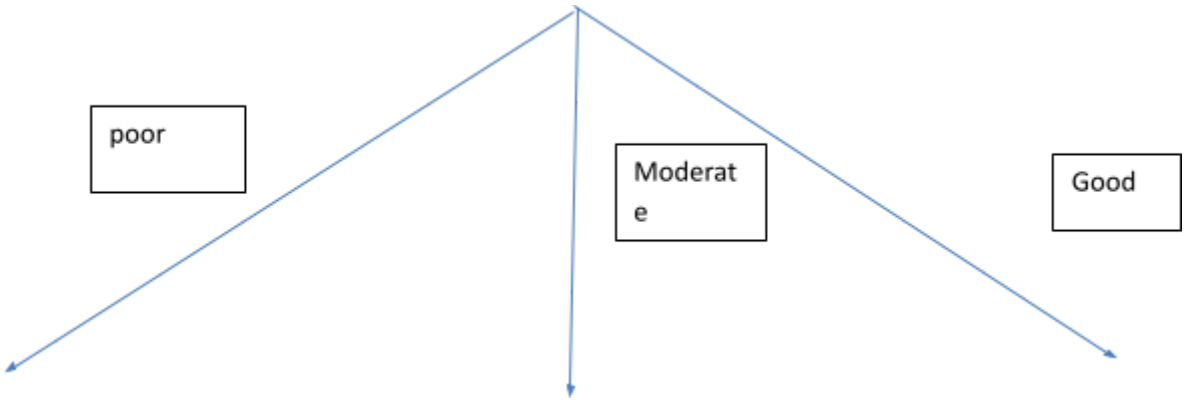
Inter Active



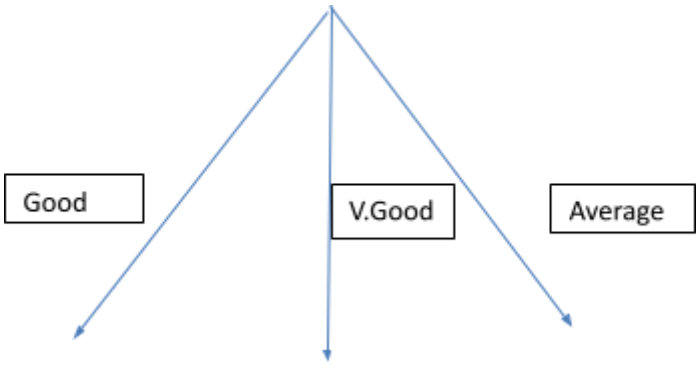
No/Yes

The Final Graph Will be look like this

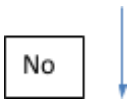
Comm skill



Practical skill



Inter Active



Yes/NO

Question N0 2

Instance	A1	A2	A3	Classification
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	Yes
4	False	Cool	Normal	Yes
5	False	Cool	Normal	Yes
6	True	Cool	High	No
7	True	Hot	High	No
8	True	Hot	Normal	Yes
9	False	Cool	Normal	Yes
10	False	Cool	High	Yes

Step 01: Entropy Of Entire Data

S{+6,-4}=

$-\frac{6}{10}\log_2\frac{6}{10}-\frac{4}{10}\log_2\frac{4}{10}$ $-\frac{6}{10}\log_2\frac{6}{10}-\frac{4}{10}\log_2\frac{4}{10}$

= 0.9709

Step 02: Entropy of all Attributes.

A1 :

$\frac{5}{10}*[-\frac{4}{5}\log_2\frac{4}{5}-\frac{1}{5}\log_2\frac{1}{5}]+\frac{5}{10}*[-\frac{5}{5}\log_2\frac{5}{5}]$ $\frac{5}{10}*[-\frac{4}{5}\log_2\frac{4}{5}-\frac{1}{5}\log_2\frac{1}{5}]+\frac{5}{10}*[-\frac{5}{5}\log_2\frac{5}{5}]$

= 0.3609

GAIN(A1)=0.9709-0.3609=0.61

A2 :

$510*[-35\log_2 35-25\log_2 25]+510[-15\log_2 15-45\log_2 45]$ $510*[-35log235-25log225]+510[-15log215-45log245]$

0.485 +0.3609= 0.845

GAIN(A2)= 0.845-0.9709=0.125

A2 :

$610*[-46\log_2 46-26\log_2 26\ ]+410[-44\log_2 44]$ $610*-46log246-26log226\ +410[-44log244]$

=0.5509

GAIN(A3)= 0.5509-0.9709=0.41

GAIN(A1)= 0.61

GAIN(A2)= 0.125

GAIN(A3)= 0.41

A1

/ \Yes

True False

A2	A3	Classification
Hot	High	No
Hot	High	No
Cool	High	No
Hot	High	No
Hot	Normal	Yes

Step 01: Entropy Of Entire Data.

S{1,-4}=

$-\frac{4}{5}\log_2\frac{4}{5}-\frac{1}{5}\log_2\frac{1}{5}$ $-\frac{4}{5}\log_2\frac{4}{5}-\frac{1}{5}\log_2\frac{1}{5}$

= 0.7219

Step 02: Entropy of all Attributes.

A2=

$45*[-\frac{3}{4}\log_2\frac{3}{4}-\frac{1}{4}\log_2\frac{1}{4}]+15*[-\frac{1}{1}\log_2\frac{1}{1}]$ $45*[-\frac{3}{4}\log_2\frac{3}{4}-\frac{1}{4}\log_2\frac{1}{4}]+15*[-\frac{1}{1}\log_2\frac{1}{1}]$

0.6490+0 = 0.6490

GAIN(A2)= 0.7219-0.6490 =0.0729

A3=

$45*[\frac{1}{4}\log_2\frac{1}{4}]+15[\frac{1}{1}\log_2\frac{1}{1}]$ $45*[\frac{1}{4}\log_2\frac{1}{4}]+15[\frac{1}{1}\log_2\frac{1}{1}]$

= 0.4

GAIN(A3)= 0.4-0.7219 = 0.3219

GAIN(A2)= 0.0729

GAIN(A3)= 0.3219

A3

/ \No

Hot Cool

A3	Classification
High	No
High	No

High	No
Normal	Yes

Step 01: Entropy Of Entire Data.

S{1,-3}=

$-\frac{3}{4}\log_2\frac{3}{4}-\frac{1}{4}\log_2\frac{1}{4}$

= 0.8112

Step 02: Entropy of all Attributes.

A3=

$\frac{3}{4}*[-\frac{3}{3}\log_2\frac{3}{3}]+\frac{1}{4}[-\frac{1}{1}\log_2\frac{1}{1}]$ $\frac{3}{4}*-\frac{3}{3}\log_2\frac{3}{3}+\frac{1}{4}[-\frac{1}{1}\log_2\frac{1}{1}]$

= 0

GAIN(A3)= 0.8112-0= 0.8112

A1

/ \Yes

True False

A2

/ \No

Hot Cool

A3

no / \yes

High Normal

Gini

Same Data

Instance	A1	A2	A3	Classification
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	Yes
4	False	Cool	Normal	Yes
5	False	Cool	Normal	Yes
6	True	Cool	High	No
7	True	Hot	High	No
8	True	Hot	Normal	Yes
9	False	Cool	Normal	Yes
10	False	Cool	High	Yes

Gini A1:

Gini A1(true) =
$1-(4/5)_2+(15)_2 1-(4/5)2+(15)2$
= 0.4

Gini A1(false) =
$1-(5/5)_2 1-(5/5)2$
= 0

Gini A1(weighted avg) =
$510*0.4+510*0 510*0.4+510*0$

= 0.2

Gini A2:

Gini A2(hot) =
$1-(2/5)^2+(35)^21-(2/5)2+(35)2$
= 1.2

Gini A2(cool) =
$1-(4/5)^2+(15)^21-(4/5)2+(15)2$
= 0.4

Gini A2(weighted avg) =
$510*1.2+510*0.4510*1.2+510*0.4$
= 0.8

Gini A3:

Gini A3(high) =
$1-(4/6)^2+(26)^21-(4/6)2+(26)2$
= 0.666

Gini A3(normal) =
$1-(44)^21-(44)2$
= 0

Gini A3(weighted avg) =
$610*0.666+410*0610*0.666+410*0$
= 0.399

A1

/ \yes

True false

A2	A3	Classification
Hot	High	No

Hot	High	No
Cool	High	No
Hot	High	No
Hot	Normal	Yes

Gini A2:

Gini A2(Hot) =
$1-(35)2+(15)21-(35)2+(15)2$
= 0.68

Gini A2(cool) =
$1-(1/5)21-(1/5)2$
= 0.96

Gini A2(weighted avg) =
$45*0.68+15*0.9645*0.68+15*0.96$
= 0.73

Gini A3:

Gini A3(hot) =
$1-(4/5)21-(4/5)2$
= 0.36

Gini A3(cool) =
$1-(1/5)21-(1/5)2$
= 0.96

Gini A3(weighted avg) =
$45*0.36+15*0.9645*0.36+15*0.96$
= 0.48

A3

/ \No

Hot Cool

A3	Classification
High	No
High	No
High	No
Normal	Yes

Gini A3:

Gini A3(high) =
$1-(3/3)^2 1-(3/3)2$
= 0

Gini A3(normal) =
$1-(1/1)^2 1-(1/1)2$
= 0

Gini A3(weighted avg) =
$34*0+14*0 34*0+14*0$
= 0

A1

/ \Yes

True False

A2

/ \No

Hot Cool

A3

no / \yes

High Normal

Question NO 3

Column1	Column2	Column3	Column4	Column5
Weekend	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Tennis
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Tennis
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

D1 = Cinema = 6, D2 = Tennis = 4

Gain (Overall): -

−6/10.log2(6/10)−4/10.log2(4/10) = 0.97

Calculate the Entropy for the "Weather"

$Gain_{sunny}$ = Entropy=−(1/3 ·log2 (1/3)+2/3 ·log2 (2/3)) = 0.918
$Gain_{windy}$ = Entropy=−(3/4 ·log2 (3/4)+1/4 ·log2 (1/4)) = 0.811
$Gain_{rainy}$ = Entropy=−(2/3 ·log2 (2/3)+1/3 ·log2 (1/3)) = 0.918

Calculate the weighted for the "Weather"

$Gain_{weighted}$ =3/10*0.918+4/10*0.811+3/10*0.918 = 0.875

Information Gain for "Weather":

IGweather = 0.97-0.875 = 0.095

Calculate the Entropy for the "Parents"

$Gain_{Yes}$ = Entropy=−(5/5 ·log2 (5/5) = 0
$Gain_{No}$ = Entropy=−(4/5 ·log2 (4/5)+1/5 ·log2 (1/5)) = 0.722

Calculate the weighted for the "Parents"

$Gain_{weighted}$ =5/10*0+5/10*0.722 = 0.875 = 0.361

Information Gain for "Parents":

IGparents = 0.97-0.722 = 0.248

Calculate the Entropy for the "Money"

$Gain_{Poor}$ = Entropy=−(3/3 ·log2 (3/3) = 0
$Gain_{Rich}$ = Entropy=−(3/7 ·log2 (3/7)+4/7 ·log2 (4/7)) = 0.985

Calculate the weighted for the "Money"

$Gain_{weighted}$ =3/10*0+7/10*0.985 = 0.689

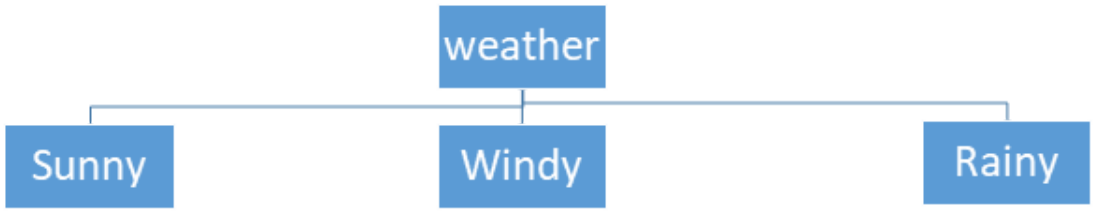
Information Gain for "Money":

IGmoney = 0.97-0.689 = 0.281

Gains: -
Gain(Weather) = 0.095
Gain(Parents) = 0.248
Gain(Money) = 0.281

Hence weather attribute has the highest gain so it's a root node

Column1	Column2	Column3
Parents	Money	Decision
Yes	Rich	Cinema
No	Poor	Cinema
No	Rich	Tennis
Yes	Rich	Cinema



Column1	Column2	Column3
Parents	Money	Decision
Yes	Rich	Cinema
No	Rich	Tennis
No	Rich	Tennis
Column1	**Column2**	**Column3**
Parents	Money	Decision
Yes	Poor	Cinema
No	Rich	Tennis
Yes	Poor	Cinema

For Sunny: -

Gain (Overall): -

$-1/3.\log2(1/3)-2/3.\log2(2/3) = 0.918$

Calculate the Gain for the "Parents"

$Gain_{Yes}$ = Entropy=$-(1/1 \cdot \log2\ (1/1\ ) = 0$
$Gain_{No}$ = Entropy=$-(2/2 \cdot \log2\ (2/2) = 0$

Calculate the weighted for the "Parents"

$Gain_{weighted} = 1/3*0+2/3*0 = 0$

Information Gain for "Parents":

IGmoney = 0.918-0 = 0.918

Calculate the Gain for the "Money"

$Gain_{Rich}$ = Entropy=$-(1/3 \cdot \log2\ (1/3)+2/3 \cdot \log2\ (2/3)) = 0.918$

Calculate the weighted for the "Money"

$Gain_{weighted} = 0/3*0+3/3*0.918 = 0.918$
Information Gain for "Money":

IGmoney = 0.918-0.918 = 0

Hence Parents attribute has the highest gain so it's a root node

Column1	Column2
Money	Decision
Rich	Tennis
Rich	Tennis
Column1	**Column2**

Money	Decision
Rich	Cinema

For Windy: -

Gain (Overall): -

−3/4.log2(3/)−1/4.log2(1/4) = 0.811

Calculate the Gain for the "Parents"

$Gain_{Yes}$ = Entropy=−(2/2 ·log2 (2/2) = 0
$Gain_{No}$ = Entropy=−(1/2 ·log2 (1/2)+(1/2.log2(1/2)) = 1

Calculate the weighted for the "Parents"

$Gain_{weighted}$ =2/4*0+2/4*1 = 0.5

Information Gain for "Parents":

IGmoney = 0.811-0.5 = 0.311

Calculate the Gain for the "Money"

$Gain_{Rich}$ = Entropy=−(2/3 ·log2 (2/3)+1/3 ·log2 (1/3)) = 0.918
$Gain_{Poor}$ = Entropy=−(1/1 ·log2 (1/1) = 0

Calculate the weighted for the "Money"

$Gain_{weighted}$ =1/4*0+3/4*0.918 = 0.688

Information Gain for "Money":

IGmoney = 0.811-0.688 = 0.123

Hence Money attribute has the highest gain so it's a root node

Column1	Column2
Parents	Decision
Yes	Cinema
No	Tennis
Yes	Cinema
Column1	**Column2**
Parents	Decision
No	Cinema

For Rainy: -

Gain (Overall): -

$-2/3.\log2(2/3)-1/3.\log2(1/3) = 0.918$

Calculate the Gain for the "Parents"

$Gain_{No}$ = Entropy=$-(1/1 \cdot \log2\ (1/1\ ) = 0$
$Gain_{Yes}$ = Entropy=$-(2/2 \cdot \log2\ (2/2) = 0$

Calculate the weighted for the "Parents"

$Gain_{weighted}$ =2/3*0+1/3*0 = 0

Information Gain for "Parents":

IGmoney = 0.918-0 = 0.918

Calculate the Gain for the "Money"

$Gain_{Poor}$ = Entropy=−(1/2 ·log2 (1/2)+1/2 ·log2 (1/2)) = 1
$Gain_{Rich}$ = Entropy=−(1/1 ·log2 (1/1) = 0

Calculate the weighted for the "Money"

$Gain_{weighted}$ =1/3*0+2/3*1 = 0.66

Information Gain for "Money":

IGmoney = 0.918-0.66 = 0.258

Hence Parents attribute has the highest gain so it's a root node

Column1	Column2
Money	Decision
Poor	Cinema
Poor	Cinema
Column1	**Column2**
Money	Decision
Rich	Tennis

GINI: -

D1 = Cinema = 6, D2 = Tennis = 4

GINI: Weather: -

$GINI_{sunny}$ = 1-((1/3)^2+(2/3)^2) = 0.44
$GINI_{Windy}$ = 1-((3/4)^2+(1/4)^2)= 0.375

$GINI_{Rainy} = 1-((2/3)^2+(1/3)^2)= 0.44$

Weighted for the "Weather"

$GINI_{Weighted} = 3/10 * 0.44 + 4/10 * 0.375 + 3/10 * 0.44 = 0.414$

GINI: Parents: -

$GINI_{Yes} = 1-((5/5)^2+(0/5)^2) = 0$
$GINI_{No} = 1-((4/5)^2+(1/5)^2)= 0.32$

Weighted for the "Parents"

$GINI_{Weighted} = 5/10 * 0 + 5/10 * 0.32 = 0.16$

GINI: Money: -

$GINI_{Rich} = 1-((3/7)^2+(4/7)^2) = 0.489$
$GINI_{Poor} = 1-((3/3)^2+(0/3)^2)= 0$

Weighted for the "Money"

$GINI_{Weighted} = 7/10 * 0.489 + 3/10 * 0 = 0.343$

In GINI, we take lowest attribute so Parents is a root node.

Column1	Column2	Column3
Weather	Money	Decision
Sunny	Rich	Cinema
Windy	Rich	Cinema
Rainy	Poor	Cinema
Rainy	Poor	Cinema
Windy	Rich	Cinema
Column1	**Column2**	**Column3**

Weather	Money	Decision
Sunny	Rich	Tennis
Rainy	Rich	Tennis
Windy	Poor	Cinema
Windy	Rich	Tennis
Sunny	Rich	Tennis

In Parents (Yes) all decisions are cinema so we are not splitting it further now split Parents (No) further to calculate decision.

GINI: Weather: -

$GINI_{sunny} = 1-((2/2)^2+(0/2)^2) = 0$
$GINI_{Windy} = 1-((1/2)^2+(1/2)^2)= 0.5$
$GINI_{Rainy} = 1-((1/1)^2+(0/1)^2)= 0$

Weighted for the "Weather"

$GINI_{Weighted} = 2/5 * 0 + 2/5 * 0.5 + 1/5 * 0 = 0.2$

GINI: Money: -

$GINI_{Rich} = 1-((4/4)^2+(0/4)^2) = 0$
$GINI_{Poor} = 1-((1/1)^2+(0/1)^2)= 0$

Weighted for the "Money"

$GINI_{Weighted} = 4/5 * 0 + 1/5 * 0 = 0$

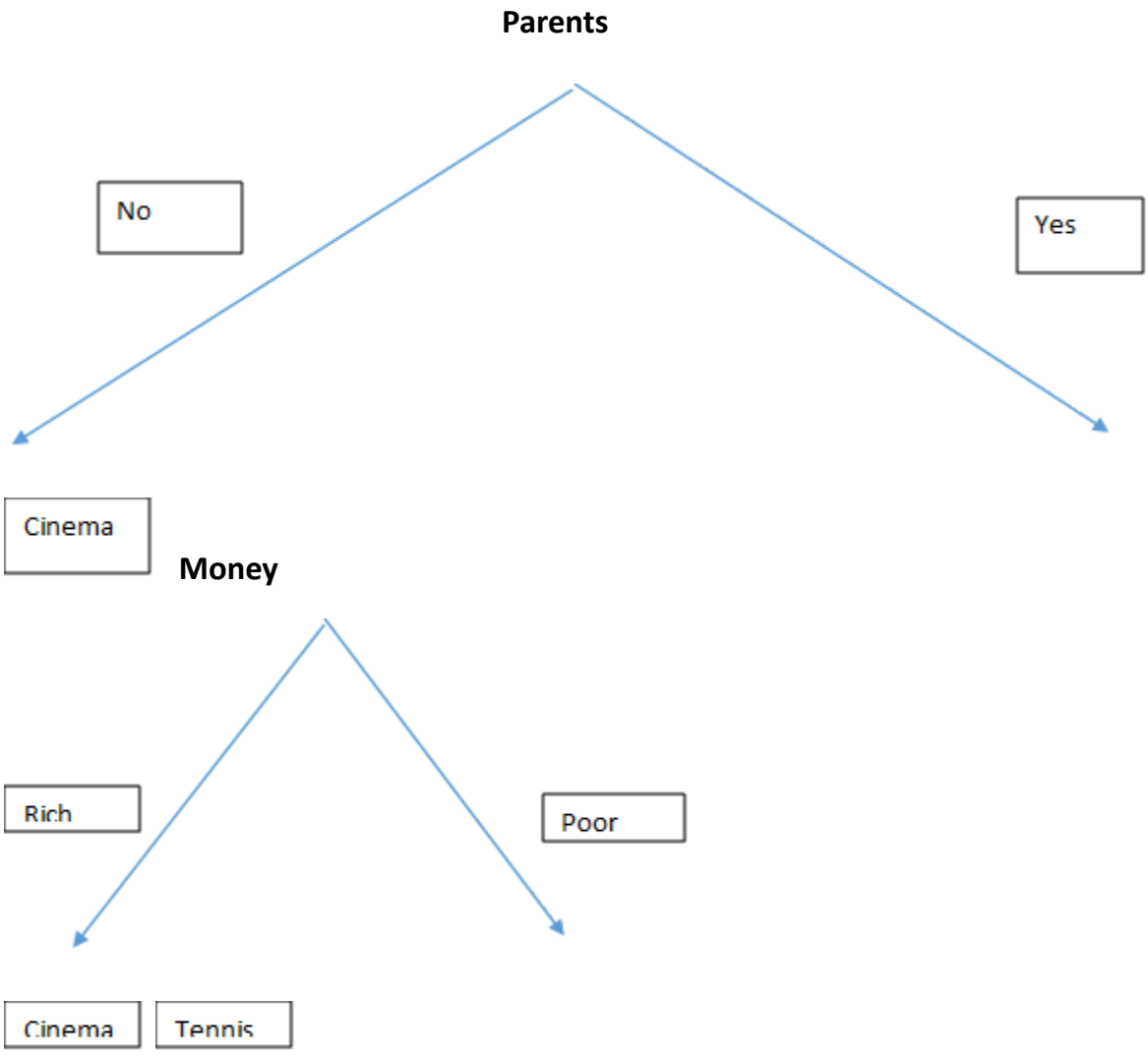
In GINI, we take lowest attribute so Money is a root node.

Weather	Decision
Sunny	Tennis

Rainy	Tennis
Windy	Tennis
Sunny	Tennis
Column1	**Column2**
Weather	Decision
Windy	Cinema

Now It cannot be split more so now draw a decision tree of GINI algorithm.

The final Graph will be look like this



Unsupervised

Question NO 1

Point	Distance from Centroid 1 (2,6)	Distance from Centroid 2 (5,10)	Distance from Centroid 3 (6,11)	Assigned Cluster
A1 (2,10)	4	3	4.123106	Cluster 2
A2 (2,6)	0	5	6.403124	Cluster 1
A3 (11,11)	10.29563	6.082763	5	Cluster 3
A4 (6,9)	5	1.414214	2	Cluster 2
A5 (6,4)	4.472136	6.082763	7	Cluster 1
A6 (1,2)	4.123106	8.944272	10.29563	Cluster 1
A7 (5,10)	5	0	1.414214	Cluster 2
A8 (4,9)	3.605551	1.414214	2.828427	Cluster 2
A9 (10,12)	10	5.385165	4.123106	Cluster 3
A10 (7,5)	5.09902	5.385165	6.082763	Cluster 1
A11 (9,11)	8.602325	4.123106	3	Cluster 3
A12 (4,6)	2	4.123106	5.385165	Cluster 1
A13 (3,10)	4.123106	2	3.162278	Cluster 2
A14 (3,8)	2.236068	2.828427	4.242641	Cluster 1
A15 (6,11)	6.403124	1.414214	0	Cluster 3

Centroid 1=(2,6) is associated with cluster 1.

Centroid 2=(5,10) is associated with cluster 2.

Centroid 3=(6,11) is associated with cluster 3

In cluster 1, we have 6 points i.e. A2 (2,6), A5 (6,4), A6 (1,2), A10 (7,5), A12 (4,6), A14 (3,8). To calculate the new centroid for cluster 1, we will find the mean of the x and y coordinates of each point in the cluster. Hence, the new centroid for cluster 1 is (3.833, 5.167).

In cluster 2, we have 5 points i.e. A1 (2,10), A4 (6,9), A7 (5,10) , A8 (4,9), and A13 (3,10). Hence, the new centroid for cluster 2 is (4, 9.6)

In cluster 3, we have 4 points i.e. A3 (11,11), A9 (10,12), A11 (9,11), and A15 (6,11). Hence, the new centroid for cluster 3 is (9, 11.25).

Question NO 2

Point	Distance from Centroid 1 (3.833, 5.167)	Distance from centroid 2 (4, 9.6)	Distance from centroid 3 (9, 11.25)	Assigned Cluster
A1 (2,10)	5.169	2.040	7.111	Cluster 2
A2 (2,6)	2.013	4.118	8.750	Cluster 1
A3 (11,11)	9.241	7.139	2.016	Cluster 3
A4 (6,9)	4.403	2.088	3.750	Cluster 2
A5 (6,4)	2.461	5.946	7.846	Cluster 1
A6 (1,2)	4.249	8.171	12.230	Cluster 1
A7 (5,10)	4.972	1.077	4.191	Cluster 2
A8 (4,9)	3.837	0.600	5.483	Cluster 2
A9 (10,12)	9.204	6.462	1.250	Cluster 3
A10 (7,5)	3.171	5.492	6.562	Cluster 1
A11 (9,11)	7.792	5.192	0.250	Cluster 3
A12 (4,6)	0.850	3.600	7.250	Cluster 1
A13 (3,10)	4.904	1.077	6.129	Cluster 2
A14 (3,8)	2.953	1.887	6.824	Cluster 2
A15 (6,11)	6.223	2.441	3.010	Cluster 2

In cluster 1, we have 5 points i.e. A2 (2,6), A5 (6,4), A6 (1,2), A10 (7,5), and A12 (4,6). To calculate the new centroid for cluster 1, we will find the mean of the x and y coordinates of each point in the cluster. Hence, the new centroid for cluster 1 is (4, 4.6).

In cluster 2, we have 7 points i.e. A1 (2,10), A4 (6,9), A7 (5,10) , A8 (4,9), A13 (3,10), A14 (3,8), and A15 (6,11). Hence, the new centroid for cluster 2 is (4.143, 9.571)

In cluster 3, we have 3 points i.e. A3 (11,11), A9 (10,12), and A11 (9,11). Hence, the new centroid for cluster 3 is (10, 11.333).

Point	Distance from Centroid 1 (4, 4.6)	Distance from centroid 2 (4.143, 9.571)	Distance from centroid 3 (10, 11.333)	Assigned Cluster
A1 (2,10)	5.758	2.186	8.110	Cluster 2
A2 (2,6)	2.441	4.165	9.615	Cluster 1
A3 (11,11)	9.485	7.004	1.054	Cluster 3
A4 (6,9)	4.833	1.943	4.631	Cluster 2
A5 (6,4)	2.088	5.872	8.353	Cluster 1
A6 (1,2)	3.970	8.197	12.966	Cluster 1
A7 (5,10)	5.492	0.958	5.175	Cluster 2
A8 (4,9)	4.400	0.589	6.438	Cluster 2
A9 (10,12)	9.527	6.341	0.667	Cluster 3
A10 (7,5)	3.027	5.390	7.008	Cluster 1
A11 (9,11)	8.122	5.063	1.054	Cluster 3
A12 (4,6)	1.400	3.574	8.028	Cluster 1
A13 (3,10)	5.492	1.221	7.126	Cluster 2
A14 (3,8)	3.544	1.943	7.753	Cluster 2
A15 (6,11)	6.705	2.343	4.014	Cluster 2

In cluster 1, we have 5 points i.e. A2 (2,6), A5 (6,4), A6 (1,2), A10 (7,5), and A12 (4,6). To calculate the new centroid for cluster 1, we will find the mean of the x and y coordinates of each point in the cluster. Hence, the new centroid for cluster 1 is (4, 4.6).

In cluster 2, we have 7 points i.e. A1 (2,10), A4 (6,9), A7 (5,10) , A8 (4,9), A13 (3,10), A14 (3,8), and A15 (6,11). Hence, the new centroid for cluster 2 is (4.143, 9.571)

In cluster 3, we have 3 points i.e. A3 (11,11), A9 (10,12), and A11 (9,11). Hence, the new centroid for cluster 3 is (10, 11.333).

Date

K-Mean Algorithm:-

1st Data set:-

$A_1(2,10), A_2(2,5), A_3(4,8)$
$B_1(3,1), B_2(2,6), B_3(6,4)$
$C_1(1,2), C_2(4,9)$.

	Data points		Distance				cluster	cluster (New
			2 | 10	3 | 1	1	2		
A_1	2	10	0	9.2	9.9		1	
A_2	2	5	8.6	4.8	4.8		2	
A_3	4	8	4.8	6.4	8.86		1	
B_1	3	1	9.6	0	2.2		2	
B_2	2	6	8	5.4	5.9		2	
B_3	6	4	7.2	4.2	6.8		2	
C_1	1	2	9.9	3.3	0		3	
C_2	4	9	2.6	8.5	9.85		1	

New centroid.

$A = (2+4+4/3, 10+8+9/3)$
$A = (3.33, 9)$

$B = (2+3+2+6/4, 5+1+6+4/4)$
$(3.25, 4)$
$C = (1, 2)$

Date

Data points			Distance 3.33	9	3.25	4	1	2	cluster	New cluster
A_1	2	10	5.1		9.5		9.9		1	1
A_2	2	5	7.9		3.9		4.8		2	2
A_3	4	8	3.4		6.5		8.6		1	1
B_1	3	1	9.6		4		2.3		2	3
B_2	2	6	7.2		5.15		5.3		2	2
B_3	6	4	6.3		5		7.4		2	2
C_1	1	2	9.3		4.6		0		3	3
C_2	4	9	3.1		7.7		8.9		1	1

New centroid

$A = (2+4+4/3 \; , \; 10+8+9/3)$

$A = (33, 9)$

$B = (2+2+6 \; / \; 3 \; , \; 5+6+4 \; / \; 3)$

$B = (3.3, 5)$

$C = (3+1 \; / \; 2 \; , \; 2+1 \; / \; 2)$

$C = (2, 1.5)$

New centroids are

$(3.3, 9) \quad (3.3, 5) \quad (2, 1.5)$

Data point			Distance 3.3	9	3.3	5	2	1.5	cluster	New Cluster
A_1	2	10	5.1		9.6		9.8		1	1
A_2	2	5	7.4		2.6		4.7		2	2
A_3	4	8	3.4		5.8		8.5		1	1
B_1	3	1	9.6		5.0		2.6		3	3
B_2	2	6	7.2		4.2		5.8		2	2
B_3	6	4	6.3		5.8		6.7		2	2
C_1	1	2	9.3		5.5		2.1		3	3
C_2	4	9	3.1		7.81		9.5		1	1