ML ASSIGNMENT

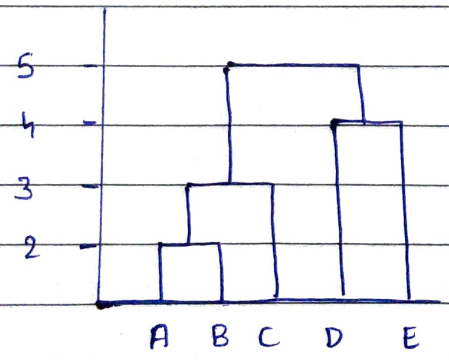
18DCS007

Q1. Applying single link clustering

	A	B	C	D	E
A	0				
B	[2]	9			
C	6	9	0		
D	10	8	7	0	
E	9	5	4	0	

	AB	C	D	E
AB	0			
C	[3]	0		
D	9	7	0	
E	8	5	4	0

	ABC	DE
ABC	0	
DE	[5]	0



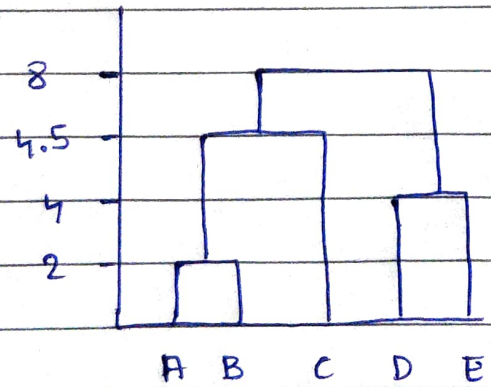
Miracle
Page
Date

Applying average link clustering

	A	B	C	D	E
A	0				
B	[2]	0			
C	6	3	0		
D	10	9	7	0	
E	9	8	5	4	0

	AB	C	D	E
AB	0			
C	4.5	0		
D	9.5	7	0	
E	8.5	5	[4]	0

	ABC	DE
ABC	0	
DE	[8]	0



Miracle
Page
Date

Q2. Computer applications for which ML is appropriate

(1) Stock prediction system - prediction system which has high uncertainity & large data

(2) Intrusion / Fraud Detection - pattern recognition plays imp. role

(3) Spam Filtering

→ Computer applications for which ML is inappropriate

(1) Shortest path finding - we already have great algorithm like A* and dijkstras algorithm

(2) Simple face detection - image processing is capable of it

(3) Rule based systems/games - We don't need AI to play tic-tac.

Q3. Min support = 30%, min confidence = 80%

ITEM	COUNT
Butter	4
Bread	5
Milk	4
Cheese	4
Biscuits	1
Clothes	3

Support threshold = $0.3 \times 7 = 2.1 \approx 3$

→ It shows biscuits does not meet min support so it is deleted

(4)

→ JOIN STEP :

ITEMS	COUNT
Butter, Bread	3
Butter, Milk	2
Butter, Cheese	3
Butter, Clothes	1
Bread, Milk	4
Bread, Cheese	2
Bread, Clothes	3
Milk, Cheese	1
Milk, Clothes	3
Cheese, Clothes	1

→ PRUNE STEP :

ITEMS	COUNT
Butter, Bread	3
Butter, Cheese	3
Bread, Milk	4
Bread, Clothes	3
Milk, Clothes	3

→ JOIN STEP :

ITEMS	COUNT	
Butter, Bread, Milk	2	
Butter, Bread, Cheese	2	
Butter, Bread, Clothes	0	
Bread, Milk, Cheese	1	
Bread, Milk, Clothes	3	✓ {Bread, Milk}, {Milk, Clothes}, {Bread, Clothes}
Milk, Cheese, Clothes	1	

∴ Frequent item set = {Bread, Milk, Clothes}

→ Association Rule :-

{ Bread, Clothes } → { Milk }

$$\text{confidence} = \frac{\text{support}\{\text{Bread, Clothes, Milk}\}}{\text{support}\{\text{Bread, Clothes}\}} = \frac{3}{3}$$

$$= 100\% > 80\%$$

{ Bread, Milk } → { Clothes }

$$\text{confidence} = \frac{\text{support}\{\text{Bread, Clothes, Milk}\}}{\text{support}\{\text{Bread, Milk}\}} = \frac{3}{4}$$

$$= 75\% < 80\%$$

{ Clothes, Milk } → { Bread }

$$\text{confidence} = \frac{\text{support}\{\text{Bread, Clothes, Milk}\}}{\text{support}\{\text{clothes, milk}\}} = \frac{3}{3} = 100\% > 80\%$$

∴ Strong association rules -

{ Bread, Clothes } ⇒ { Milk }
{ Clothes, Milk } ⇒ { Bread }

Q4. In feed forward networks, inputs are fed to the network and transformed into an output.
They are used for classification.

In feed forward networks, predictions does not depend on the predictions in the previous moment.

But the situation is different in reocurred neural networks. The decision the reurrent neural net gets at this moment, depends on the decisions which the net got at the previous moment

Moreover, the inputs and outputs to a fied forward network should be two dimensional with shape [nos of examples input/output size]
And input and outputs for recurrement neural network should be three dimensional with shape [no. of example, input size/ time series length]

Q5.

x	2	3	5	7	9
y	4	5	7	10	15

→ We know linear regression hypothesis

$$h_\theta(x) = \theta_0 + \theta_1 x$$

→ Lets consider $y = a + bx$

→ Using method of least squares

$$\Sigma y = a\Sigma 1 + b\Sigma x$$

$$\Sigma xy = a\Sigma x + b\Sigma x \quad \text{(multiply with } x\text{)}$$

→ Calculating parameters, $\Sigma x = 26$ $\Sigma x^2 = 168$

$\Sigma y = 41$ $\Sigma xy = 263$

$\therefore 5a + 26b = 41$

$26a + 168b = 263$

→ By solving the equation, $a = 0.305$

$b = 1.518$

→ For $x = 8.3$, $y = 0.305 + 1.518$

$= 12.905$

$= 13$

Q6. $H(\text{Diabetic}) = -\Sigma P(x_i) \log_b P(x_i)$

$= -\frac{6}{9} \log_2\left(\frac{6}{9}\right) - \frac{3}{9} \log_2\left(\frac{3}{9}\right)$

$= -0.66(-0.599) - 0.33(-1.599)$

$= 0.92301$

Weight

/ \

<80 >=80

YES NO YES NO

1 4 2 2 4 (weight >= 80) = 1

$H(\text{weight}_{<80}) = \frac{-1}{5} \log_2\left(\frac{1}{5}\right) - \frac{4}{5} \log_2\left(\frac{4}{5}\right)$

$= 0.7219$

$IG(\text{Diabetic}, \text{Weight}) = H(\text{Diabetic}) - \frac{5}{9} \omega(\text{Weight} < 80) - \frac{4}{5} H(\text{Weight} >= 80)$

$= 0.92301 - \frac{5}{9}(0.7219) - \frac{4}{9}(1) = 0.25478$

FOOD INTAKE

LOW		MEDIUM		HIGH	
YES	NO	YES	NO	YES	NO
1	3	0	2	2	1
H(low)		H(medium)		H(high)	
=.81125		=0		=0.9230)	

IG(food intake)

$= 0.92301 - \frac{4}{9}(0.81125) - \frac{2}{9}(0) - \frac{3}{9}(0.92301)$

$= 0.25478$

EXERCISING

Never		Occassionally		Regulary	
YES	NO	YES	NO	YES	NO
3	2	0	2	0	2
H(never)		H(occassionaly)		H(regularly)	
=0.970		=0		=0	

IG (Exercising)

$= 0.97301 - \frac{5}{9}(0.970)$

$= 0.383616$

→ Exercise will be root node.

IG (Diabetic, exercising) is highest amonst all

Q7.

	C_1	C_2	C_3		
A1	0	3.6	5.38	→	C_1
A2	3.6	0	5.38	→	C_2
A3	5.38	5.38	0	→	C_1
A4	5.1	2.23	8.06	→	C_2
A5	3.16	4.12	2.23	→	C_3
A6	5	2	5.1	→	C_2
A7	6.08	4.47	3.16	→	C_3
A8	8.24	6.08	5	→	C_3

$K = 3$

$C_1 = A_1$

$C_2 = A_2$

$C_3 = A_3$

→ New centroid value

$C_1 = (1,2)$

$C_2 = ((3+2+5)/3, (5+7+5)/3) = (3.33), (5.66)$

$C_3 = ((6+4+7+9)/4, (0+1+3+4)/4) = (6.5, 2)$

→ Taking distance from new centroids,

	C_1	C_2	C_3	
A1	0	4.33	5.5	→ C_1
A2	3.6	0.73	4.6	→ C_2
A3	5.38	6.25	2.06	→ C_3
A4	5.1	1.88	6.72	→ C_2
A5	3.16	4.70	2.69	→ C_4
A6	5	1.79	3.35	→ C_2
A7	6.08	4.53	1.118	→ C_3
A8	8.246	5.9	3.2	→ C_3

→ No changes, Cluster 1 = $C_1 = (1,2) = \{A1\}$

Cluster 2 = $C_2 = (3.33, 5.66) = \{A2, A4, A6\}$

Cwster 3 = $C_3 = (6.5, 2) = \{A3, A5, A7, A8\}$

Q8. K = 5, Manhattan Distance :

	8	9
1	(19)	[17]
2	(18)	[34]
3	21	35
4	26	38
5	(11)	[5]
6	(7)	[19]
7	(7)	[23]

(10)

→ The closest 5 neighbours of 8 are {1,2,5,6,7}
which means Tiger : 3
Leopard : 2
∴ 8 will be classified as tiger acc to KNN

→ The closest 5 neighbour of 9 are {1,2,5,6,9}
which means Tiger : 3
Leopard : 2
∴ 9 will be classified as tiger are to KNN