COMSATS UNIVERSITY Islamabad,

ATTOCK CAMPUS.

Mid Term Solution

'ML'

Submitted By:-

Fakiha Shahzael.

Reg#

Fa19-bcs-019

Section

'A'

Submitted To:-

Dr. Muhammad Sharif

Date:-

20th May, 22.

Q1):- Write short answer of the following question.

(A) Explain the Difference between classification, clustering and regression with the help of example?

Ans:- **Classification:-**

In machine learning, classification refers to a predictive modeling problem where a class label is predicted for a given example of input data.

Example:-

Given an example, classify if it is spam or not.

classification model include logistic regression, decision tree, multilayer perception.

Clustering:-

clustering is used to identify groups of similar objects in datasets with two or more variable quantities.

Example:-

- Agglomerative Approach.
- centroid based clustering.
- K-mean clusteering.

Regression:-

Regression is supervised machine learning technique which is used to predict continous values.

Example:-

- Linear Regression.
- Regression trees
- lasso regression.
- multivariate regression.

(B) Elaborate the few limitation C4.5 and ID3?

Ans:- **C4.5:-**

The limitation of C4.5 is its information entropy, it gives poor results for a larger distinct attributes.

C4.5 work with both Discrete and continous data.

C4.5 removed the restriction that featuee must be categorical by dynamically defining a discrete attribute!

ID3:-

- There exist a problem of multi-value bias in the process of attribute.
- It is not easy to calculate entropy.
- ID3 do not guarantee an optimal solution.

QNO3:- Apply linear regression technique using Batch gradient descent method ...?

Ans:- $\alpha = 0.05$

$\theta_0 = 0.1, \theta_1 = 0.2, \theta_2 = 0.1$

$h = \theta_0 + \theta_1 x_1 + \theta_2 x_2$

x1	x2	y(output)
1	2	0.1
4	5	0.2
1	4	0.4

$= 0.1 + 0.2(1) + 0.1(2) \Rightarrow 0.5$

$h2 = 0.1 + 0.2(4) + 0.1(5) \Rightarrow 1.4$

$h3 = 0.1 + 0.2(1) + 0.1(4) \Rightarrow 0.7$

x1	x2	h	h-y	(h-y)x1	(h-y)x2	$(h-y)^2$
1	2	0.5	0.4	0.4	0.8	0.16
4	5	1.4	1.2	4.8	6	1.44
1	4	0.7	0.3	0.3	1.2	0.09
			1.9	5.5	8	

For Cycle 1 :-

$$\theta_0 = \theta_0 - \frac{1}{n} \alpha \Sigma(h-y)$$

$$\theta_0 = 0.1 - \frac{1}{3}(0.05)(1.9)$$

$$\theta_0 = 0.1 - 0.0316$$

$$\theta_0 = 0.06$$

$$\theta_1 = \theta_1 - \alpha \frac{1}{n} \Sigma(h-y)x_1$$

$$= 0.2 - 0.05\left(\frac{1}{3}\right)(5.5)$$

$$= 0.2 - 0.0916$$

$$\theta_1 = 0.108$$

$$\theta_2 = \theta_2 - \alpha \frac{1}{n} \Sigma(h-y)x_2$$

$$\theta_2 = 0.1 - 0.05\left(\frac{1}{3}\right)(8)$$

$$= 0.1 - 0.133$$

$$\theta_2 = -0.03$$

(i) Cost value?

$$h1 = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$$= 0.06 + 0.108(1) + (-0.03)2$$

$$h1 = 0.208$$

$$h2 = 0.06 + 0.108(4) + (-0.03)5$$

$$h2 = -0.224$$

$h3 = 0.06 + 0.108(1) + (-0.03)4$

$h3 = 0.048$

h	h-y	$(h-y)^2$
0.108	0.008	0.000064
-0.124	-0.324	0.104
0.048	-0.352	0.123

$\Sigma(h-y)^2 = 0.2270$

(ii) hypothesis? Predicted values?

$h = \theta_0 + \theta_1 x_1 + \theta_2 x_2$

$h_1 = 0.06 + 0.108(3) + (-0.03)1$

$h_1 = 0.354$

$h_2 = 0.06 + 0.108(4) + (-0.03)3$

$h_2 = 0.402$

x1	x2	y
3	1	0.3
4	3	0.4

(iii) hypothesis?

$h = \theta_0 + \theta_1 x_1 + \theta_2 x_2$.

nits	Yellow	Sweet	Long	Total
Orange	350	450	0	650
Banana	400	300	350	400
Others	50	100	50	150
Total	800	850	400	1200

Not long = 650 - 0 , = 650

Not sweet = 650 - 4500

= 200

$Orange_{NB}$ = P(orange) x P(yellow) x P(not sweet) x P(not long)

$= \frac{650}{1200} \times \frac{350}{650} \times \frac{200}{650} \times \frac{650}{650}$

$= 0.09$

$Banana_{NB}$ = P(Banana) x P(yellow) x P(not sweet) x P(not long)

$= \frac{400}{1200} \times \frac{400}{400} \times \frac{100}{400} \times \frac{50}{400}$

$= 0.01$

$Others_{NB}$ = P(others) x P(yellow) x P(not sweet) x P(not long)

$= \frac{150}{1200} \times \frac{50}{150} \times \frac{100}{150} \times \frac{150}{150} = 0.02$

QNO2:- Apply Hierarchical [illegible] algorithm?

A	0	4
B	2	1
C	1	0
D	2.5	1.5
E	3	3

Distance

(A, B)(0, 4)(2, 1)

$= \sqrt{(0-2)^2 + (4-1)^2} = \sqrt{4+9} = \sqrt{13} = 3.60$

(A, C)(0, 4)(1, 0)

$= \sqrt{(0-1)^2 + (4-0)^2} = \sqrt{1+16} = \sqrt{17} = 4.12$

(A, D)(0, 4)(2.5, 1.5)

$= \sqrt{(0-2.5)^2 + (4-1.5)^2} = \sqrt{6.25+6.25} = 3.53$

(A, E)(0, 4)(3, 3)

$= \sqrt{(0-3)^2 + (4-3)^2} = \sqrt{9+1} = \sqrt{10} = 3.16$

(B, C)(2, 1)(1, 0)

$\sqrt{(2-1)^2 + (1-0)^2} = \sqrt{1+1} = \sqrt{2} = 1.41$

(B, D)(2, 1)(2.5, 1.5)

$= \sqrt{(2-2.5)^2 + (1-1.5)^2} = \sqrt{0.25+0.25} = 0.70$

(B, E)(2, 1)(3, 3)

$\sqrt{(2-3)^2 + (1-3)^2} = \sqrt{1+4} = \sqrt{5} = 2.23$

(C, D)(1, 0)(2.5, 1.5)

$\sqrt{(1-2.5)^2 + (0-1.5)^2} = \sqrt{(1.5)^2 + (-1.5)^2} = 2.12$

$)(1,0)(3,3)$

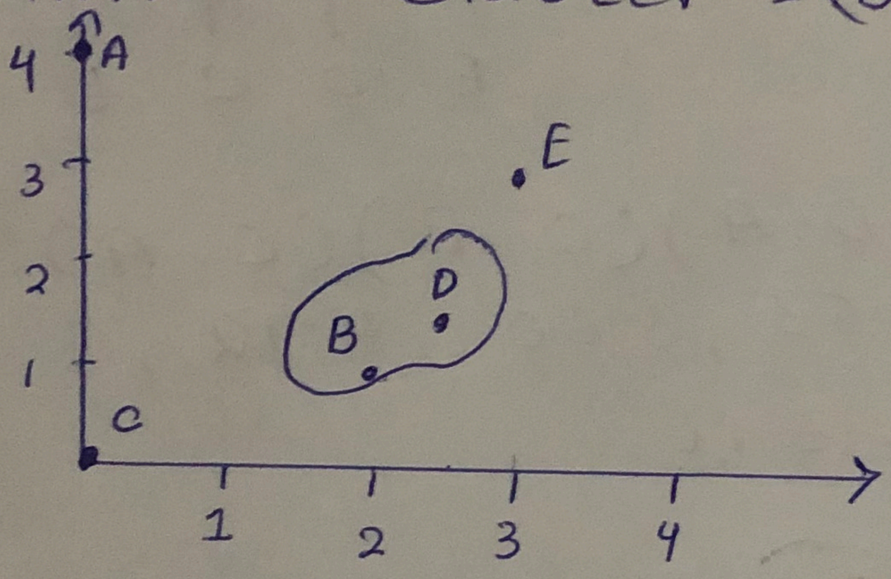
$$\sqrt{(1-3)^2+(0-3)^2} = \sqrt{4+9} = \sqrt{13} = 3.60$$

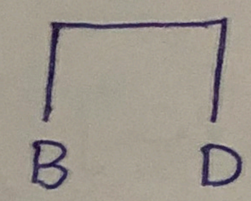
$(D,E)(2.5,1.5)(3,3)$

$$\sqrt{(2.5-3)^2+(1.5-3)^2} = \sqrt{0.25+2.25} = 1.58$$

	A	B	C	D	E
A	0				
B	3.60	0			
C	4.12	1.41	0		
D	3.53	0.70	2.12	0	
E	3.16	2.23	3.60	1.58	0

Minimum cluster = (B, D)





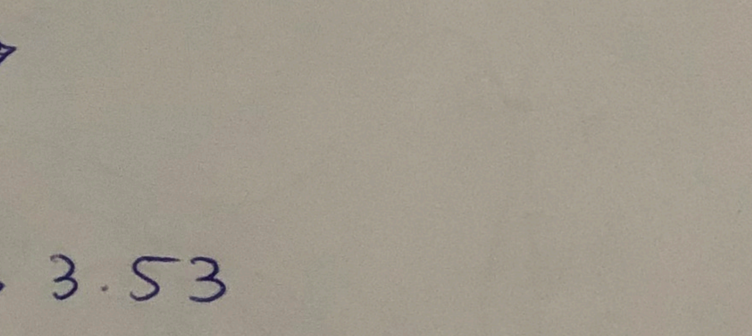
$(B,D)A = (B,A),(D,A)$

$= 3.60, 3.53 \Rightarrow 3.53$

$(B,D)C = (B,C)(D,C) = 1.41, 2.12 = 1.41$

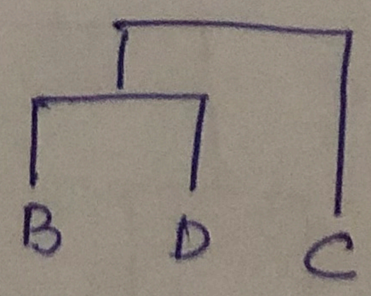
$(B,D)E = (B,E)(D,E) = 2.23, 1.58 = 1.58$

	A	B,D	C	E
A	0			
B,D	3.53	0		
C	4.12	1.41	0	
E	3.16	1.58	3.60	0



Minimum cluster

(B, D), C



$((B,D),C)A = (B,D,A)(C,A)$

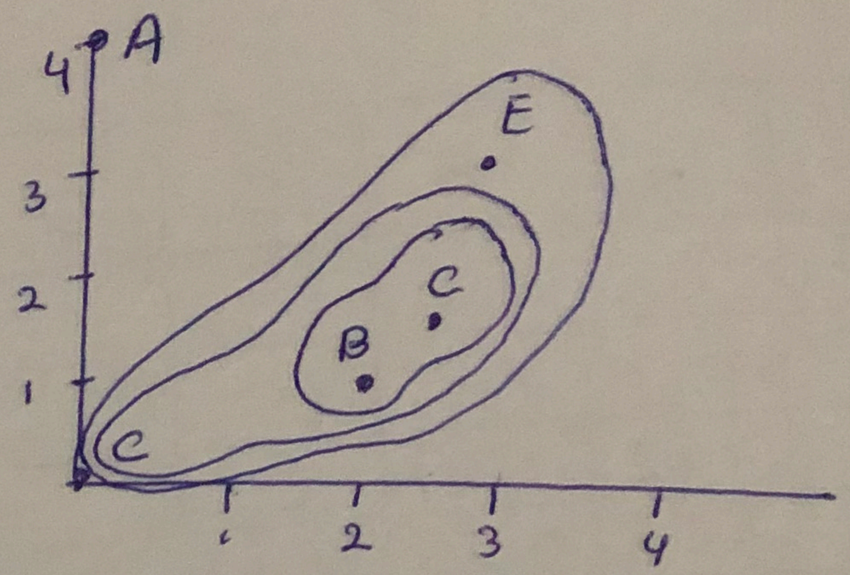
$= 3.53, 4.12$

$= 3.53$

$((B,D),C)E = (B,D,E)(C,E) = 1.58, 3.60$

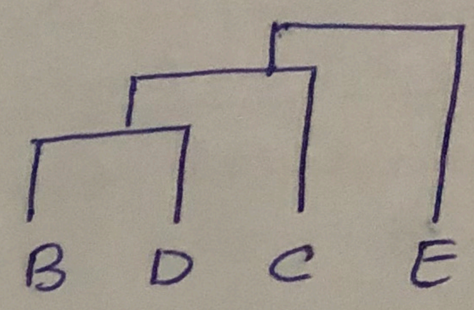
$= 1.58$

	A	(B,D),C	E
A	0		
(B,D)C	3.53	0	
E	3.16	1.58	0



Minimum cluster

$= [(B,D),C]E$



$(([(B,D),C]E)A = (B,D,A)(C,E)(C,A)(E,A)$

$= 3.53, 3.60, 4.12, 3.16$

$= 3.16$