

--	--	--	--	--	--	--	--

MCA DEGREE FIFTH SEMESTER EXAMINATION, JANUARY 2021
16-381-0566 DATA MINING
(Regular)

Time: 3 Hours

Maximum Marks: 50

(Answer ANY FIVE questions)
 (All questions carry EQUAL marks)

- 1 A What is an attribute? Explain different types of attributes with examples (5 Marks)
- B The table below describes the rate of economic growth (x_i) and the rate of return on the S&P 500 (y_i). Using the covariance formula, determine whether economic growth and S&P 500 returns have a positive or inverse relationship. Create a covariance matrix for the given data

Economic growth	S&P 500
3.1	9
3.5	13
5.0	15
4.6	11

(5 marks)

- 2 A. Explain briefly about Data Transformation and Data Discretization (5 Marks)
- B. Use these methods to normalize the following group of data: 350, 450, 650, 750, 1500
- min-max normalization by setting min=0 and max=1
 - Z score normalization (5 Marks)
- 3 A. What is a data cube? With examples explain typical OLAP operations. (5 Marks)
- B. A metal rod was gradually heated and its length, L was measured at various temperatures. The table below shows the details. Find if there is any correlation between the temperature condition and length. (5Marks)

Temperature	length
15.2	216
17.5	326
12.9	180
16.2	333
19.5	407
23.1	523
19.5	413
26.2	615
24.4	545
19.1	422
23.6	446
18.2	409
15.2	216

4. A. Differentiate between OLAP and OLTP (4 Marks)

B. Given the following data (in increasing order) for the attribute age: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70

i. Use smoothing by bin means to smooth these data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data. (3 Marks)

ii. Show a boxplot of the data. (3Marks)

5. A Consider the database given below. Find which attribute would be chosen as the first splitting attribute if a decision tree is constructed for the database Use information gain as the attribute selection measure (6Marks)

outlook	temp.	humidity	windy	play
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
rainy	mild	high	true	no

B. Explain in detail the algorithm for classification by back propagation. (4 Marks)

6. A. Explain in detail about the steps in Apriori algorithm (4 marks)

B. Consider a transaction database given below. Find the frequent item set using Apriori algorithm, where the minimum support count is 2. (6 Marks)

TID	List of item_IDs
T100	11,12,15
T200	12,14
T300	12,13
T400	11,12,14
T500	11,13
T600	12,13
T700	11,13
T800	11,12,13,15
T900	11,12,13
T900	11,12,13

- 7 A Suppose we have 4 types of medicines and each has two attributes (pH and weight index). Group these objects into $K=2$ group of medicine.

(5 Marks)

Medicine	Weight	pH Index
A	1	1
B	2	1
C	4	3
D	5	4

- B Explain in detail about Density Based clustering

(5 Marks)

#####