M.Sc. Examination 2018
Semester-IV
Computer Science
Course : MCSO-48
(Data Mining and Warehousing)

Time : 3 Hours **Full Marks : 40**

Questions are of value as indicated in the margin.

Answer Question No.**1** and **any four** from the rest.

1. a) "Dimensionality reduction is an important task in data pre-processing." Why?
   b) Differentiate classification, regression, prediction and clustering.
   c) What is the role of confusion matrix for classifiers.
   d) How does CLARA improve PAM? [4×2=8]
2. (a) Explain the KDD Process.
   b) How does Hunt's algorithm work? Explain with an example. [3+5=8]
3. a) Write a general algorithm for classification problem.
   b) How is the zero frequency problem handled in Bayesian classification method? Illustrate it.
   c) Distinguish between agglomerative and divisive algorithms. [4+2+2=8]
4. a) Describe $K-$medoids clustering algorithm. Analyze its time complexity.
   b) Consider 5 sample data items $\{A,B,C,D,E\}$. The following table shows the distance between any two elements:

Item	A	B	C	D	E
A	0	2	3	3	4
B	2	0	3	5	4
C	3	3	0	2	6
D	3	5	2	0	4
E	4	4	6	4	0

   By applying above algorithm, cluster the data items into two groups where A and C are the initial medoids. [(3+1)+4=8]
5. a) Define association rule, support, confidence and frequent itemset.
   b) Design an algorithm to generate all association rules from frequent itemsets.
   c) "Backpropagation algorithm is also known as two pass method." Justify. [2+3+3=8]
6. a) Suppose the confidences of the rules $A \to B$ and $B \to C$ are larger than some threshold, *minconf*. Is it possible that $A \to C$ has a confidence less than *minconf*? Justify your answer.

P.T.O.

b) Consider the following market-basket data transaction:

TID	**Items bought**
T10	*{K, E, Y, M, O, N}*
T20	*{K, I, E, O, C}*
T30	*{K, E, Y, N, O, D}*
T40	*{C, K, E, M, A}*
T50	*{I, C, Y, M, K}*

Using Apriori algorithm find frequent itemsets for the above database. Use support threshold =2. Generate all association rules for confidence = 80% [2+6=8]

7. a) Describe different steps of Bayesian Classifier.

b) Summarise the role of data mining in web mining.

c) Describe in detail about the applications of data mining in biomedical science. [4+2+2=8]

_____