



MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL

Paper Code : PEC-IT602B Data Warehousing and Data Mining

UPID : 006584

Time Allotted : 3 Hours

Full Marks : 70

The Figures in the margin indicate full marks.

Candidate are required to give their answers in their own words as far as practicable

Group-A (Very Short Answer Type Question)

1. Answer any ten of the following :

[1 x 10 = 10]

- (I) A star schema has what type of relationship between a dimension and fact table?
- (II) K-Means clustering is what type of learning?
- (III) Manhattan distance also called what?
- (IV) What is the full form of DSMS?
- (V) Web mining does not include what?
- (VI) AGM Approach is what type of candidate generation method?
- (VII) "FP tree does not need candidate generation. " – True/False
- (VIII) The clustering technique k-means is based on Centroid. True/False.
- (IX) The best-fitted trend line is one for which sum of squares of residuals or errors is minimum/maximum?
- (X) A stream data query processing architecture does not include which server?
- (XI) Which of the following frequent pattern mining technique mines without candidate generation?
 - a) Partitioning
 - b) Apriori
 - c) FP-growth
 - d) Dynamic intensive counting
- (XII) Choose correct alternatives from the following options:
 - i) The attribute with the highest information gain is chosen as the splitting attribute
 - ii) The attribute with the lowest information gain is chosen as the splitting attribute
 - iii) The attribute with the Highest Gini index is chosen as the splitting attribute
 - iv) The attribute with the lowest Gini index is chosen as the splitting attribute
 - a) Both (i) and (iii) is true
 - b) Both (ii) and (iii) is true
 - c) (i) is true and (iv) is false
 - d) (i) is true and (iii) is false

Group-B (Short Answer Type Question)

Answer any three of the following :

[5 x 3 = 15]

2. Define Support, Confidence, frequent itemset, lift and Association rule. [5]
3. Discuss briefly the tree construction principle. [5]
4. What is Clustering? Briefly describe the following approaches of clustering: partitioning methods, hierarchical methods, density-based methods, and grid-based methods. [5]
5. What is a time-series database? How time series data is different from sequential Data? [5]
6. Write k-means clustering algorithm/procedure. [5]

Group-C (Long Answer Type Question)

Answer any three of the following :

[15 x 3 = 45]

7. (a) Suppose that the data mining task is to cluster the following ten points (with (x, y) representing location) [7]
 into two clusters. Use distance function as $|x_i - x_j| + |y_i - y_j|$. Use k-medoid algorithm to determine the two clusters.

X1	2	6
X2	3	4
X3	3	8
X4	4	7
X5	6	2
X6	6	4
X7	7	3
X8	7	4
X9	8	5
X10	7	6

- (b) What are the four axioms of distance Metrics? [4]
- (c) Show that Manhattan distance satisfies all four Distance Metrics. [4]
8. (a) What is Data Stream? [2]
- (b) What are the challenges of stream data mining? [3]
- (c) What is Synopsis and synopsis data structures in context of stream data mining? [2+2]
- (d) Briefly describe the following stream data processing technique a) reservoir sampling, b)sliding window model [3+3]
9. (a) Given a dataset $X = \{(5.9, 3.2), (4.6, 2.9), (6.2, 2.8), (4.7, 3.2), (5.5, 4.2), (5.0, 3.0), (4.9, 3.1), (6.7, 3.1), (5.1, 3.8), (6.0, 3.0)\}$, perform a k-means clustering on this dataset using the Euclidean distance as the distance function. Here (K) is chosen as 3. The center of the 3 clusters is initialized as red (6.2, 3.2), green (6.6, 3.7) and blue (6.5, 3.0). Provide the final cluster centers. [9]
- (b) Describe CLARA and CLARANS. [3+3]
10. (a) What are the application fields for similarity search in time-series analysis? [3]
- (b) Why normalization can be necessary for similarity search? [2]
- (c) Define Min-Max Scaling and Z-Score Normalization. [2+2]
- (d) Convert the random variable $X = \{12, 19, 21, 23, 25, 35, 47, 48, 59, 65\}$ using Min-Max Scaling and Z-Score Normalization. [3+3]
11. (a) Briefly describe Supervised and unsupervised learning? [6]
- (b) Explain KNN algorithm with suitable example? [9]

*** END OF PAPER ***

<https://www.makaut.com>
Whatsapp @ 9300930012
Send your old paper & get 10/-
अपने पुराने पेपर्स भेजे और 10 रुपये पायें,
Paytm or Google Pay से