

Total No. of Questions : 8]

SEAT No. :

P3557

[Total No. of Pages : 4

[4959] - 1157

**B.E. (Computer Engineering) (Semester - I)**  
**Data Mining Techniques and Applications**  
**(2012Pattern)**

*Time : 2 1/2 Hours]*

*[Max. Marks : 70*

*Instructions to the candidates:*

- 1) *Answer Q.1 or Q.2, Q.3 or Q.4, Q.5 or Q.6, Q.7 or Q.8.*
- 2) *Neat diagrams must be drawn wherever necessary.*
- 3) *Figures to the right side indicate full marks.*
- 4) *Assume suitable data if necessary.*

**Q1) a)** Discuss whether or not each of the following activities is a data mining task. **[6]**

- i) Computing the total sales of a company.
- ii) Predicting the future stock price of a company using historical records.
- iii) Predicting the outcomes of tossing a pair of dice.

**b)** Explain the following terms: **[6]**

- i) Closed and maximal frequent itemsets.
- ii) Multilevel association rules.

**c)** Consider the following set of frequent 3-itemsets: **[8]**

{1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4}, {1, 3, 5}, {2, 3, 4}, {2, 3, 5}, {3, 4, 5}. Assume that there are only five items in the data set.

- i) List all candidate 4-itemsets obtained by the candidate generation procedure in Apriori algorithm.
- ii) List all candidate 4-itemsets that survive the candidate pruning step of the Apriori algorithm.

OR

**P.T.O.**

- Q2)** a) Describe the various methods for handling the missing values. [6]  
b) Consider the market basket transactions shown below: [6]

Transaction ID	Items-bought
T1	{Mango, Apple, Banana, Dates}
T2	{Apple, Dates, Coconut, Banana, Fig}
T3	{Apple, Coconut, Banana, Fig}
T4	{Apple, Banana, Dates}

Assuming the minimum support of 50% and minimum confidence of 80%

- i) Find all frequent itemsets using Apriori algorithm.  
ii) Find all association rules using Apriori algorithm.
- c) Explain with suitable example. [8]  
i) Confusion matrix.  
ii) K-Nearest-Neighbor Classifier.

- Q3)** a) Perform the single and complete link hierarchical clustering using the similarity matrix given below: [6]

	P1	P2	P3	P4	P5
P1	1.00	0.10	0.41	0.55	0.35
P2	0.10	1.00	0.64	0.47	0.98
P3	0.41	0.64	1.00	0.44	0.85
P4	0.55	0.47	0.44	1.00	0.76
P5	0.35	0.98	0.85	0.76	1.00

Show your results by drawing a dendrogram.

- b) Explain with suitable example the k-means algorithm. [7]
- c) Differentiate between Hierarchical and Partitional clustering. [4]

OR

**Q4)** a) Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8): [4]

- i) Compute the Euclidean distance between the two objects.
- ii) Compute the Manhattan distance between the two objects.

b) What are the requirements of clustering in data mining? [5]

c) Explain the following terms: [8]

- i) Density based clustering.
- ii) Agglomerative hierarchical clustering.
- iii) Grid based clustering.
- iv) Minkowski distance.

**Q5)** a) Explain the following terms: [6]

- i) Term frequency.
- ii) Stop list.
- iii) Inverse document frequency.

b) Enlist the dimensionality reduction techniques for text. Explain any one of them in brief. [6]

c) What is Weblog records? How it is used in Web usage mining? [5]

OR

- Q6)** a) Explain the following terms: [6]
- i) Precision.
  - ii) Recall.
  - iii) F-Score
- b) Differentiate between Web content and Web usage mining. [6]
- c) Explain Hyperlink-Induced Topic Search (HITS) algorithm. [5]

- Q7)** a) “Reinforcement learning is different from supervised learning”. Justify the above statement. [5]
- b) How different perspectives are represented in multi-perspective learning? [5]
- c) Explain with suitable diagram influence diagram. [6]

OR

- Q8)** a) What are the similarities and differences between reinforcement learning and systematic machine learning? [5]
- b) What are the issues and challenges in big data mining? [5]
- c) Write short note on multi-perspective decision making. [6]

