

Total No. of Questions : 10]

SEAT No. :

P3979

[5561]-683

[Total No. of Pages : 3

B.E. (Computer Engg.)

DATA MINING AND WAREHOUSING

(2015 Course) (Semester - I) (Elective - I) (410244D)

Time : 2½ Hours]

[Max. Marks : 70

Instructions to the candidates:

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

- Q1)** a) How to compute dissimilarity for categorical attributes with examples. [4]
b) Explain data cleaning techniques. [6]

OR

- Q2)** a) Explain types of attributes with examples. [4]
b) Suppose a group of 12 sales price records has been sorted as follows: [6]
5, 10, 11; 13; 15, 35, 50; 55; 72; 92; 204; 215:
Partition them into three bins by each of the following methods.
i) equal-frequency partitioning
ii) equal-width partitioning

- Q3)** a) Explain data discretization techniques. [4]
b) Explain OLAP operations in Multidimensional data model. [6]

OR

- Q4)** a) Explain following : [4]
i) Minkowski Distance
ii) Euclidean distance
b) Briefly compare the following concepts. You may use an example to explain your point(s). [6]
Snowflake schema, fact constellation, star schema, star schema query model

P.T.O.

- Q5) a)** Explain Steps of Apriori Algorithm and Steps of FP Growth Algorithm. [6]
- b) Explain mining Multilevel association rules. What is Uniform support? [6]
- c) Compare and contrast FP-Frowth algorithm with Apriori algorithm. [4]

OR

- Q6) a)** Consider an example with following set of transactions [6]

| TID | Items bought |
|-----|--------------|
| T1 | A,B,C |
| T2 | A,B,C,D,E |
| T3 | A,C,D |
| T4 | A,C,D,E |
| T5 | A,B,C,D |

Assume that we wish to find the association rules with at least 40% support and 40% confidence. Find the frequent itemsets and then association rule using Apriori algorithm.

- b) Explain the following terms: [6]
- Constraint based rule mining
 - Closed and maximal frequent itemsets
- c) What do you mean by frequent item set, Closed item set? Explain with example. [4]
- Q7) a)** Define Classification and Prediction. Explain decision tree based Classification method with suitable example. [8]
- b) Write and explain K-Nearest-Neighbour Classification algorithm with suitable example. [6]
- c) Write short note on Rule Induction Using a Sequential Covering Algorithm. [4]

OR

- Q8) a)** Explain the following : [8]
- Gini index
 - Gain ratio
 - Information gain
- b) Differentiate between Supervised and unsupervised Learning. [6]
- c) What are Bayesian classifiers? [4]

Q9) a) Explain following with example [8]

- i) Accuracy
- ii) Error Rate
- iii) Sensitivity
- iv) Specificity

b) Describe following. [8]

- i) Multiclass classification
- ii) Reinforcement learning

OR

Q10)a) Explain in detail following techniques to evaluate the accuracy of a Classifier. [8]

- i) Holdout method
- ii) Random subsampling

b) Explain following. [8]

- i) Multi-perspective learning
- ii) Wholistic learning