

B.TECH DEGREE EXAMINATIONS: APRIL/MAY 2014

(Regulation 2009)

Seventh Semester

INFORMATION TECHNOLOGY

ITY117: Data Warehousing and Data Mining

Time: Three Hours

Maximum Marks: 100

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. A----- for an attribute A partitions the data distribution of A into disjoint subsets or buckets.
 - a) Sampling
 - b) Clustering
 - c) Histogram
 - d) Classification
2. The mean and the standard deviation of the values for the attribute income are €54,000 and €16,000 respectively. With Z-score normalization a value €73,600 for income is transformed to
 - a) 2.225
 - b) 2.225
 - c) 3.335
 - d) 4.125
3. -----is a pragmatic collection of related facts, but does not have to be exhaustive or exclusive.
 - a) Data mart
 - b) Knowledge Database
 - c) Data Set
 - d) Database
4. A data warehouse _____.
 - a) Can be updated by end users.
 - b) Contains numerous naming conventions and formats.
 - c) Organized around important subject areas.
 - d) Contains only current data.
5. KDD
 - a) Knowledge Databases.
 - b) Discovery of knowledge
 - c) Knowledge Discovery in Databases.
 - d) Knowledge Data Dictionary
6. A cube for the highest level of abstraction is the
 - a) Top Cuboids
 - b) Apex Cuboids

- c) Base Cuboids
 - d) Bottom Cuboids
7. Clustering techniques consider data tuples as-----
 - a) Clustering data
 - b) Objects
 - c) Aggregate data
 - d) Constraints
8. Classification is _____
 - a) Supervised Learning
 - b) Unsupervised Learning
 - c) Physical Learning
 - d) Base Learning
9. The weighted outputs of the last hidden layer are input to units making up the ----- layer, which emits the network's prediction for given samples.
 - a) Input layer
 - b) Output layer
 - c) Hidden layer
 - d) Base layer
10. _____ focuses on analysis of the link structure of the web and one of its purposes is to identify more preferable documents.
 - a) Web content Mining
 - b) Web Structure Mining
 - c) We b Usage Mining
 - d) Document retrieval

PART B (10 x 2 = 20 Marks)

11. Explain the various OLAP operations.
12. Distinguish between data mart and data ware housing.
13. What is the importance of dimensionality reduction in data mining application.
14. Define Association rule.
15. Mention the types of Clustering.
16. Compare Supervised and Unsupervised Learning.
17. Define Descriptive Data Mining.
18. Why Naïve Bayesian classification is called 'Naïve'?
19. Mention some of the Data Mining Applications.
20. Describe text Mining.

PART C (5 x 14 = 70 Marks)

21. a) i) Describe the architecture of data mining system with block diagram. (8)

ii) Compare Operational Data Base Systems and Data Ware houses. (6)

(OR)

b) i) Describe the architecture of Data ware house (7)

ii) What is Data Mining? Explain the steps in Knowledge Discovery? (7)

22. a) Discuss about various strategies for data reduction operation

(OR)

b) (i) Suppose that the data for analysis include the attribute the frequency of stop words in documents. The values are given in increasing order: 14,15,16,16,19,20,20,21,23,23,27,27,35,37,40,45,46,52,70. Apply the following methods:

(i) Use Min-max normalization to transform the value 35 into the range from 0.0 to 1.0

(ii) Use z-score normalization to transform the value 35 where the standard deviation of the above frequency is 12.94

(ii) Briefly describe about smoothing techniques (7)

23. a) A data base has four transactions with min_sup=60%

TID	date	Items_bought
T100	10/15/99	{K,A,D,B}
T200	10/15/99	{D,A,C,E,B}
T300	10/15/99	{C,A,B,E}
T400	10/15/99	{B,A,D}

Find all frequent item set using apriori respectively

(OR)

b) Discuss about Statistical based outlier analysis

24. a) Discuss about K-means & K-Medoids Partition algorithms

(OR)

b) i) Given two objects represented by the tuples (22,1,42,10) and (20,0,36,8) .compute Euclidean distance, Manhattan distance and Minkowski distance using q=3 (10)

ii) Explain about nominal, ordinal and Ratio-scaled variables (4)

25. a) i) Discuss Spatial data mining (7)

ii) Discuss about Social impacts in data mining (7)

(OR)

b) Summarize the role of data mining in text databases
