53 (IT 701) DMDW

2014

DATA MINING AND DATA WAREHOUSING

Paper : IT 701

Full Marks : 100

Time : Three hours

The figures in the margin indicate full marks for the questions.

All questions are ***compulsory.***

1. Answer the following : 2×5=10

   (a) Define support and confidence in Association Rule Mining.

   (b) Specify the advantages of concept hierarchy in Data Mining.

   (c) What is market-basket analysis ?

   (d) Define a strong association rule.

   (e) Provide examples for incomplete and inconsistent data.

2. *(a)* Differentiable between OLAP and OLTP. 5

(b) Explain Roll up and Slice OLAP operations. 5

3. *(a)* Explain the data cube model for representing a data warehouse. 5

(b) Explain the star-schema model for a Data-warehouse design. 5

4. *(a)* Find the distance between :
$R = (22, 2, 45, 10)$ and $Q = (21, 0, 34, 9)$
using Manhattan distance and Minkowski distance for $P = 4$. 6

(b) Differentiate between supervised and unsupervised learning. 4

5. What is clustering ? Describe the partition around medoid (PAM) algorithm for clustering. Mention the strengths and weakness of the algorithm.

Explain how the presence of an outlier can affect *K*-means clustering algorithm. 2+5+4+4=15

6. *(a)* Find all the frequent itemsets for the following database, using Apriori algorithm. The minimum support count threshold in 3. Show the candidate and frequent itemsets for each database scan. 9

Tid	Items
1	milk, bread, eggs
2	bread, sugar
3	bread, cereal
4	milk, bread, sugar
5	milk, cereal
6	bread, cereal
7	milk, cereal
8	milk, bread, cereal, eggs
9	milk, bread, cereal

(b) List all the Association Rules generated and highlight the strong ones, min conf $= 50\%$. 6

7. Say BCDE, CEF, CDEG, ADEF, CDH, AEJ, BCD, ABDE, DJ, ABE represent 10 transactions of a dataset where every letter represents an item. Construct the FP-tree with a minimum support count 2. 10

Derive all the frequent itemsets, where item 'A' is involved. 5

Or

(a) State Bayes theorem and explain the Bayesian belief network in detail. 10

(b) Explain in detail the decision tree classifier. 5

8. Write short notes on the following : 5×3=15

(a) Data pre-processing techniques

(b) Properties of a Data-Warehouse

(c) Accuracy of a classifier.

Or

(a) Robustness of a clustering algorithm

(b) Apriori property

(c) Data Reduction. 5×3=15
