



Sixth Semester B Tech C3 Examination June 2017

Course : DATA MINING TECHNIQUES

Course Code: BTCS15F6410

Time: 3 hours

Max. Marks: 100

Note: Answer ONE FULL question from each section.

SECTION-I (UNIT - I)

REVA - LIBRARY

1. a) Data Mining is an integral part of KDD, Explain the process of Knowledge discovery in databases with a neat diagram. 09
- b) Explain the different challenges that motivated the development of data mining technologies. 10
- c) Define and give examples for: i) Data set ii) Attribute iii) Measurement scale 06

OR

2. a) Data preprocessing is a broad area and consists of a number of different strategies and techniques that are interrelated in complex ways. List and explain different steps involved in data preprocessing. 10
- b) For the following vectors, X and Y calculate the indicated similarity or distance measures. 15
 - i. $X=(1,0,0,0,0,0,0,0,0,0), Y=(0,0,0,0,0,0,1,0,0,1)$ Find SMC and J.
 - ii. $X=(3,2,0,5,0,0,0,2,0,0), Y=(1,0,0,0,0,0,0,1,0,2)$ Find Cos(X,Y).

SECTION-II (UNIT - II)

3. a) Consider the following transaction database for a super market in the given table : 15

Customers	Items
C1	Milk,egg,bread,chip
C2	Egg,popcorn,chip,beer
C3	Egg,bread,chip
C4	Milk,egg,bread,popcorn,chip,beer
C5	Milk,bread,beer
C6	Egg,bread,beer
C7	Milk,bread,chip
C8	Milk,egg,bread,butter,chip
C9	Milk,egg,butter,chip

Generate Apriori rules by assuming the minimum support of 30% (atleast three transactions) and minimum confidence of 60%.

- b) The number of frequent itemset produced from a transaction data set can be large. It is useful to identify a small representative set of itemsets from which all other frequent itemsets can be derived. Explain the two compact representation of frequent itemset, with example? 10
- OR**
4. a) Define and give examples for: i) itemset ii) support count ii) Association Rule 06
 - b) Write an algorithm to find frequent itemset generation of the Apriori algorithm. 12
 - c) State and explain the different ways to reduce the computational complexity of frequent itemset generation. 07

PTO

SECTION-III (UNIT – III)

5. a) Define classification. Explain general approach for solving a classification problem 08
 b) Explain how the decision tree works with an example. 08
 c) Define Rule coverage and Accuracy .Consider the dataset given below and Find 09
 Rule coverage and Accuracy for the rules given:
 i)(STATUS=SINGLE) -----→NO
 ii)(STATUS=MARRIED) ---→NO
 iii)(STATUS=DIVORCED)--→YES

TID	REFUND	MARITAL STATUS	TAXABLE INCOME	CLASS
1	Yes	Single	125K	NO
2	No	Married	700K	NO
3	No	Single	70K	NO
4	Yes	Married	120K	NO
5	No	Divorced	95K	YES
6	No	Married	60K	NO
7	Yes	Divorced	220K	NO
8	No	Single	85K	YES
9	No	Married	75K	NO
10	No	Single	90K	YES

OR

6. a) Write an algorithm for decision tree induction, also explain each step in detail? 07
 b) Explain the different strategies for growing a classification rule. 08
 c) Decision tree induction algorithms must provide a method for expressing an attribute test condition and its corresponding outcomes for different attribute types. Explain in detail. 10

SECTION-IV (UNIT – IV)

7. a) Clustering aims to find useful groups of objects, where usefulness is defined by the goals of the data analysis. List and explain different types of clusters. 10
 b) Write and explain basic K-Means algorithm to find clusters. 08
 c) Write an algorithm for Agglomerative clustering technique and explain with an example. 07

OR

8. a) Explain in detail, DBSCAN algorithm, with example. 10
 b) Differentiate between K-means and DBSCAN. 08
 c) List and explain the important characteristics of clusters such as prototype, graph and density based clusters. 07
