

**School of Information Technology
IIT Kharagpur**

Course Id: IT60107 Data Warehousing and Data Mining

Date: November 21, 2006

Total Time: 3 Hours

Max. Marks: 100

Instructions: Answer any five questions. You may answer the questions in any order. However, all parts of the same question must be answered together. Clearly state any reasonable assumption you make.

1. Consider the 5 transactions given below. If minimum support is 40% and minimum confidence is 50%, determine the frequent itemsets and association rules using the *FP-Tree* algorithm. **[15+5=20]**

Transaction	Items
T1	Bread, Jam, Butter, Ketchup, Egg, Milk
T2	Bread, Butter, Egg
T3	Milk, Butter
T4	Egg, Bread, Ketchup, Milk
T5	Egg, Milk

2. Consider the following table of transactions. Each row represents a transaction and each column represents an item. If an item is present in a transaction, it is marked as '1', else it is marked as '0'. Determine the Frequent Itemsets using the Dynamic Itemset Counting algorithm and all the association rules. Use intervals of 5 transactions, minimum support = 50% and minimum confidence = 80%. **[18+2=20]**

	A1	A2	A3	A4	A5	A6	A7	A8	A9
1	1	1	1	1	1	1	0	1	1
1	1	0	1	0	1	0	0	1	0
1	1	0	1	1	1	1	1	0	0
1	1	1	1	1	1	1	1	0	0
1	1	0	0	1	1	1	1	0	0
1	1	1	1	1	0	1	0	0	1
1	1	0	1	0	1	1	1	0	1
0	0	0	0	0	1	1	0	0	0
1	1	0	1	1	1	1	0	1	0
0	0	1	0	1	0	1	0	1	1
1	0	1	0	1	0	1	0	0	0
0	0	0	0	0	1	1	0	1	1
0	1	0	1	0	1	1	1	0	0
1	0	1	0	1	0	1	0	0	0
0	1	1	1	1	0	0	1	1	1

3. Consider the following set of transactions for a number of customers. Determine the maximal sequences that have at least 50% support. [20]

Transaction Date	Customer Id	Items Bought
11/11/2006	1	A, B
11/11/2006	2	B, D
11/11/2006	3	A, C, B
11/11/2006	4	B, A
11/11/2006	5	C
12/11/2006	1	A, D
12/11/2006	2	B, D
12/11/2006	3	A, D, B
12/11/2006	4	A, B
13/11/2006	1	B, D
14/11/2006	2	A, D
14/11/2006	3	B, A
15/11/2006	1	A, D
15/11/2006	5	C

4. Build a Decision Tree using the training data in the table given below. Divide the Height attribute into ranges as follows: (0,1.7], (1.7,1.9], (1.9, 2.5] and Weight into the ranges (0,50], (50, 70], (70, 80], (80,200) [20]

Gender	Height	Weight	Class
F	1.60 m	58	RW
M	1.95 m	120	OW
F	1.70 m	62	UW
F	1.75 m	72	RW
F	1.68 m	40	UW
M	1.85 m	99	OW
F	1.60 m	35	UW
M	1.69 m	60	UW
M	1.99 m	120	OW
M	2.00 m	120	OW
F	1.80 m	75	RW
M	1.95 m	110	OW
F	1.89 m	70	UW
M	1.80 m	78	RW
F	1.75 m	55	UW

5. There are two clusters C1 and C2 formed from a dataset. The Clustering Feature (CF) vectors of these two clusters are: $CF1 = (4, 13, 51)$ and $CF2 = (4, 34, 294)$. Determine the following:

- a) Centroids of C1 and C2
- b) Radius of C1
- c) Diameter of C2
- d) Define a valid inter-cluster distance and find the corresponding inter-cluster distance between C1 and C2

[2+6+6+6=20]

6. Cluster the following 5 data points using CLARANS with $maxneighbor = 2$ and $numlocal = 1$. Assume that the initial set of selected medoids is $\{2, 3\}$. Show the steps in detail but you need not explain them. They should be self-explanatory.

2, 3, 4, 9, 10

[20]