**Birla Institute of Technology & Science, Pilani**

**Work-Integrated Learning Programmes Division**

**First Semester 2017-2018**

**Mid-Semester Test**

**(EC-2 Regular)**

Course No. : SS ZG548

Course Title : ADVANCED DATA MINING

Nature of Exam : Closed Book

No. of Pages = 2

# No. of Questions = 4

Weightage : 35%

Duration : 2 Hours

Date of Exam : 24/09/2017 (FN)

Note:

1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.
4. Explain the following concepts with respect to data mining. [2 + 2 + 2 = 6]
5. Difference between error and noise in data
6. Predictive and Descriptive data mining tasks with example
7. Overfitting and regularization.
8. Describe incremental association mining setting. How fast update algorithm (FUP) differs from fast update 2 algorithm (FUP2). Explain how one can maintain certain data structures required by FUP2 in an incremental way. [1 + 1 + 2 = 4]
9. Consider association rule mining for incremental databases. What are the main advantages of using Compact Pattern Stream tree (CPS Tree) over FP-Tree. Assuming a pan contains two transactions (that is restructuring happens after arrival of two transactions). And each window to contain two pans. Consider step by step arrival of following six transactions
10. A, B, C, D, E
11. B, C, E, F
12. E, C, D, F
13. B, F, C, A
14. D, E, C
15. F, B, D

Starting from a null CPS tree, draw all intermediate CPS tree on incremental arrival of every transaction. Clearly explain modifications to the CPS tree in each step. [2 + 8 = 10]

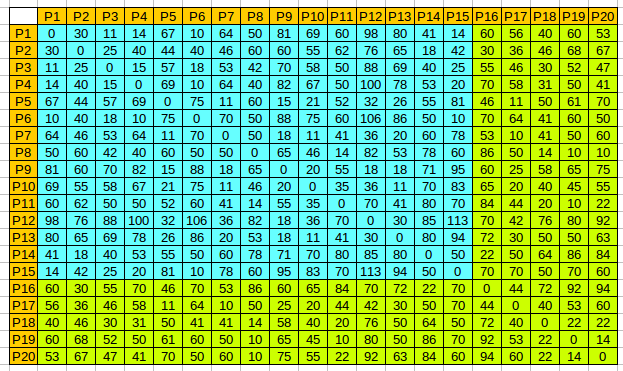
1. Consider the problem of clustering an evolving database. Assume database updates happens at a regular interval. Let initially there be 15 datum points P1, P2, …, P15 spread in 2D space. Distances between every pair of datum point is provided in Table-1. A Density-Based Spatial Clustering algorithm DBSCAN is applied with parameters **Eps=31** and **MinPts=3**. [3 + 1 + 3 + 3 + 1 + 4 = 15]
2. Identify Core, border and noise points
3. How many clusters are there?
4. Determine the cluster membership for each of the core or border point

***SS ZG548 (EC-2 Regular) First Semester 2017-2018 Page 2***

Consider the arrival of five more data points P16, P17, P18, P19, P20 in the database. Now we have 20 datum points. Refer Table-1 to get the distance between any two datum point of the database. Apply incremental DBSCAN on updated database to report the following.

1. Identify Core, border and noise points
2. How many clusters are there?
3. Determine the cluster membership for each of the core or border point

**Table 1.: Distance between every pair of datum points is given below**



**\*\*\*\*\*\*\*\*\*\***