**BT 3041 Analysis and Interpretation of Biological Data**

Class : Btech

Date : 21-2-2017

Time : 8:00-8:50 am

QUIZ 1 Examination

Marks: 20

15/20

Part A

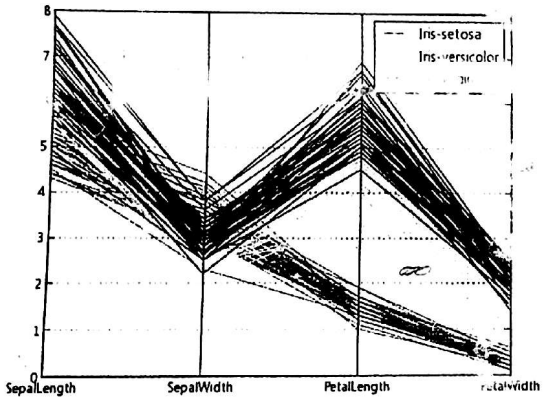
1. For the data set X = {(1,0), (0,0), (0,1), (2,1) (3,0), (3,-1), (4,-1)}. Applying DBSCAN method to this data, label the points as CORE, BORDER and NOISE (eps = 1.1; minpts = 3). Cluster the data using DBSCAN algorithm. (5 marks)

Part B

2. Classify the following clustering algorithms into one of two categories. Explain your answer.
   a) K-Means: exclusive or overlapping? exclusive
   b) Fuzzy-k means: exclusive or overlapping? overlapping
   c) Scale-based clustering: hierarchical or partitional? hierarchical
   d) Self-organizing map: hierarchical or partitional? partitional (4 marks)

3. If you want to compare the genomes of two human beings, which of the following distance measures will you use? (1 mark)
   A) Euclidean distance, B) Jaccard distance, C) Hamming distance, D) Manhattan distance.

4. What kind of a plot is the following depiction of IRIS DATA? (1 mark)




A)Circle segment display, B) Dimensional stacking, C) Parallel coordinates, D) Cone trees

5. Which of the following visualization methods can best depict hierarchical data? (1 mark)
   a) Circle segment display, b) Dimensional stacking, c) Chernoff faces, c) Cone trees

6. Which of the following visualization methods is NOT a depiction of high dimensional data? (1 mark)
   b) Circle segment display, b) Dimensional stacking, c) Chernoff faces, c) Cone trees

7. Which of the following statements best describes the purpose of a Pareto chart (1 mark)?

A) It depicts the most important factors to an event
B) It is similar to a bar chart wherein the widths of the bars are unequal
C) It depicts a vector field in 2D or 3D spaces
D) It depicts the schedule according to which an activity will be performed

8. Features of two vehicles are defined below. The table shown below is relevant for both the subproblems (2 marks).

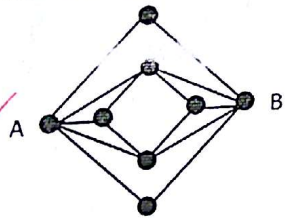
| | wheels | wings | fuel | rudder | tail |
|---|---|---|---|---|---|
| Airplane | Yes | Yes | Yes | Yes | No |
| Car | Yes | No | Yes | No | No |

8.1) The Jaccard similarity of the two vehicles is:
A) ½, B) 1/3, C) ¼, D) 2/5

8.2) The Jaccard distance of the two vehicles is:
B) ½, B) 1, C) 2, D) 4

9. If you fit a single Gaussian density function to the following data set, S = {-2, -1, 0, 1, 2}, using Expectation Maximization using Gaussian mixture model, what are the values of μ, σ, and α?
A) μ=1, σ=2, α=0.5, B) μ=0, σ=1.4, α=1, C) μ=0, σ = 1.7, α=1, D) μ=1, σ=1.4,α=1.

10. Figure below shows the graph representation of a data set. (Only links with strength greater than a threshold are shown). The shared nearest neighbor similarity between nodes A and B is,




A) 2, B) 3, C) 4, D) 5.

11. A key weakness of DBSCAN algorithm is inability to deal with,
A) Variability in cluster size, B) variability in cluster shape, C) variability in cluster density, D) high dimensional data.

12. Which of the following is a graph based clustering method?
A) DBSCAN, B) shared nearest neighbor method, C) k-means clustering, D) k-medoids clustering.