DEPARTMENT OF BIOTECHNOLOGY, IIT, MADRAS

CHENNAI – 36

BT 3041 Analysis and Interpretation of Biological Data

Class : B.Tech Date : 02-8-2020

Time : 10:00 AM- 1:00 PM END SEMESTER Examination Marks: 50

Part A (1 mark each):

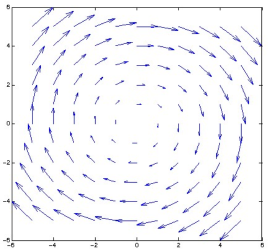
1. Which of the following is a graph based clustering method?
   A) DBSCAN, B) shared nearest neighbor method, C) k-means clustering, D) k-medoids clustering.

2. The definiteness of matrix A is:
   A = [1 1
   0 -1]
   A) Positive definite, B) negative definite, C) indefinite, D) positive semi-definite

3. The five-number summary of a distribution has the following five quantities:
   A) [Q1, Q2, Q3, Q4, Q5]
   B) [Minimum, Q1, Median, Q3, Maximum]
   C) [Q1, Q2, Q3, first moment, second moment]
   D) [Mean, median, mode, minimum, maximum]

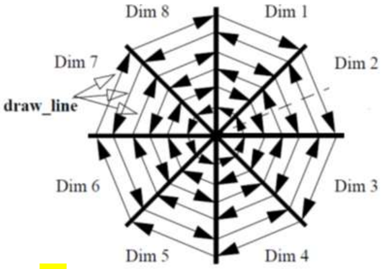
4. Gradient w.r.t. ‘x’ of the scalar function f(x) = $a^T x + x^T Ax$, where ‘x’ is an n-dimensional vector is,
   A) a + Ax, B) 2a + Ax, C) a + 2Ax, D) $2a^T + 2Ax$.
5. Which of the following sentences best describes the purpose of Pareto charts?
   A) To clearly describe the timelines of execution of a project
   B) To depict the comparative progression of two quantities with time
   C) To present higher dimensional data in two dimensions
   D) To highlight the most important ones from among a large set of factors

6. Name the type of the plot shown below:



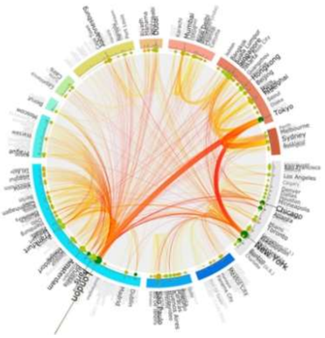
   A) circle plot, B) vector field plot, C) surface plot, D) mesh plot

7. The type of the plot shown below is,



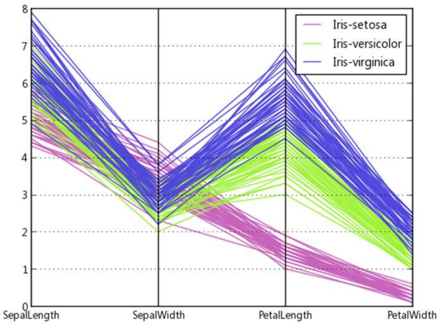
A) circle segment display, B) cone tree, C) scatter plot, D) surface plot

8. The type of the plot shown below is,



A) circle segment display, B) cone tree, C) circos, D) surface plot

9. What kind of a plot is the following depiction of IRIS DATA?



A) Scatter plot, B) Parallel coordinates plot, C) Stream flow plot, D) cone tree

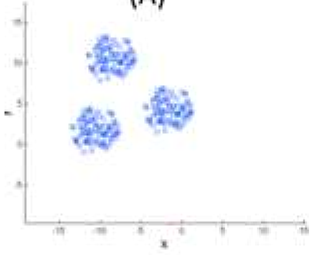
10. Which of the following distance measures are particularly suited for asymmetric binary variables?
    A) Euclidean distance, B) Minkowski distance, C) Jaccard distance, D) Hamming distance
11. Which of the following is NOT a property of k-means clustering?
    A) Existence of Local minima, B) Unique solution, C) existence of an objective function, D) offers a way of finding the “right” number of clusters.
12. Agglomerative clustering is a type of,
    A) Fuzzy clustering, B) hierarchical clustering, C) probabilistic clustering, D) density-based clustering.
13. Which of the following is a dimensionality reduction method?
    A) K-means clustering, B) Self-organizing Map, C) Fuzzy clustering, D) DBSCAN

14. The number of edges in a simple, complete directed graph with N vertices is,

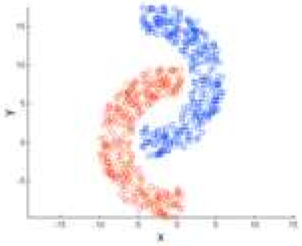
*A) N * (N-1) / 2*
*B) N * (N+1) / 2*
*C) N * (N-1)*
*D) N * N*

15. For which of the following data sets will K-means be able to detect the right clusters most successfully?

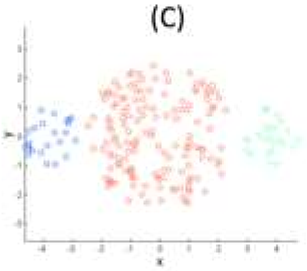
(A)



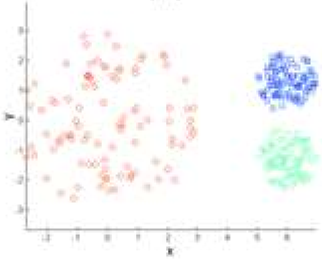
(B)



(C)

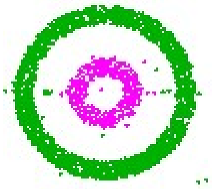


(D)



16. Which of the following clustering methods offers a way of finding the “right” number of clusters?
A) Scale-based clustering, B) agglomerative clustering, C) fuzzy clustering, D) DBSCAN

17. Which of the following clustering methods is ideal for discovering the two clusters present in the data set shown below?



A) K-means clustering, B) Expectation maximization using Gaussian mixture model, C) Fuzzy clustering, D) Shared nearest neighbor clustering

18. What is the purpose of ‘early stopping’ in a Multilayer Perceptron’s training?
A) Improve training accuracy
B) Improve testing accuracy
C) Save on training time
D) Network pruning

19. Which of the following is NOT a necessary property of the hidden layer 'nonlinearity' of a MLP for the universal approximation theorem to apply?
   A) Continuity
   B) Boundedness
   C) Differentiability
   D) Monotonicity

20. Which of the following is NOT a radial basis function?
   A) $\exp(-||x-w||^2)$, B) $(1 + (w^Tx)^2)$, C) $||x-w||^2$, D) $1/(1+||x-w||^2)$

21. The 'interpolation theorem' that assures unique training solution for a Radial Basis Function network requires that,
   A) The data points are all distinct
   B) The autocorrelation matrix of the training data has a 'flat spectrum'
   C) The data points are within a n-dimensional hypercube (n = input dim)
   D) The number of data points is greater than the number of hidden nodes

22. The training algorithm for a Radial Basis Function network is best described as,
   A) Backpropagation algorithm, B) delta rule, C) k-means and pseudoinverse, D) none of the above.

23. The uniqueness of the learning algorithm of a Support Vector Machine compared to backpropagation algorithm consists in optimizing:
   A) Output error, B) cross entropy, C) margin, D) Kullback-Leibler distance

24. Which of the following is the correct formulation of learning for a linear Support Vector Machine? (N is the number of data points.)

   A) $\max \|w\|^2$, *subject to* $y_i(w \bullet x_j + b) \geq 1,\ j = 1, \ldots, N$
   B) $\min \|w\|^2$, *subject to* $y_i(w \bullet x_j + b) \geq 0,\ j = 1, \ldots, N$
   C) $\max \|w\|^2$, *subject to* $y_i(w \bullet x_j + b) \leq 1,\ j = 1, \ldots, N$
   D) $\min \|w\|^2$, *subject to* $y_i(w \bullet x_j + b) \geq 1,\ j = 1, \ldots, N$

25. The space and time complexities for agglomerative clustering are:
   A) $O(N^2)$ for time, $O(N^3)$ for space
   B) $O(N^3)$ for time, $O(N^2)$ for space
   C) $O(N^3)$ for time, $O(N^4)$ for space
   D) $O(N^4)$ for time, $O(N^3)$ for space

26. If a deep network with 3 hidden layers (HL = HL#1,2,3 from input towards output layer) is trained on recognizing faces, what is the pattern of responses of hidden layer neurons that you expect to see?
   A) HL#1 recognizes whole faces, HL#2 recognizes edges and HL#3 recognizes parts of the face
   B) HL#1 recognizes whole faces, HL#2 recognizes parts of a face and HL#3 recognizes edges

C) HL#1 recognizes edges, HL#2 recognizes parts of the face, and HL#3 recognizes whole faces

D) HL#1 recognizes edges, HL#2 recognizes whole faces and HL#3 recognizes parts of the face

27. Specificity of a classifier is defined as: (where, TP = True positives, TN= True Negatives; FP = False positives, FN = False Negatives)

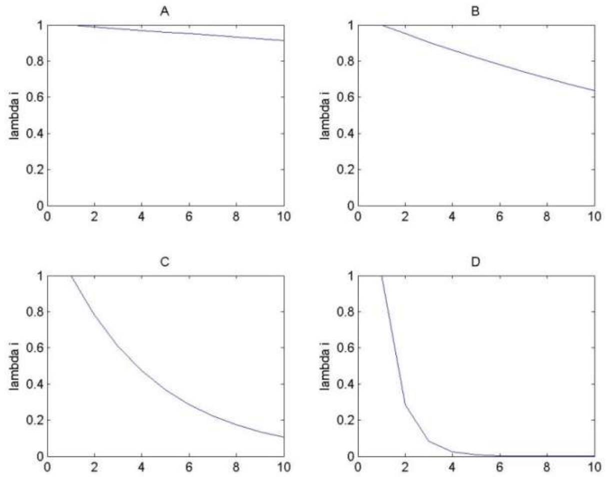
A) = TP/(TP+FP)

B) = TP/(TP+FN)

C) =TN/(TN+FP)

D) =TN/(TN+FN)

28. The "spectra" (sorted eigenvalues) of autocorrelation matrices of 4 data sets (dim = 10) are shown in figure below. Which of them allows the greatest scope for compression?



A) A, B) B, C) C, D) D

29. Which of the following visualization methods can best depict hierarchical data? (1 mark)

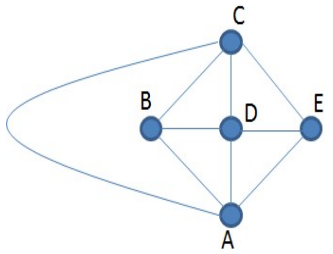
a) Circle segment display, b) Dimensional stacking, c) Chernoff faces, c) Cone trees

30. 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

Part B (2 marks each)

1. A graph representation of a dataset S consisting of 5 points (A to E) is shown below. A pair of points are connected by a link only if their similarity exceeds a certain threshold.



The number of shared nearest neighbors of node pairs (A,D), (B,D), (C,D), (E,D) are:

A) 3, 2, 2, 3, B) 3,2,3,2, C) 3,3,2,2 D) 2,3,3,2

2. The data set S = {(0,0), (0,1), (3,0), (3,2), (7,0), (7,5)} is being clustered using Agglomerative clustering. After the first step the points (0,0) and (0,1) are combined to form Cluster #1. The next merger is between the following: (The inter-cluster similarity measure is MIN.)
   A) (0,0) and (3,0)
   B) Cluster #1 and (3,0)
   C) (3,0) and (3,2)
   D) Cluster # 1 and (3,2)
3. The problem setting is the same as in the previous problem. But The inter-cluster similarity measure is MAX. The next merger is between the following:
   A) (0,0) and (3,0)
   B) Cluster #1 and (3,0)
   C) (3,0) and (3,2)
   D) Cluster # 1 and (3,2)
4. Consider the two discriminant functions,
   $y_1 = g(x_1 - x_2 - 0.5)$ and $y_2 = g(x_1 + x_2 - 0.5)$, g(x) is the logistic function.
   If $y_1 > y_2$, X→ Class #1, else X → Class #2. The decision surface separating the two classes is described as,
   A) x1 = 0, B) x1-x2 = 0, C) x1 + x2 = 0, D) x2 = 0.

5. Consider the 3 layer MLP whose input/output function is defined as follows. $y_1 = g(x_1)$, $y_2 = g(x_2)$, and $y = g(y_1 + y_2 - 1.3)$ where g(x) is the step function. Which of the following statements is TRUE?
   A) y is non-zero in the first and second quadrants.
   B) y is non-zero in the first quadrant alone.
   C) y is non-zero in the third quadrant alone.
   D) y is zero in the first, second and fourth quadrants.

6. A 3-layer MLP is defined as,
   $V_1 = \sigma(x_1 + x_2 - 1.5)$ $V_2 = \sigma(x_1 + x_2 - 0.5)$ $y = \sigma(V_1 - V_2 - 0.5)$
   where $\sigma()$ is the step function. The function $y = f(x_1,x_2)$ represents which of the following?
   A) OR gate, B) AND gate, C) XOR gate, D) XNOR gate

7. The maximum value of the cost $C = x^2 + y^2$, under the constraint $3x + 4y = 10$, is:

   A) 1, B) 2, C) 3, D) 4.

8. 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) $\mu=1$, $\sigma=2$, $\alpha=0.5$, B) $\mu=0$, $\sigma=1.4$, $\alpha=1$, C) $\mu=0$, $\sigma = 1.7$, $\alpha=1$, D) $\mu=1$, $\sigma=1.4$,$\alpha=1$.

9. A linear discriminant function defined as $y = w^T x - w_0$, is used for a binary classification task such that if y >0, X→ Class #1 (d = +1), else X → Class #2 (d = -1). The learning equations for this task, which are invoked only when there is an erroneous classification, are: (1 mark)
A) $\Delta w = \eta d_j x_j$; $\Delta w_0 = -\eta d_j$, B) $\Delta w = -\eta d_j x_j$; $\Delta w_0 = -\eta d_j$; C) $\Delta w = -\eta d_j x_j$; $\Delta w_0 = \eta d_j$, D) $\Delta w = \eta d_j x_j$; $\Delta w_0 = \eta d_j$.

10. Consider a MLP defined as, $y_1 = g(x_1 + x_2 - 1.5)$, $y_2 = g(x_1 + x_2 - 0.5)$, and $y = g(y_2 - y_1 - 0.5)$. The region defined by y >0 is best described as,
A) An infinitely long rectangular strip, B) a semi-infinite region, C) an angular region with an infinite area, D) none of the above.