MACHINE LEARNING

ANSWER SHEET

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following is an application of clustering?
   a. All of the above
2. On which data type, we cannot perform cluster analysis?
   a. None
3. Netflix's movie recommendation system uses-
   a. Reinforcement learning and Unsupervised learning
4. The final output of Hierarchical clustering is-
   a. The tree representing how close the data points are to each other
5. Which of the step is not required for K-means clustering?
   a. None
6. Which is the following is wrong?

   a. k-nearest neighbour is same as k-means

7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?
   i. Single-link
   ii. Complete-link
   iii. Average-link
   Options:

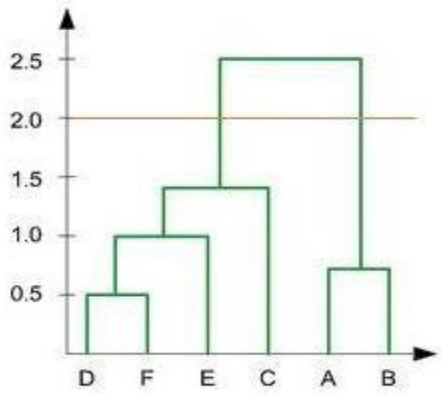
   b. 1, 2 and 3

8. Which of the following are true?
   i. Clustering analysis is negatively affected by multicollinearity of features
   ii. Clustering analysis is negatively affected by heteroscedasticity
   Options:

   a. 1 only

MACHINE LEARNING

9. In the figure above, if you draw a horizontal line on y-axis for y=2. What will be the number of clusters formed?



a. 2

10. For which of the following tasks might clustering be a suitable approach?

a. Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.

11. Given, six points with the following attributes:

point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

Table : X-Y coordinates of six points.

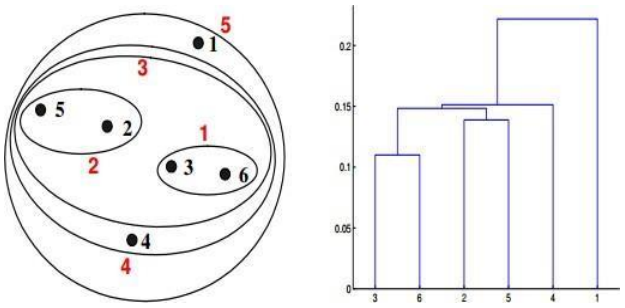
	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points

MACHINE LEARNING

Which of the following clustering representations and dendrogram depicts the use of MIN or Single link proximity function in hierarchical clustering:

a.



MACHINE LEARNING

12. Given, six points with the following attributes:

point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

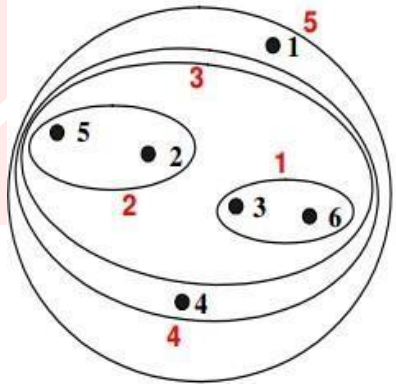
Table : X-Y coordinates of six points.

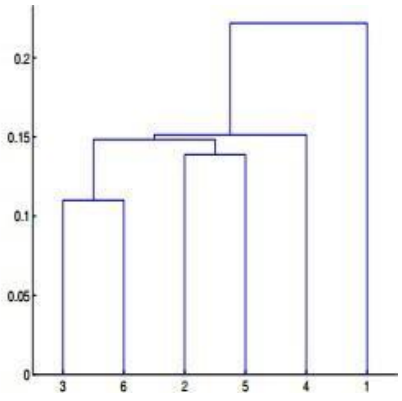
	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points

Which of the following clustering representations and dendrogram depicts the use of MAX or Complete link proximity function in hierarchical clustering.

a.





MACHINE LEARNING

Q13 to Q14 are subjective answers type questions, Answers them in their own words briefly

13. What is the importance of clustering?
14. How can I improve my clustering performance?

ANSWER Q13 TO Q14

13. Clustering is a widely used unsupervised learning technique that allows us to find hidden pattern or relationships between the data points based on the common attributes in the data.

14. Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance.