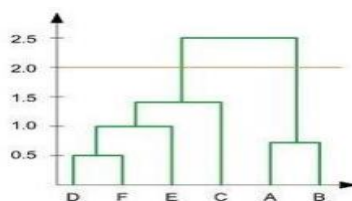


MACHINE LEARNING

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?
d. All of the above
2. On which data type, we cannot perform cluster analysis?
d. None
3. Netflix's movie recommendation system uses-
c. Reinforcement learning and Unsupervised learning
4. The final output of Hierarchical clustering is -
b. The tree representing how close the data points are to each other
5. Which of the step is not required for K-means clustering?
d. None
6. Which of the following is wrong?
c. 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?
 - 1) Single-link
 - 2) Complete-link
 - 3) Average-linkd. 1, 2 and 3
8. Which of the following are true?
 - 1) Clustering analysis is negatively affected by multicollinearity of features
 - 2) Clustering analysis is negatively affected by heteroscedasticitya.1 only
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?

b. Given a database of information about your users, automatically group them into different market segments

11. Given, six points with the following attributes:

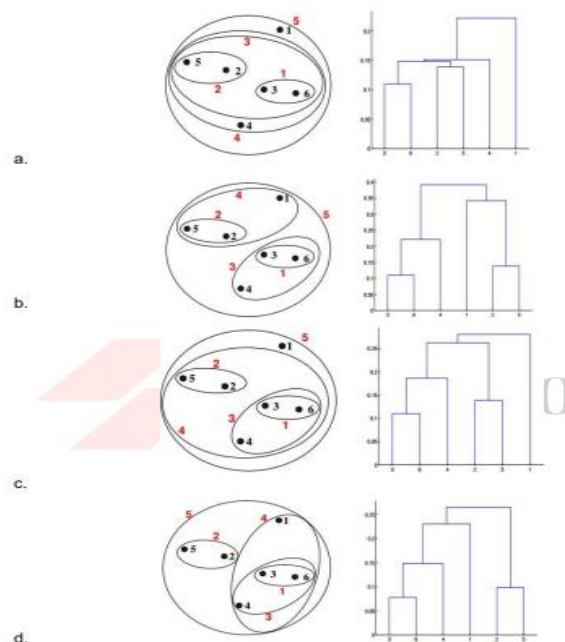
point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

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

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



Option: a

12. Given, six points with the following attributes:

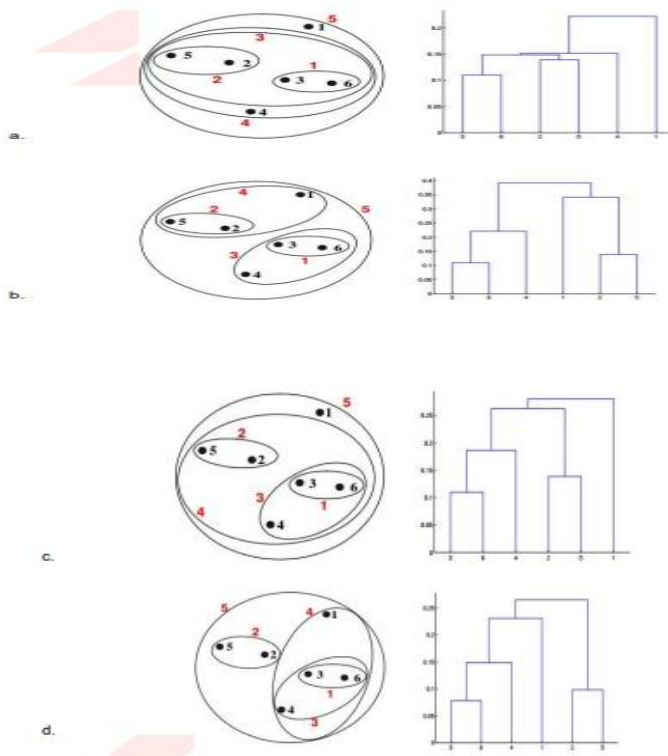
point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

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.



Option: b

13. What is the importance of clustering?

Ans: Clustering is important as, as a data scientist you will not get labelled data in your datasets. Most of the problems are not segregated into classes. Clustering helps to find out the patterns in the data without taking any particular variable in considerations. This is could very well help us to find the relations between the features of the data. After clustering we could treat the records based on its group. We can also find similarly patterns in the records which could be uses for recommendation system applications.

14. How can I improve my clustering performance?

Ans: Clustering is an unsupervised technique. Based on the data, we create cluster and represent into graphical methods for better visualizations and also for formation of optimal clusters. Once the cluster are formed, we perform inferential as well as descriptive statistical analysis on the cluster and find relational patterns between them. Once this is done, we have to present and explain the cluster analysis to the client or the team, this is known as cluster profiling. We create a table of features that each of the cluster's trait. We have to describe the founding to the client of the team in a form of a story and give valuable perks which will help them to work on each cluster with respect to the traits we found in the statistical analysis steps.