

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?
- Biological network analysis
 - Market trend prediction
 - Topic modeling
 - All of the above

Ans : d

2. On which data type, we cannot perform cluster analysis?
- Time series data
 - Text data
 - Multimedia data
 - None

Ans:d

3. Netflix's movie recommendation system uses-
- Supervised learning
 - Unsupervised learning
 - Reinforcement learning and Unsupervised learning
 - All of the above

Ans: c

4. The final output of Hierarchical clustering is-
- The number of cluster centroids
 - The tree representing how close the data points are to each other
 - A map defining the similar data points into individual groups
 - All of the above

Ans: b

5. Which of the step is not required for K-means clustering?
- A distance metric
 - Initial number of clusters
 - Initial guess as to cluster centroids
 - None

Ans: d

6. Which is the following is wrong?
- k-means clustering is a vector quantization method
 - k-means clustering tries to group n observations into k clusters
 - k-nearest neighbour is same as k-means
 - None

Ans:c

7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?

- Single-link
- Complete-link
- Average-link

Options:

- 1 and 2
- 1 and 3
- 2 and 3
- 1, 2 and 3

Ans: d

8. Which of the following are true?
-

MACHINE LEARNING

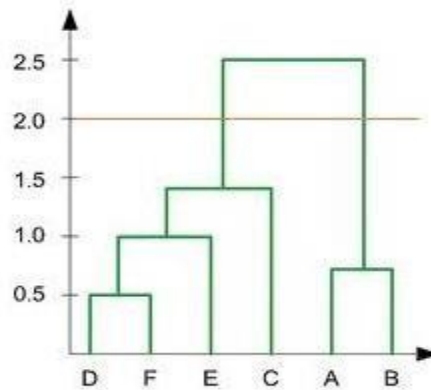
- i. Clustering analysis is negatively affected by multicollinearity of features
- ii. Clustering analysis is negatively affected by heteroscedasticity

Options:

- a. 1 only
- b. 2 only
- c. 1 and 2
- d. None of them

Ans: a

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
- b. 4**
- c. 3
- d. 5

Ans : 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.
 - b. Given a database of information about your users, automatically group them into different market segments.
 - c. Predicting whether stock price of a company will increase tomorrow.
 - d. Given historical weather records, predict if tomorrow's weather will be sunny or rainy.

Ans : c

11. Given, six points with the following attributes:

MACHINE LEARNING

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.

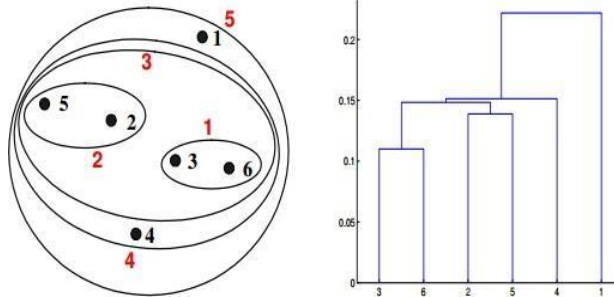
	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points

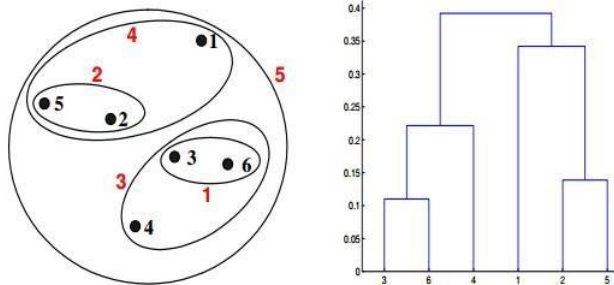
MACHINE LEARNING

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

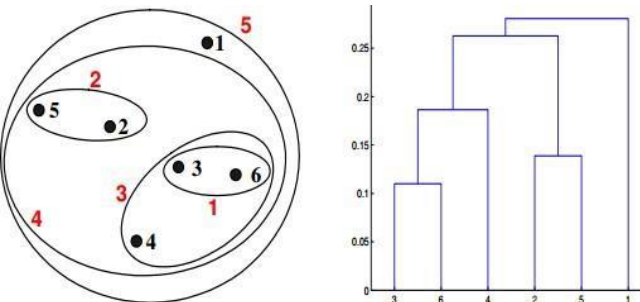
a.



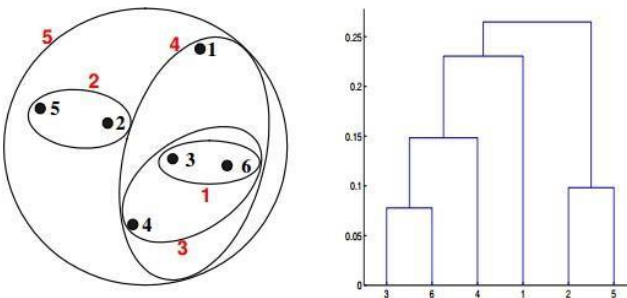
b.



c.



d.



MACHINE LEARNING

Ans: 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

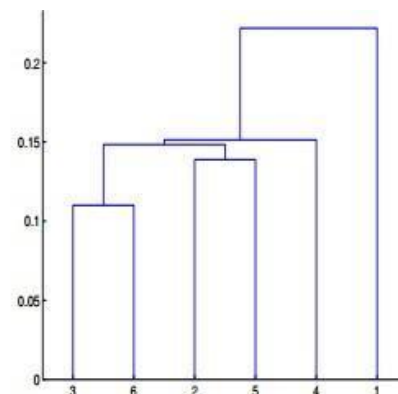
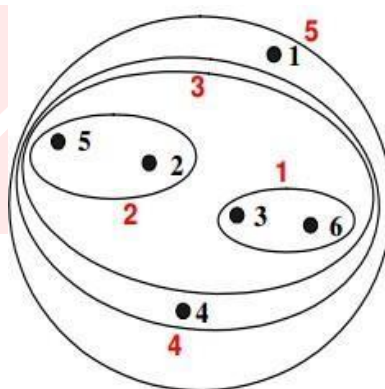
Table : X-Y coordinates of six points.

	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

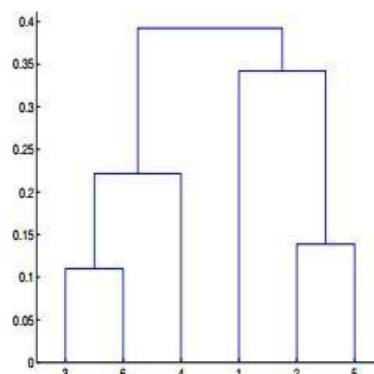
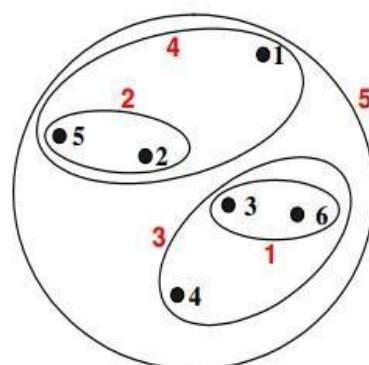
Table : Distance Matrix for Six Points

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

a.

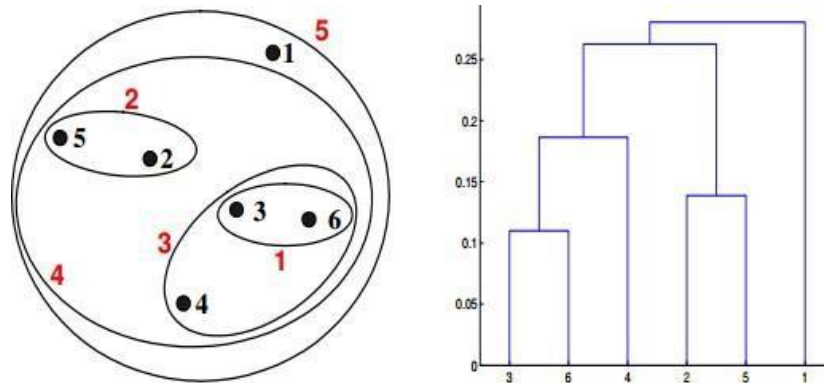


b.

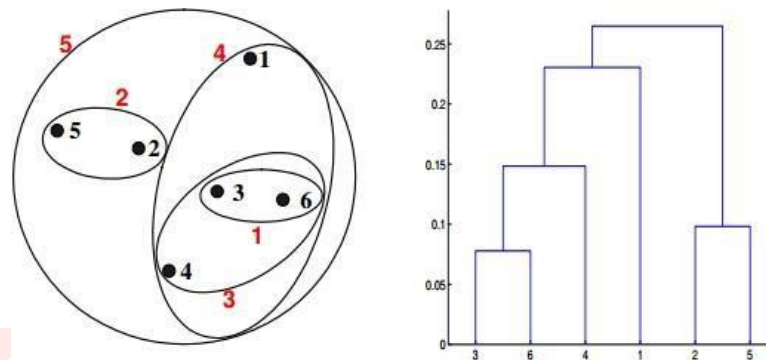


MACHINE LEARNING

c.



d.



Ans : b

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

13. What is the importance of clustering?

Clustering is a machine learning technique, which is used to group the unlabeled dataset. It is a way of grouping of datapoints into different clusters which consists of similar data points. Possible similarities group together as a group and thus forms clusters. Clustering algorithms are essential for data scientists to discover innate groupings among unlabeled and labeled data presets.

14. How can I improve my clustering performance?

Allow max variance in each cluster

Merge neighboring cluster if the resulting cluster' variance is below the threshold

Move some nearby clusters