

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?

- a. Biological network analysis
- b. Market trend prediction
- c. Topic modeling
- d. All of the above

Answer:- (D)

2. On which data type, we cannot perform cluster analysis?

- a. Time series data
- b. Text data
- c. Multimedia data
- d. None

Answer:- (D)

3. Netflix's movie recommendation system uses-

- a. Supervised learning
- b. Unsupervised learning
- c. Reinforcement learning and Unsupervised learning
- d. All of the above

Answer:- (C)

4. The final output of Hierarchical clustering is-

- a. The number of cluster centroids
- b. The tree representing how close the data points are to each other
- c. A map defining the similar data points into individual groups
- d. All of the above

Answer:- (B)

5. Which of the step is not required for K-means clustering?

- a. A distance metric
- b. Initial number of clusters
- c. Initial guess as to cluster centroids
- d. None

Answer:- (D)

6. Which of the following is wrong?

- a. k-means clustering is a vector quantization method
- b. k-means clustering tries to group n observations into k clusters
- c. k-nearest neighbour is same as k-means
- d. None

Answer:- (C)



MACHINE LEARNING

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:

- a. 1 and 2
- b. 1 and 3
- c. 2 and 3
- d. 1, 2 and 3

Answer:- (D)

8. Which of the following are true?

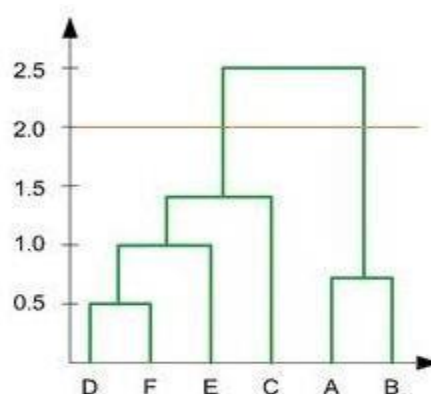
- i. Clustering analysis is negatively affected by multicollinearity of features
- ii. Clustering analysis is negatively affected by heteroscedasticity

Options:

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

Answer:- (D)

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

Answer:- (A)

MACHINE LEARNING

10. For which of the following tasks might clustering be a suitable approach?
- Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.
 - Given a database of information about your users, automatically group them into different market segments.
 - Predicting whether stock price of a company will increase tomorrow.
 - Given historical weather records, predict if tomorrow's weather will be sunny or rainy.
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

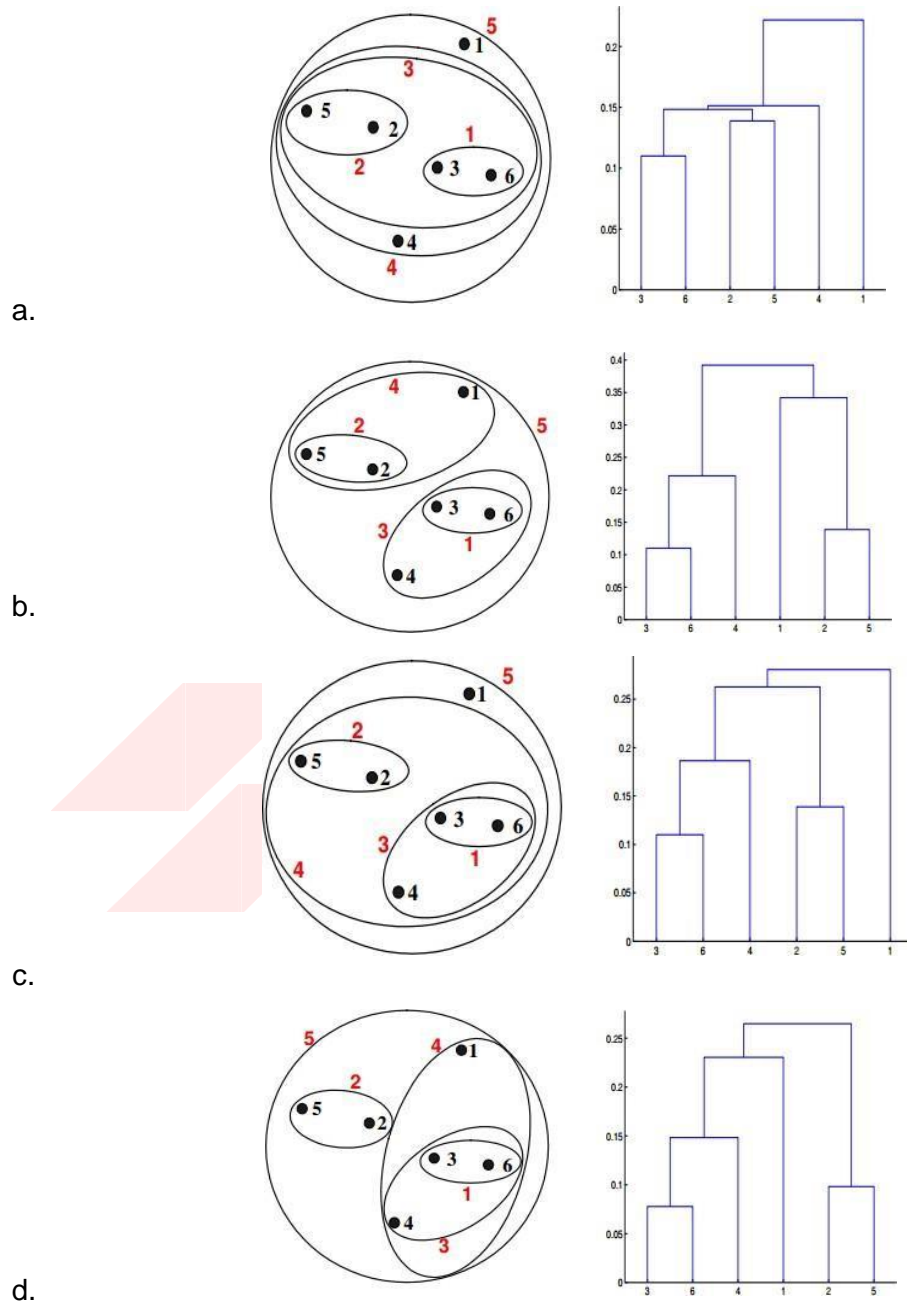
Table : X-Y coordinates of six points.

	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:



Answer:- (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

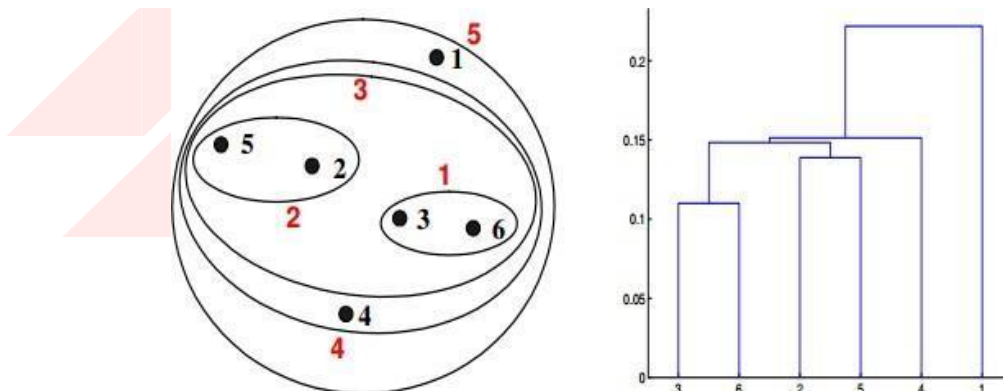
Table : X-Y coordinates of six points.

	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

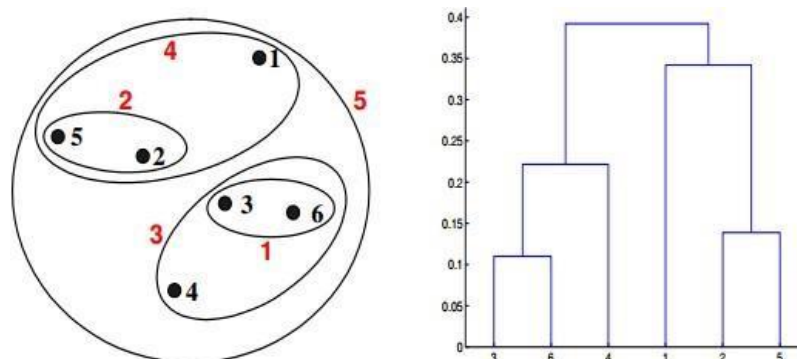
Table : Distance Matrix for Six Points

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

a.

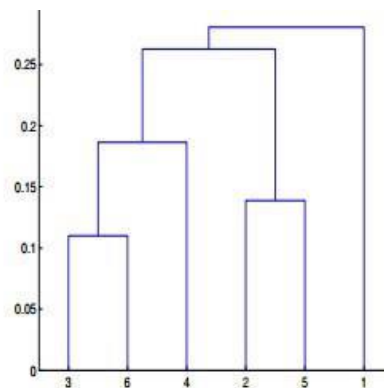
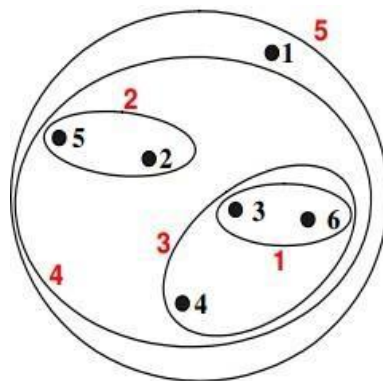


b.

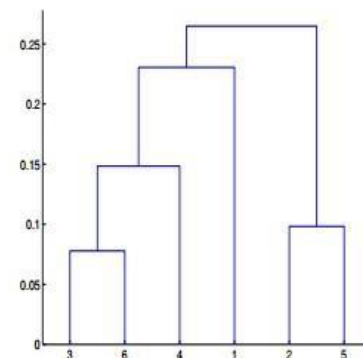
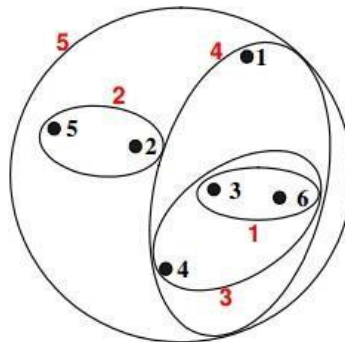


MACHINE LEARNING

c.



d.



Answer:- (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 method of data analysis that groups similar data points together. The importance of clustering lies in its ability to uncover underlying patterns and structures in data that may not be immediately apparent. It can be used for a variety of applications, such as market segmentation, image segmentation, anomaly detection, and text summarization. Additionally, clustering can be a useful tool for reducing the dimensionality of data and making it more manageable for further analysis. Overall, it is an unsupervised learning technique that can be used to extract insights from data without the need for labeled data, or when the labels are not known.

14. How can I improve my clustering performance?

Following ways to improve the performance of a clustering algorithm:

Feature Selection: Selecting the most relevant features for the clustering task can improve the performance by reducing the dimensionality of the data and eliminating noise.

Scaling: Scaling the features can help to ensure that all features are on the same scale, which can help to improve the performance of some clustering algorithms.

Choosing the right number of clusters: A common problem in clustering is determining the number of clusters. A general method is to use elbow method, silhouette method, gap statistic, etc.

Choosing the right clustering algorithm: Different clustering algorithms are suited to different types of data and different problem domains. It's important to choose the algorithm that is most appropriate for

MACHINE LEARNING

your data.

Validation: The performance of a clustering algorithm should be evaluated using a suitable validation method such as adjusted rand index, silhouette score, etc.

Hyperparameter tuning: Some clustering algorithms have various hyperparameters that can be tuned to improve performance.