

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

Ans) D

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

- b. Text data
- c. Multimedia data
- d. None

Ans) 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

Ans) 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

Ans) B

5. Which of the steps 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

Ans) 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

Ans) C

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

Ans) 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

Ans) B

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) A

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) A

11. Given, six points with the following attributes

Ans) B

12. Given, six points with the following attributes:

Ans) C

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

13. What is the importance of clustering?

Cluster analysis is the grouping of objects such that objects in the same cluster are more similar to each other than they are to objects in another cluster. The classification into clusters is done using criteria such as smallest distances, density of data points, graphs, or various statistical distributions. Cluster analysis has wide applicability, including in unsupervised machine learning, data mining, statistics, Graph Analytics, image processing, and numerous physical and social science applications.

14. How can I improve my clustering performance?

There are two important elements in improving the quality of clustering: improving the weights of the features in a document vector and creating a more appropriate distance measure. A good weighting technique can promote the good features of an object, and an appropriate distance measure can help bring similar features together. The next two sections explain how you can create custom feature-selection and distance-measurement classes.