

MACHINE Learning Assignment 3

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) All of the above

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

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

Answer d) None

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) Reinforcement learning and Unsupervised learning

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

Answer d) None

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

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

Answer a) 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
- b. 4
- c. 3
- d. 5

Answer b) 4

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.

Answer d)

11. Given, six points with the following attributes.

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

Answer A)

12. Given, six points with the following attributes:

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

Answer B)

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

13. What is the importance of clustering?

Ans:- Clustering is important in data analysis and data mining applications. It is the task of grouping a set of objects so that objects in the same group are more similar to each other than to those in other groups

14. How can I improve my clustering performance?

Ans:- 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.