MACHINE LEARNING - Answers

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 is 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?

d. 1, 2 and 3

8. Which of the following are true?

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

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:

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.

a.

13. What is the importance of clustering?

Clustering is a crucial part of machine learning and makes the process of developing new machine learning techniques more simpler. It primarily separates multiple unstructured data sets into clusters and aids in the creation of increasingly more clusters based on the shared characteristics found in each cluster.

Making sense of massive quantities of structured and unstructured data and extracting value from them are the goals of clustering and classification algorithms. It only makes sense to try to split the data into some form of logical groupings before attempting to analyse it if you're working with large volumes of unstructured data.

Exploring data with clustering is helpful. Clustering techniques can be used to find natural groupings when there are many cases but no evident divisions. Clustering can also be used to discover homogeneous groups on which to base supervised models as a stage in the pre-processing of data.

14. How can I improve my clustering performance?

Clustering is an unsupervised machine learning methodology that divide data into different groups, or clusters. Hierarchical, density-based, and similarity-based types are just a few of the variations. Each also has a few distinct algorithms connected to it.

By utilising a better initiation method, restarting the algorithm, avoiding imbalanced cluster size, or any combination of these, the K-means clustering algorithm can be considerably improved. When there are overlapping clusters in the data, K-means can enhance the initialization technique's output. When the data is divided into distinct clusters, the effectiveness of k-means solely depends on the quality of the initialization. The initialization using simple furthest point heuristic (Maxmin), reduces the average clustering error of k-means from 15% to 6%.