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?

Ans- b. Market trend prediction

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

Ans - a. Time series data

3. Netflix's movie recommendation system uses

Ans- d. All of the above

4. The final output of Hierarchical clustering is

Ans- c. A map defining the similar data points into individual groups

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

Ans- a. A distance metric

6. Which is the following is wrong?

Ans- b. k-means clustering tries to group n observations into k clusters

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

Ans- a.1 and 2

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

Ans - 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?

Ans- b. 4

10. For which of the following tasks might clustering be a suitable approach?

Ans - a.Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.

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:

Ans- B

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.

Ans- C

13. What is the importance of clustering?

Ans-Clustering is used **to identify groups of similar objects in datasets with two or more variable quantities**. In practice, this data may be collected from marketing, biomedical, or geospatial databases, among many other places.

14. How can I improve my clustering performance?

Ans- Graph-based clustering performance can easily be improved by applying ICA blind source separation during the graph Laplacian embedding step. Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance.