MACHINE LEARNING

1. Which of the following is an application of clustering?

Ans. d. All of the above

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

Ans. d. None

3. Netflix's movie recommendation system uses

Ans. c. Reinforcement learning and Unsupervised learning

4. The final output of Hierarchical clustering is ?

Ans. c. The tree representing how close the data points are to each other

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

Ans. d. None

6. Which is the following is wrong?

Ans. d. None

7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?
i.   Single-link
ii.   ii. Complete-link
iii.   iii. Average-link

Ans. 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

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. b.Given a database of information about your users, automatically group them into different market segments

11. Given, six points with the following attributes:

Ans. a

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

Ans. b

13. What is the importance of clustering?

Ans. Clustering or cluster analysis is a machine learning technique, which groups the unlabelled dataset. They can cluster different customer types into one group based on different factors, such as purchasing patterns. The factors analysed through clustering can have a big impact on sales and customer satisfaction, making it an invaluable tool to boost revenue, cut costs, or sometimes even both.

14. How can I improve my clustering performance?

Ans. Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance. Surprisingly for some cases, high clustering performance can be achieved by simply performing K-means clustering on the ICA components after PCA dimension reduction on the input data.