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)

a. Supervised learning

4. The final output of Hierarchical clustering is

Ans)

d. All of the above

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-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?

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)

   a. 2

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)

a)

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** (clusters).

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.