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?

a. Biological network analysis

b. Market trend prediction

c. Topic modelling

d. All of the above

Ans: d

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

a. Time series data

b. Text data

c. Multimedia data

d. None

Ans: d

3. Netflix's movie recommendation system uses

a. Supervised learning

b. Unsupervised learning

c. Reinforcement learning and Unsupervised learning

d. All of the above

Ans: c

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

Ans: b

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

Ans: d

6. Which is 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

Ans: c

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

Ans: d

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

Ans: a

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

Ans: b

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

Ans: b

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: c

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: d

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 a widely used unsupervised learning technique that allows us to find hidden patterns or relationships between the data points based on the common attributes in the data. 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. 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.

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

14. How can I improve my clustering performance?

Ans: There are several ways to improving the performance of clustering are follows:

- Improving clustering performance using independent component analysis and unsupervised feature learning.
- Improving Clustering Performance Using Feature Weight Learning.

- Improving Clustering Performance by Using Feature Selection and Extraction Techniques.
- improve the clustering performance for high dimensional data by Principal Component Analysis.
- Improve the clustering performance using Dataset transformation.
- Writing a custom distance measure.
- Improving document vector generation.