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

a. Biological network analysis

b. Market trend prediction

c. Topic modeling

d. All of the above

**Ans: D all of the above**

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

a. Time series data

b. Text data

c. Multimedia data

d. None

**Ans**: **A. Time series data**

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. Reinforcement learning and Unsupervised learning**

**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: D.** **All of the above**

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: C.** **Initial guess as to cluster centroids**

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** . **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?**

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 . 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 Options:

a. 1 only

b. 2 only

c. 1 and 2

d. None of the

**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?

a. 2

b. 4

c. 3

d. 5

**Ans: A 2**

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:**

**Ans: A**

12. **Given, six points with the following attributes:**

**Ans: B**

13**. What is the importance of clustering?**

**Clustering:**

A cluster is the collection of data objects which are similar to each other within the same group.

The data objects of a cluster are dissimilar to data objects of other groups or clusters.

**Importance of clustering:**

* Having clustering methods helps in restarting the local search procedure and remove the inefficiency. In addition, clustering helps to determine the internal structure of the data.
* This clustering analysis has been used for model analysis, vector region of attraction.
* Clustering helps in understanding the natural grouping in a dataset. Their purpose is to make sense to partition the data into some group of logical groupings.
* Clustering quality depends on the methods and the identification of hidden patterns.
* They play a wide role in applications like marketing economic research and weblogs to identify similarity measures, Image processing, and spatial research.
* They are used in outlier detections to detect credit card fraudulence.

14**. How can I improve my clustering performance?**

**Ans:** We can iteratively apply to improve the quality of our clustering. First, perform a visual check that the clusters look as expected. Improving clustering performance using independent component analysis and unsupervised feature learning Abstract.