MACHINE LEARNING 3

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

d. All of the above

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

a. Time series data

3. Netflix's movie recommendation system uses-

c. Reinforcement learning and Unsupervised learning

4. The final output of Hierarchical clustering is-

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

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

d. None

6. Which is the following is wrong?

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

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

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

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

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

11. Which of the following clustering representations and dendrogram depicts the use of MIN or Single link proximity function in hierarchical clustering:

a

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

b

13. What is the importance of clustering?

Clustering methods help in grouping the data points into clusters, using the different techniques are used to pick the appropriate result for the problem, these clustering techniques helps in grouping the data points into similar categories, and each of these subcategories is further divided into subcategories to assist the exploration of the queries output.

Importance:

1. 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.
2. This clustering analysis has been used for model analysis, vector region of attraction.
3. 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.
4. Clustering quality depends on the methods and the identification of hidden patterns.
5. They play a wide role in applications like marketing economic research and weblogs to identify similarity measures, Image processing, and spatial research.
6. They are used in outlier detections to detect credit card fraudulence.

14. How can I improve my clustering performance?

First, perform a visual check that the clusters look as expected, and that examples that you consider similar do appear in the same cluster. Then check these commonly-used metrics as described in the following sections:

- Cluster cardinality
- Cluster magnitude
- Performance of downstream system

Cluster cardinality

Cluster cardinality is the number of examples per cluster. Plot the cluster cardinality for all clusters and investigate clusters that are major outliers.

Cluster magnitude

Cluster magnitude is the sum of distances from all examples to the centroid of the cluster. Similar to cardinality, check how the magnitude varies across the clusters, and investigate anomalies.

Performance of Downstream System

Since clustering output is often used in downstream ML systems, check if the downstream system's performance improves when your clustering process changes. The impact on your downstream performance provides a real-world test for the quality of your clustering. The disadvantage is that this check is complex to perform