Machine learning

Answer 1-

a -2

Answer -2

d- 1,2 and 4

Answer-3

d- formulating the clustering problem

Answer-4

A- Euclidean distance

Answer -5

b- Divisive clustering

Answer-6

d- all of the above

Answer-7

a-divide the data points into groups

Answer-8

b- unsupervised learning

Answer 9

d- all of the above

Answer-10

A- K-means clustering algorithm

Answer-11

D- all of the above

Answer-12

A -labeled data

Answer-13

Cluster analysis that tries to spot structures among the information. Cluster analysis is additionally known as segmentation analysis or taxonomy analysis.

In this form of formula, the information divide or segregate the information into “K disjoint clusters”. You need to decide on the amount of clusters(K) in line with your information. Cluster centers or centroids represent every cluster.

the formula works like :

Step 1: 1st of all, opt for the cluster centers or the amount of clusters.

Step 2: Delegate every purpose to its nearest cluster center by shrewd the geometrician distance.

Step 3:The cluster centroids are going to be optimized supported the mean of the points appointed thereto cluster.

Step 4: Once we have a tendency to see that the cluster centroids don't seem to be creating several movements or moving tiny distances, we will safely say that the K-means cluster has converged.

Answer 14

A clustering-quality measure (CQM) is **a function that, given a data set and its par- tition into clusters, returns a non-negative real number representing how strong or conclusive the clustering is**.

We can measure the quality of clustering by using the Dissimilarity/Similarity metric in most situations. But there are some other methods to measure the Qualities of Good Clustering if the clusters are alike.

1. **Dissimilarity/Similarity metric:** The similarity between the clusters can be expressed in terms of a distance function, Distance functions are different for various data types and data variables. Distance function measure is different for continuous-valued variables, categorical variables, and vector variables.

2. **Cluster completeness:** Cluster completeness is the essential parameter for good clustering, if any two data objects are having similar characteristics then they are assigned to the same category of the cluster according to ground truth. Cluster completeness is high if the objects are of the same category.

3. **Ragbag:** In some situations, there can be a few categories in which the objects of those categories cannot be merged with other objects. Then the quality of those cluster categories is measured by the Rag Bag method. According to the rag bag method, we should put the heterogeneous object into a rag bag category.

4. Small cluster preservation: If a small category of clustering is further split into small pieces, then those small pieces of cluster become noise to the entire clustering and thus it becomes difficult to identify that small category from the clustering. The small cluster preservation criterion states that are splitting a small category into pieces is not advisable and it further decreases the quality of clusters as the pieces of clusters are distinctive.

Answer 15

Cluster analysis is groups the unlabelled dataset. It is a type of unsupervised machine learning algorithm. It is used for data that do not have any proper labels.

Its main goal is to groups objects .

- Types of clustering
- Hierarchical Cluster Analysis

- Centroid-based Clustering
- Distribution-based Clustering
- Density-based Clustering