MACHINE LEARNING

1. B
2. D
3. D
4. A
5. B
6. D
7. A
8. B
9. D
10. A
11. D
12. A
13. How is cluster analysis calculated?

    Cluster Analysis is calculated using Euclidian Distance. Euclidian distance is a distance between data points (observations). This distance is used to separate out data points into different groups. Take a case if observations are having values A (2, 3), B (3, 5) and C (5, 8). The Pythagoras theorem is basically used to calculate distance between two data points. If there are two centre points then distance between each centre point and observed point are calculated and tabulated in the tables are mentioned in below as for one centre point i.e. A. The matrix distances are as

	X	Y	Distance from Point A or any other point (centre point)	Second centre point (7,4)
A	2	3	0	1
B	3	5	2.34606	4.12310
C	5	8	5.83095	4.47214

$$ab = \sqrt{(2-3)^2 + (3-5)^2}$$

$$ab = 2.23606$$

Similarly distances are calculated for bc and ac. Then find out the proximity matrix which is nothing but a symmetric matrix in which upper triangular and lower triangular values are identical. Based on the minimum distance and number of groups, clusters are created.

14. How is cluster quality measured?

    The quality of cluster is measured using different techniques. If the data is evaluated by itself and created cluster then it is called as Extrinsic Evaluations.

1. Rand Index: Rand index is used to measure the similarity index of the clusters created by the algorithm
2. Mutual Information Based Score: It measures agreements of two clusters ignoring permutations. The standard MI-based value is some generalised mean of each clustering's entropies.
3. Homogeneity, completeness and V-measure: Homogeneity is related to cluster contains only one members of single class. Completeness is nothing more than assigning the same class to every member of a given class. The symmetry of the V-measure allows for the evaluation of the agreement between two independent assignments on the same dataset.

Measures that are intrinsic that don't need ground truth labels. The Calinski-Harabasz Index, Davies-Bouldin Index, and Silhouette Coefficient are a few examples of clustering performance indicators.

15. What is cluster analysis and its types?

Cluster analysis in statistics is a method to organize data by clustering data points in a particular cluster. Rightly put, cluster analysis is a way of putting data points with similar characteristics in one group so that they differ from other data points of other clusters. It must be noted that the level of similarity between two data points in one cluster is maximum while the similarity between two data points of different clusters is minimum. Broadly, there are 2 types of cluster analysis methods. On the basis of the categorization of data sets into a particular cluster, cluster analysis can be divided into 2 types - hard and soft clustering. They are as follows -

1. Hard Clustering

The data scientist can cluster datasets so that each dataset is included in only one of the total number of clusters accessible. This demonstrates the importance of carefully classifying the dataset in order to properly arrange and classify the data. A clustering algorithm, for example, divides data points into clusters based on their similarity to one another. However, there is no other reason to compare it to other groups' databases.

2. Soft Clustering

The second group includes soft clustering. It differ from hard clustering in that they only allow one data point to be attached to one cluster at a time. In the case of soft clustering, data points can include more than one cluster at the same time. The residue separation distinguishes the thin dough. The fuzzy group technique in machine learning is a fuzzy algorithm for classifying data into fuzzy groups.