MACHINE LEARNING

Q.1] B

Q.2] D

Q.3] D

Q.4] A

Q.5] C

Q.6] D

Q.7] A

Q.8] B

Q.9] A

Q.10] A

Q.11] D

Q.12] B

Q.13] How is cluster analysis calculated?

ANS. = Cluster analysis, also known as clustering, is a technique used to group similar data points together based on their features. The process of cluster analysis involves several steps:

1. Data preparation: The first step is to prepare the data for analysis by cleaning, transforming, and normalizing the data.

2. Selection of variables: The next step is to select the variables on which clustering will be based. The choice of variables can greatly affect the outcome of the clustering and the interpretation of the results.

3. Selection of clustering method: The next step is to choose the appropriate clustering method based on the characteristics of the data and the goals of the analysis. There are several different types of clustering methods, such as k-means, hierarchical, and density-based clustering.

4. Clustering: Once the data is prepared and the clustering method is chosen, the clustering algorithm is applied to the data to group the data points into clusters.

5. Evaluation: The last step is to evaluate the quality of the clusters and determine whether they are meaningful and useful for the goals of the analysis. This may involve measuring the internal cohesion of the clusters and the separation between clusters.

6. Interpretation: Finally, the characteristics of the clusters and the patterns in the data that led to their formation are understood.

Q.14] How is cluster quality measured ?

ANS. = Cluster quality can be measured in several ways, depending on the characteristics of the data and the goals of the analysis. Some common methods include:

1. Internal evaluation: This method involves measuring the internal cohesion of the clusters, which is a measure of how similar the data points are within each cluster. This can be done using measures such as the mean distance between data points within a cluster, or the silhouette coefficient which measures the similarity of a data point to its own cluster compared to other clusters.

2. External evaluation: This method involves measuring the separation between clusters and the degree to which they correspond to predefined classes or categories, if they exist. This can be done using measures such as the Rand index or the Jaccard coefficient.

3. Scattering measures: This method is used to evaluate the quality of the clusters based on the distance between the cluster centroid and the data points assigned to it. The most commonly used measures are Within-Cluster-Sum-of-Squares (WCSS) and the Davies-Bouldin Index (DBI)

4. Visualization: This method involves plotting the data points and the clusters in a two-dimensional space, which can help to visually assess the quality of the clustering and identify any issues or patterns in the data.

Q.15] What is cluster analysis and its types ?

ANS. = Cluster analysis, also known as clustering, is a technique used to group similar data points together based on their features. Clustering algorithms group similar data points together based on their features and do not require labeled data. The algorithm attempts to find natural groupings or clusters in the data based on similarity. Clustering is an unsupervised learning technique that can be used for a variety of applications, such as market segmentation, image segmentation, anomaly detection, and text mining.

There are several different types of clustering methods, including:

1. Centroid-based clustering: This method involves defining a centroid for each cluster, and assigning data points to the cluster whose centroid is closest to it. The most common centroid-based clustering method is k-means.

2. Hierarchical clustering: This method involves creating a hierarchy of clusters, where each node is a cluster. The most common hierarchical clustering methods are Agglomerative and Divisive.

3. Density-based clustering: This method involves identifying clusters as dense regions of the data. The most common density-based clustering method is DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

4. Distribution-based clustering: This method involves modeling the probability distribution of the data and using it to identify clusters. The most common distribution-based clustering method is Gaussian Mixture Model (GMM)

5. Subspace and Projection-based clustering: This method involves identifying clusters based on subspace or projection of the data. The most common subspace and projection-based clustering method is CLIQUE (Clustering in Quest)

6. Graph-based clustering: This method involves representing the data points as nodes in a graph and grouping them based on the edges connecting them.