MACHINE LEARNING

Q.1] B

Q.2] D

Q.3] B [FALSE]

Q.4] D

Q.5] C

Q.6] B

Q.7] A

Q.8] D

Q.9] A

Q.10] D

Q.11] D

Q.12] Is K sensitive to outliers?

ANS= It depends on the type of algorithm and the specific implementation. K-means is a centroid-based clustering algorithm which uses mean-based approach to calculate the centroids, it can be sensitive to outliers in the data. Outliers can have a large impact on the mean, pulling the centroid away from the bulk of the data and causing the algorithm to form clusters that are not representative of the underlying structure of the data.

The K-Means clustering is sensitive to outliers, because a mean is easily influenced by extreme values.

Q.13] Why is K means better?

ANS= 1] Simplicity: K-means is a simple and easy-to-implement algorithm. The algorithm only requires specifying the number of clusters, k, and the algorithm proceeds to group the observations into k clusters.

2] Scalability: K-means can handle large datasets with a relatively fast computational time. The algorithm is linear in the number of observations and the number of variables, making it well-suited for high-dimensional data.

3] Intuitive: K-means algorithm is based on the idea of centroids, or central points, which makes it easy to understand and interpret the resulting clusters.

4] Flexibility: K-means can be used with different distance metrics, such as Euclidean, Manhattan, or Cosine similarity which makes it flexible and adaptable to different types of data.

5] Generalization: K-means can be used to cluster data in any domain, it's widely used in various fields such as image processing, bioinformatics, and text mining.

Q.14] Is K means a deterministic algorithm?

ANS= The basic k-means clustering is based on a non-deterministic algorithm. This means that running the algorithm several times on the same data, could give different results. K-means is a deterministic algorithm in the sense that given the same input data and the same initial centroids, it will always produce the same output clusters. However, the algorithm uses a random initialization method to select the initial centroids, and the final clustering solution depends on the initial centroids selected. Because the initial centroids are chosen randomly, it is possible that the algorithm will converge to different solutions when run multiple times. To overcome this issue, different techniques like k-means++ can be used to choose the initial centroids, which reduces the chance of getting stuck in a local minimum, but still there is no guarantee that running the algorithm multiple times will always produce the same clustering results.