MACHINE LEARNING

Assignment - 1

Answers:

1. b) 4
2. c) 2 and 4
3. d) formulating the clustering problem
4. a) Euclidean distance
5. c) Agglomerative clustering
6. d) All answers are correct.
7. a) Divide the data points into groups
8. b) Unsupervised learning
9. a) K- Means clustering
10. a) K-means clustering algorithm is most sensitive to outliers.
11. d) All of the above.
12. b) Unlabeled data

13. Cluster analysis is a method to group similar data points together into clusters. It involves using algorithms that examiine the characterictics of the data points and then group them based on those characteristics. The process is done in a way that data points within a cluster are more similar to each other than to data points in other clusters. It is a technique used in unsupervised learning.

14. Cluster quality is measured by evaluating how well the data points are grouped together within the clusters. Evaluation metrics like silhouette score, Davies-Bouldin index, Rand index, Normalized Mutual Information, and Fowlkis-Mallows index are commonly used to assess the quality of the clustering. These metrics compare the similarity of data points within a cluster with the similarity of data points in other clusters.

15. Cluster analysis is a method to group similar data points together into clusters. It is a technique used in unsupervised learning; no predefined labels are needed.
    Types of cluster analysis:
    i.   **Hierarchical Clustering**: Groups data points by building hierarchy of clusters
    ii.   **Centroid-based Clustering**: Divides data into clusters based on center points
    iii. **Density-based Clustering**: Groups data points based on density
    iv.   **Distribution-based Clustering**: Clusters data points based on probability distribution
    v.   **Spectral Clustering**: Clusters data points based on similarity matrix
    vi.   **Grid-based Clustering**: Clusters data points by dividing feature space into cells.

    The choice of method depends on the nature of data and the desired outcome of the analysis.