

MLAS-1

1. We can Create 4 Clusters (Ans—B)
2. K-Means clustering fails to give good results when Data contains outliers, Data points with different densities and Data points follow non-convex shapes (Ans—D).
3. formulating the clustering problem (Ans—D).
4. Euclidean distance is most commonly used measure of similarity (Ans—A).
5. Divisive is a clustering procedure where all objects start out in one giant cluster (Ans—B).
6. All are correct (Ans—D).
7. Divide the data points into groups (Ans—A).
8. Clustering is a Unsupervised Learning (Ans—B).
9. K- Means clustering (Ans—A).
10. K-means clustering algorithm (Ans—A).
11. All are correct (Ans—D).
12. For clustering, we do not require Labelled Data (Ans—A).
13. Cluster Analysis process requires to quantify the degree of similarity, or dissimilarity, between observations. The results of the analysis is strongly dependent on the kind of the used similarity metric. Various Algorithms are used to find cluster analysis . Three important factors by which clustering can be evaluated are:- (a) Clustering tendency (b) Number of clusters, k (c) Clustering quality etc.
14. We have a few methods to choose from for measuring the quality of a clustering. In general, these methods can be categorized into two groups according to whether ground truth is available (i) **Extrinsic** Measures which require ground truth labels. Examples are Mutual information based scores, Homogeneity, Completeness and V-measure.
(ii) **Intrinsic** Measures that does not require ground truth labels. Some of the clustering performance measures are Silhouette Coefficient etc.
15. Cluster Analysis divide Data into groups that are meaningful and useful. It is also useful in data mining, Machine Learning, Pattern recognition etc. Partitioning and segmentation are also synonym of clustering.
Types of Cluster Analysis -:
 - a) Agglomerative Hierarchical clustering
 - b) K-means
 - c) DBSCAN