

Machine Learning Assignment-1

A.1) Option b- 4 is the most appropriate no. of clusters for the data points represented by the dendrogram.

A.2) Option d- 1,2 and 4, The K-Means clustering fail to give good results when **Data points with outliers,** **Data points with different densities** and **Data points with non-convex shapes.**

A.3) Option d- formulating the clustering problem.

A.4) Option a- Euclidean distance

A.5) Option b- Divisive clustering

A.6) Option d- All answers are correct

A.7) Option a- Divide the data points into groups

A.8) Option b- Unsupervised learning

A.9) Option d- All of the above

A.10) Option a- K-means clustering algorithm

A.11) Option d- All of the above

A.12) Option a- Labeled data

A.13) It is calculated by measuring the distance between each data point and its centroid, squaring this distance, and summing these squares across one cluster.

A.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. Here, *ground truth* is the ideal clustering that is often built using human experts.

If ground truth is available, it can be used by **extrinsic methods**, which compare the clustering against the group truth and measure. If the ground truth is unavailable, we can use **intrinsic methods**, which evaluate the goodness of a clustering by considering how well the clusters are separated. Ground truth can be considered as supervision in the form of “cluster labels.” To measure a cluster's quality within a clustering, we can compute the average silhouette coefficient value of all objects in the cluster. To measure the quality of a clustering, we can use the average silhouette coefficient value of all objects in the data set.

A.15) Cluster Analysis is a data analysis technique that explores the naturally occurring groups within a data set known as clusters. Cluster analysis doesn't need to group data points into any predefined groups, which means that it is an unsupervised learning method. **Types of Cluster Analysis**-The types of cluster analysis are Centroid Based/ Partition Clustering, Hierarchical Based Clustering, Distribution Based Clustering, Density-Based Clustering, and Fuzzy Based Clustering.