Machine Learning Assignment - 1

1. B
2. D
3. D
4. A
5. B
6. C
7. A
8. C
9. A
10. A
11. D
12. A
13. Cluster analysis can be calculated using K-Means and K-Medoids:

    K-Means: This algorithm establishes the presence of clusters by finding their centroid points. A centroid point is the average of all the data points in the cluster. By iteratively assessing the Euclidean distance between each point in the dataset, each one can be assigned to a cluster. The centroid points are random to begin with and will change each time as the process is carried out.

    K-Medoids: This works in a similar way to k-means, but rather than using mean centroid points which don't equate to any real points from the dataset, it establishes medoids, which are real interpretable data-points.K-medoids offers an advantage for survey data analysis as it is suitable for both categorical and scalar data. This is because rather than measuring Euclidean distance between the medoid point and its neighbors, the algorithm can measure distance in multiple dimensions, representing a number of different categories or variables.
14. Two methods are there for measuring clustering quality: Extrinsic and Intrinsic

    Extrinsic: Supervises i.e., the ground truth is available. Compare a clustering against the ground truth using certain clustering quality measure.

    Ex. Purity, precision and recall metrics, normalized mutual information.

    Intrinsic: Unsupervised i.e., the ground truth is unavailable. Evaluate the goodness of a clustering by considering how well the clusters are separated, and how compact the clusters are.

    Ex. Silhouette coefficient.
15. Cluster analysis or clustering is the task of grouping a set of objects in the same group are more similar in each other than to those in other groups.

    Types:

    Hierarchical cluster analysis:

    In this method, first, a cluster is made and then added to another cluster (the most similar and closest one) to form one single cluster. This process is repeated until all subjects are in one cluster. This particular method is known as Agglomerative method. Agglomerative clustering starts with single objects and starts grouping them into clusters.

    The divisive method is another kind of Hierarchical method in which clustering starts with the complete data set and then starts dividing into partitions.

Machine Learning Assignment - 1

Centroid based clustering:

In this type of clustering, clusters are represented by a central entity, which may or may not be a part of the given data set. K-Means method of clustering is used in this method, where k are the cluster centers and objects are assigned to the nearest cluster centres.

Distribution based clustering:

It is a type of clustering model closely related to statistics based on the modals of distribution. Objects that belong to the same distribution are put into a single cluster.This type of clustering can capture some complex properties of objects like correlation and dependence between attributes.

Density based clustering:

In this type of clustering, clusters are defined by the areas of density that are higher than the remaining of the data set. Objects in sparse areas are usually required to separate clusters.The objects in these sparse points are usually noise and border points in the graph.The most popular method in this type of clustering is DBSCAN.