ASSIGNMENT-1

MACHINE LEARNING

1. b)
2. d)
3. d)
4. a)
5. b)
6. d)
7. a)
8. b)
9. d)
10. a)
11. d)
12. a)
13. A cluster is a collection of homogenous data objects. There are mainly 2 types of cluster analysis, flat cluster which includes K-means cluster analysis and hierarchical cluster analysis which is the more common type used. Now to calculate hierarchical cluster analysis firstly we have to choose various clusters and then by choosing the base by which we have to group our clusters we start grouping them. For example, if we have n number of clusters initially, every time we group two clusters there will be n-1 number of clusters. We keep on grouping them till we reach to a single group that contains all the clusters. Now we calculate the distances and choose the solution by deciding on the right number of clusters.
14. A good cluster is the one that is more compact, the one that is not scattered and contains no outliers and the one whose distance is more from the other clusters. In short intra-cluster distances have to be minimized and the inter-cluster distances have to be maximized. To measure the cluster quality

there are two methods, extrinsic which is supervised and intrinsic which is unsupervised. In case of extrinsic measure, the ground truth is available and it compares the cluster against the ground truth using certain clustering quality measures like precision and recall metrics and normalized mutual information. On the other hand in the case of intrinsic measure the ground truth is unavailable and it evaluates the goodness of a clustering by considering how well the clusters are separated and how compact he clusters are for which it uses the silhouette coefficient measure.

15. Cluster analysis is a data mining technique whose main objective is to group objects which are homogenous. It is the main step of statistical data analysis. Cluster Analysis is the procedure to find alike groups of items in order to form clusters. It is also an unsupervised machine learning-based algorithm that mainly acts on unlabelled data. For example, there is a data of different types of animals. Now it's our work to form clusters among them by separating clusters of dog breeds, cat breeds, aquatic and other non-aquatic animals. They all link up to form a single cluster namely Animals which will contain various clusters of different species of animals. There are different types of clustering which includes:
    - K-means clustering
    - Partitioning method
    - Hierarchical clustering method
    - Density based clustering method
    - HDBSCAN
    - DBSCAN
    - Model based clustering method