MACHINE LEANING- WORKSHEET – 1

1. Question -1 - Answer - **B**
2. Question -2 - Answer - **D**
3. Question -3 - Answer - **D**
4. Question -4 - Answer - **A**
5. Question -5 - Answer **- C**
6. Question -6 - Answer - **D**
7. Question -7 - Answer - **A**
8. Question -8 - Answer - **B**
9. Question -9 - Answer – **A**
10. Question -9 - Answer - **A**
11. Question -9 - Answer - **D**
12. Question -9 - Answer – **A**

13. There are basic 3 steps :

    1. Calculate the distance.
    2. Link the clusters.
    3. Choose a solution by selecting the right number of clusters.

14. **Measures for Quality of Clustering :**

If all the data objects in cluster are highly similar than the cluster has high quality. We can measure all the quality of clustering by using the similarity metric in most situations. But there are some other methods to measure to qualities of good cluster clustering if the clusters are alike.

1. **Dissimilarity/Similarity Metric**: The similarity between the clusters can be expressed in terms of distance function, which is presented by D. Distance

functions or different for various data types and data variables. Distance function measure is different from continuous valued variable, categorical well variables and vector arrivals.

2. **Cluster Completeness:** Cluster Completeness is the essential parameters for good clustering. If any two data objects are having similar characteristics then they assigned to the category of cluster.

3. Ragbag: In some situations, there can be a few categories in which the objects of those categories cannot be merged with other objects. Then the quality of those cluster categories is measured by the Rag Bag method. According to the rag bag method, we should put the heterogeneous object into a rag bag category.

4. Small cluster preservation: If a small category of clustering is further split into small pieces, then those small pieces of cluster become noise to the entire clustering and thus it becomes difficult to identify that small category from the clustering. The small cluster preservation criterion states that are splitting a small category into pieces are not advisable and it further decreases the quality of clusters as the pieces of clusters are distinctive.

15. CLUSTER ANALYSIS :

1. Cluster Analysis Group data objects based only on information found in the data that describes the object and their relationship.
2. The goal is that the object within a group be similar to one another and different from the objects in other groups.
3. The greater the similarity within a group and the greater the differences between groups, the better or more distinct the clustering.

TYPES OF CLUSTER ANALYSIS:

1. Hierarchical Cluster Analysis.

2. Centroid based Clustering.
3. Distribution Based clustering.
4. Density Based Clustering.

1. **HIERACRVHICAL CLUSTER ANALYSIS**: A Cluster is made and then added to another cluster to form one single cluster. This 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.

2. **CENTROID BASED CLUSTERING:** In this type of clustering, cluster 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 is the cluster centers and objects are assigned to the nearest cluster centers.

3. **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 s single cluster. This type of clustering can capture some complex properties of objects like correlation and dependence between attributes

4. **DENSITY BASED CLUSTERING:** In this type of clustering, cluster are defined by the areas of density that are higher than the remaining of the data set. Objects in spares areas are usually required to separate cluster. The objects in these sparse points are usually noise and boarder points in the graph.