Q.1Ans :- Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of exploratory data analysis, and a common technique for statistical data analysis, used in many fields, including pattern recognition, Image analysis, information retrieval, bioinformatics, data compression, computer graphics and machine learning.

The hierarchical cluster analysis follows three basic steps :

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

Computing accuracy for clustering can be done by reordering the rows (or columns) of the confusion matrix so that the sum of the diagonal values is maxima.

Q.2 Ans :- To measure the quality of a clustering, we can use the average silhouette coefficient value of all objects in the data set.

There are some techniques to measure the quality of clusters formed after undergoing clustering techniques.

1) Clustering

2) Data Visualization
3) Advanced Statistical Analysis
4) Cluster Analysis
5) Clustering Algorithms
6) Computational Data Mining

Q.3 Ans :- ***Clustering :-*** *In this technique, which groups the unlabeled dataset.*
A way of grouping the data points into different clusters, consisting of similar data points.
The objects with possible similarities remain in a group that has less or no similarities with another group.
It is used by Amazon, Netflix in its recommendation system to provide recommendation as per the past search of products , movies respectively.

Types of clustering methods

1) **Partitioning Clustering**
2) **Density Based Clustering**
3) **Distribution Model Based Clustering**
4) **Hierarchical Clustering**
5) **Fuzzy Clustering**

1)Partitioning Clustering:-*It divides the data into non-hierarchical groups. It is also known as centroid based method.*

Eg. K-means clustering algo.

2) Density Based:- *It connects highly dense area into cluster, and arbitrarity shaped distribution are formed as long as dense region can be connected.*

3) Distribution Model Based Clustering:- *The data is divided based on probability of how a dataset belongs to particular distribution. The grouping is done by assuming some distribution commonly Gaussian Distribution.*

4) Hierarchical Clustering *:- It can be used as alternative for partitioned clustering as there is no requirement of pre-specifying no of clusters to be created. Dataset is divided into clusters to vreate dendrogram.*

5) Fuzzy Clustering:- *It type of soft method in which data object may belong to more then one cluster. Each dataset has a set of membership coefficients, which depend on degree of membership to be in a cluster.*