

MACHINE LEARNING- I

Q.1) b

Q.2) d

Q.3) d

Q.4) a

Q.5) b

Q.6) d

Q.7) a

Q.8) b

Q.9) d

Q.10) a

Q.11) d

Q.12) a

Q.13) Cluster Analysis is calculated by using following steps:

1. Select the number of clusters (K)
2. Randomly select a number of data points that matches the number of clusters
3. Measure the distances between each point to its initial cluster
4. Assign each datapoint to its nearest initial cluster
5. Repeat the calculations for each point
6. Calculate the mean of each cluster
7. Assign the mean as the new cluster centroid
8. Measure each point to the new cluster centroid
9. Redefine clusters and assign the new mean as the next cluster centroid
10. Repeat process until convergence

Q.14) Cluster Quality can be measured using:

1. Dissimilarity/Similarity Metric: The similarity between the clusters can be expressed in terms of a distance function, which is represented by $d(i, j)$. Distance functions are different for various data types and data variables.

2. Cluster Completeness: Cluster completeness is the essential parameter for good clustering, if any two data objects are having similar characteristics, then they are assigned to the same category of the cluster according to ground truth.

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.

Q.15) Cluster analysis is a multivariate data mining technique whose goal is to groups objects (e.g., products, respondents, or other entities) based on a set of user selected characteristics or attributes.

Types of Cluster Analysis:

- 1. 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.
- 2. 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.
- 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 a single cluster.