MACHINE LEARNING

ANSWERS :-

Assignment :1

- The best choice for the above cluster is b) 4 .
- K-Means clustering will fail to give good results when the data set contains the outliers ,different densities, non- convex shapes so the answer of this will be d) 1,2,&4 .
- The most important part of d) Formulating the clustering problem is selecting the variables on which clustering is based.
- The most commonly used measure of similarity is the a) Euclidean distance or its square.
- b) Divisive Clustering is a procedure where all objects start out in one giant cluster.
- d) all is correct .
- The goal of clustering is to a) Divide the data points into groups.
- Clustering is a b) Unsupervised learning .
- a) K-Means clustering algorithms suffers the problem of convergence .
- a) K-Means clustering algorithm is most sensitive to outliers.
- d) All of the above data points having outliers and different densities with non convex shapes is a bad characteristics of datasets.
- For clustering we do not require a) Labeled data.

Subjective Type Answers: -

- 13) Cluster analysis is also called segmentation analysis calculated by the distances, link the clusters, choose a solution by selecting the right number of cluster.
- 14) To measure the quality of cluster we will use the average silhouette coefficient value of all objects in the data sets.
- 15) Cluster analysis is a multivariate data mining technique whose goal is to form group objects based on the selected characteristics or attributes types of cluster analysis: -
- Hierarchical Cluster Analysis

- Centroid based Clustering
- Distribution based Clustering
- Density based Clustering