

## Assignment 1 – Machine Learning

1. a) 2.
2. d) 1, 2 and 4.
3. d) Formulating the clustering problem.
4. a) Euclidean distance.
5. b) Divisive Clustering.
6. d) All are correct.
7. a) Divide the data points into groups.
8. b) Unsupervised learning
9. a) K-Means clustering
10. a) K- means
11. d) All of the above.
12. a) Labeled Data
13. Cluster analysis is calculated by using a variety of algorithms such as k-means clustering, hierarchical clustering, and density-based clustering. These algorithms are used to identify natural groupings of data points, or clusters, within a larger dataset. The algorithms take into account the similarities and differences between data points and use this information to assign each data point to a specific cluster. The result is a set of clusters which can then be analyzed further to gain insights into the data.
14. Cluster quality can be measured using various evaluation metrics such as the Rand Index, the Silhouette Coefficient, the Dunn Index, and the Calinski-Harabasz Index. The Rand Index measures the similarity of the resulting clusters to the true, known clusters. The Silhouette Coefficient measures the cohesion and separation of the clusters. The Dunn Index measures the diameter and inter-cluster distance. Lastly, the Calinski-Harabasz Index measures the ratio of the between-cluster scatter to the within-cluster scatter.
15. Cluster analysis is a data mining technique used to group data points into clusters, or groups of similar objects. Cluster analysis is used to find patterns and relationships in data and identify clusters or natural divisions among the data. There are two types of cluster analysis: hierarchical clustering and partitioning clustering. Hierarchical clustering is a bottom-up approach that starts with individual data points, and gradually merges them into larger

clusters. Partitioning clustering is a top-down approach that starts with all data points in one group, and then splits them into smaller groups.