

Machine Learning Assignment-1

1. b) 4
 2. d) 1, 2 and 4
 3. d) formulating the clustering problem
 4. a) Euclidean distance
 5. b) Divisive clustering
 6. d) All answers are correct
 7. a) Divide the data points into groups
 8. b) Unsupervised learning
 9. d) All of the above
 10. a) K-means clustering algorithm
 11. d) All of the above
 12. a) Labeled data
- 13.** The hierarchical cluster analysis follows three basic steps: 1) calculate the distances, 2) link the clusters, and 3) choose a solution by selecting the right number of clusters.
- 14.** Measuring the quality of a clustering, We have a few methods to choose from for measuring the quality of a clustering. In general, these methods can be categorised into two groups according to whether ground truth is available. Here, ground truth is the ideal clustering that is often built using human experts.
- If ground truth is available, it can be used by extrinsic methods, which compare the clustering against the group truth and measure. If the ground truth is unavailable, we can use intrinsic methods, which evaluate the goodness of a clustering by considering how well the clusters are separated. Ground truth can be considered as supervision in the form of “cluster labels.” Hence, extrinsic methods are also known as supervised methods, while intrinsic methods are unsupervised methods.

15. Cluster analysis is a multivariate data mining technique whose goal is to groups objects (eg., products, respondents, or other entities) based on a set of user selected characteristics or attributes. It is the basic and most important step of data mining and a common technique for statistical data analysis, and it is used in many fields such as data compression, machine learning, pattern recognition, information retrieval etc.

Types of Cluster Analysis:-

- Hierarchical Cluster Analysis
- Centroid-based Clustering
- Distribution-based Clustering
- Density-based Clustering