MACHINE LEARNING ASSIGNMENT -1

ANSWERS:

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 correct
7. (a) Divide 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) Labelled data
13. Cluster Analysis is calculated in 3 main steps:
    a. Picking a clustering technique upon scaling the metric data and choosing the number of clusters with appropriate distances.
    b. Initialising the clustering model with defined clusters and fitting the dataset in the model.
    c. Checking the model/cluster quality using metrics such as silhouette_score
14. Cluster quality can be measured using any of the clustering accuracy metrics such as silhouette_score.
15. Cluster analysis is an Unsupervised Machine learning technique which tries to distribute the dataset into most appropriate number of smaller groups/clusters based on similarities with other data points. It does not require any label for analysis.
    Types:
    - Agglomerative Clustering – It is a “Bottom-Up” type of clustering where individual data points are grouped together further and further based on similarities (distance from each other) until 1 big cluster is formed containing all data points. Examples: DBSCAN, Hierarchical Clustering.
    - Divisive Clustering – It is a “Top-Down” approach where the complete data set is divided down to smaller clusters/groups based on similarities or distances of 1 data point to another until a predefined number of clusters is formed. Example: KMeans Clustering