

ASSIGNMENT – 2 MACHINE LEARNING

1. D) 2 and 3
2. D) 1,2 and 4
3. A) True
4. A) Capping and Flooring of Variables.
5. B) 1
6. A) Yes
7. A) Yes
8. D) All of the above
9. A) K-means Clustering Algorithm
10. A) 1 only.
11. D) All of the above
12. The *K*-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values. *K*-medoids clustering is a variant of *K*-means that is more robust to noises and outliers. Instead of using the mean point as the center of a cluster, *K*-medoids uses an actual point in the cluster to represent it. Medoid is the most centrally located object of the cluster, with minimum sum of distances to other points. The group of points in the right form a cluster, while the rightmost point is an outlier. Mean is greatly influenced by the outlier and thus cannot represent the correct cluster center, while medoid is robust to the outlier and correctly represents the cluster center.
13. K means is better than Hierarchical clustering in a few ways. Firstly, K means is more efficient and can process larger datasets more quickly. Additionally, K means is more effective when the clusters are distinct and clear-cut, whereas Hierarchical clustering can struggle when the clusters are not well-defined. Finally, K means is more suitable for high-dimensional datasets, whereas Hierarchical clustering may be more effective for lower-dimensional datasets.
Also, it can handle missing data and is robust to the initial set of clusters. It is often used as a first step in exploratory data analysis to identify patterns and structure in the data. However, it does have some limitations, such as assuming that clusters are spherical and equally sized, which may not always be the case in real-world data.
14. The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids. This means that running the algorithm several times on the same data, could give different results.
The key idea of the algorithm is to select data points which belong to dense regions and which are adequately separated in feature space as the initial centroids.