MACHINE LEARNING

1. d) 2 and 3
2. d) 1, 2 and 4
3. True
4. 1 only
5. 1
6. No
7. Yes
8. All of the above
9. K-means clustering algorithm
10. All of the above
11. d) All of the above
12. The sensitivity of K-nearest neighbours (KNN) algorithm to outliers depends on the value of k and the distance metric used. In general, KNN is sensitive to outliers when k is small because the classification decision is based on the labels of a few nearby points. Outliers can greatly influence the distance metric and thus affect the classification result.

However, when k is large, the effect of outliers is reduced because the algorithm considers more neighbours and the influence of any single outlier is diminished. Additionally, using a robust distance metric can helps to reduce the impact of outliers.

Therefore, whether KNN is sensitive to outliers or not depends on the specific dataset and parameters used, and it is important to carefully select the value of k and distance metric based on the nature of the data.

1. K-means is a popular clustering algorithm that is widely used in various applications, such as image processing, data mining, and machine learning. Here are some reasons why K-means is considered a better algorithm:

* Simple and easy to implement: K-means is a simple and intuitive algorithm that is easy to implement, even for large datasets. The algorithm can be implemented in various programming languages and requires minimal parameter tuning.
* Fast and scalable: K-means is a fast and scalable algorithm that can handle large datasets with millions of records. The algorithm uses an iterative approach that converges quickly, and the computational complexity is linear with the number of data points.
* Efficient in terms of memory usage: K-means is an efficient algorithm in terms of memory usage because it only stores the centroid coordinates for each cluster, rather than the entire dataset.
* Robust to noise: K-means is robust to noise and can handle datasets with outliers by assigning them to the nearest cluster centroid.
* Flexible: K-means is a flexible algorithm that can be customized to different applications by using different distance metrics, initialization methods, and stopping criteria.

Overall, K-means is a powerful and versatile clustering algorithm that can efficiently partition large datasets into meaningful groups, making it a popular choice in various fields.

1. Yes, k-means is a deterministic algorithm. This means that given the same input data and parameters, the algorithm will always produce the same output. The algorithm works by iteratively assigning data points to clusters and updating the cluster centroids until convergence is reached. The final cluster assignments and centroids will be the same every time the algorithm is run on the same data with the same parameters.

However, if the initialization of cluster centroids is changed or a different distance metric is used, the resulting clustering may be different.