ASSIGNMENT – 3 MACHINE LEARNING

1. D) All of the above.
2. D) None
3. A) Supervised Learning.
4. B) The tree representing how close the data points are to each other.
5. D) None.
6. C) k-nearest neighbor is same as k-means.
7. D) 1,2 and 3
8. A) 1 only.
9. A) 2
10. B) Given a database of information about your users, automatically group them into different market segments.
11. For the single link or MIN version of hierarchical clustering, the proximity of two clustersis defined to be the minimum of the distance between any two points in the differentclusters. For instance, from the table, we see that the distance between points 3 and 6 is0.11, and that is the height at which they are joined into one cluster in the dendrogram.As another example, the distance between clusters {3, 6} and {2, 5} is given by dist({3,6}, {2, 5}) = min(dist(3, 2), dist(6, 2), dist(3, 5), dist(6, 5)) = min(0.1483, 0.2540, 0.2843,0.3921) = 0.1483
12. For the single link or MAX version of hierarchical clustering, the proximity of two clustersis defined to be the maximum of the distance between any two points in the differentclusters. Similarly, here points 3 and 6 are merged first. However, {3, 6} is merged with{4}, instead of {2, 5}. This is because the dist({3, 6}, {4}) = max(dist(3, 4), dist(6, 4)) =max(0.1513, 0.2216) = 0.2216, which is smaller than dist({3, 6}, {2, 5}) = max(dist(3, 2),dist(6, 2), dist(3, 5), dist(6, 5)) = max(0.1483, 0.2540, 0.2843, 0.3921) = 0.3921 anddist({3, 6}, {1}) = max(dist(3, 1), dist(6, 1)) = max(0.2218, 0.2347) = 0.2347.
13. Clustering is a technique used in data analysis and machine learning to group similar data points together. It is important because it can help identify patterns and relationships in the data that might not be immediately apparent, and can also be used for tasks such as anomaly detection and data compression. Additionally, it can be used as a preprocessing step for other machine learning tasks such as classification. Clustering can also help in

understanding complex data sets and can be useful in fields such as market research, image processing and bioinformatics.

14. There are several ways to improve the performance of a clustering algorithm:

1. Feature selection: Selecting the most relevant features for the clustering task can improve the performance of the algorithm by reducing the dimensionality of the data.
2. Preprocessing: Data preprocessing techniques such as normalization and scaling can help improve the performance of clustering algorithms by ensuring that all features are on the same scale.
3. Choosing the right algorithm: Different clustering algorithms are suitable for different types of data and clustering tasks. Choosing the right algorithm for your data and task can greatly improve the performance of the clustering.
4. Choosing the right number of clusters: Determining the appropriate number of clusters for the data set can be challenging. The elbow method, silhouette analysis, and gap statistic are popular methods for determining the optimal number of clusters.
5. Hyperparameter tuning: Clustering algorithms often have several hyperparameters that can be adjusted to optimize performance.
6. Using a combination of multiple clustering algorithm: Some data set can be complex, and might not be able to be explained by a single clustering algorithm. Combining multiple algorithms and comparing the results can be beneficial in such cases.

It's also important to keep in mind that the quality of the results of clustering algorithm is highly dependent on the quality of the data. Thus, it is essential to have a good understanding of the data set, and to clean and preprocess the data accordingly before applying a clustering algorithm.