MACHINE LEARNING – WORKSHEET (CLUSTERING)

1. d. All of the above
2. d. None
3. c. Reinforcement learning
4. b. The tree representing how close the data points are to each other
5. d. None
6. c. k-nearest neighbour 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. Option- (A)
12. Option-(B)
13. What is the importance of clustering?

* Clustering is defined as same type of data points merged into similar groups and different set of data points are in different cluster (groups).
* The importance of clustering we need to first divide the data points into groups and then assign the labels to the group.
* Clustering is very useful when the marketer wants to discover distinct group out of lots and find out the consumer purchasing behaviour.
* Clustering plays a pivotal role in finding out the outliers in the dataset for example in credit card fraud detection for a customer.
* From the clustering we also get the know how the data points are distributed inside and find out the pattern in each data groups.(cluster)

1. How do you cluster a profile?

* In Data science clustering is used when we have large data sets and want to find out the characteristics of each group separately, that forms a cluster. We analyse each data group and find out the pattern or behaviour. For doing the cluster profile the following steps are required:
* Step 1: First we need to verify data is metric.
* Step 2: Scale the data.
* Step 3: Select Segmentation Variables.
* Step 4: Explain the similarity measure.
* Step 5: Visualize Pair-wise Distances of each data points from the nearest centroid value.
* Step 6: Method and Number of division.
* Step 7: Profile and understand the segments of each cluster.
* Step 8: Critically Analysis of each data group.

1. How can I improve my clustering performance?

* There are two method to evaluate the performance of clustering algorithm. Internal assessment and external assessment. In most cases the external performance is used in which the clustering result is compared with expert result. MoJo, MoJoFM is the algorithms used for external evaluation. Also we can add neighbouring clusters if the resulting cluster's variance is below the threshold level. Removing the elements that are far if the variance of the cluster is above the threshold level.