Machine Learning - Worksheet 3

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 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) (A)

(12) (B)

(13) Importance of Clustering

- Clustering analysis is broadly used in many applications such as market research, Pattern recognition, data analysis and image processing.

(14) clusters can be profile in following steps:-

1) Graphically represent your clusters according to your input variables. The selected cluster algorithm aims to maximize similarity between the datapoints in the same cluster and minimize similarity between data points in different clusters.

2) Score your cluster in a table so that you can measure and compare them on each input variable with regards to numerical or descriptive values.

Step 3) Now, it time to profile your ~~customers~~ clusters. At this step variables should be described in a type of story about the category or customer base

(15) • K means clustering algorithm can significantly improved by using a better initialization technique and by repeating the algorithm

• When the data has overlapping clusters, K means can improve the results of initialization technique.

• When data has well seperated clusters the performance of k means depends completely on the goodness of initialization.

• Initialization using simple furthest point heuristic reduces the clustering error of k-means from 15% to 6% on average.