MACHINE LEARNING ASSIGNMENT

ANSWERS:

Objective type:

1. d. All of the above
2. d. None
3. c. Reinforcement learning and Unsupervised 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. a.
12. b.

Subjective type:

13. It provides structure and visualization tools for unsupervised learning. It also gives a starting point for semi-supervised learning algorithms (cluster a small sample, use results as labels, proceed to large sample).

14. The clustering performance can be improved in following ways:

    - K-means clustering algorithm can be significantly improved by using a better initialization technique, and by repeating (re-starting) the algorithm.
    - When the data has overlapping clusters, k-means can improve the results of the initialization technique.
    - When the data has well separated clusters, the performance of k-means depends completely on the goodness of the initialization.
    - Initialization using simple furthest point heuristic (Maxmin) reduces the clustering error of k-means from 15% to 6%, on average.words, an error term is a value which represents how observed data differs from actual population data.
    - It can be improved by using feature weight learning as well. To measure feature weight importance, we will have to use a weighted euclidean distance function.