ASSIGNMENT -2

Worksheet 1

MACHINE LEARNING

1. a) 2 Only
2. d) 1, 2 and 4
3. a) True
4. a) 1 only
5. b) 1
6. b) No
7. a) Yes
8. d) All of the above
9. a) K-means clustering algorithm
10. d) All of the above
11. d) All of the above

12. . Is K sensitive to outliers?
Yes. K is sensitive to outliers.

13. Why is K means better?
K means is considered to be very simple and easy to implement. It is one of the most robust methods, especially when it comes to image segmentation. The main reason is that, K means can handle big data because the time complexity is linear.

14. Is K means a deterministic algorithm?
K means is a Non Deterministic Algorithm because compiler cannot solve the problem in polynomial time as some problems have a great degree of randomness in them.

ASSIGNMENT 3

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.

13. What is the importance of clustering?

    Clustering is the task of grouping a set of objects so that objects in same group are more similar to each other than to those in other groups. It helps in understanding the natural grouping in a dataset. Clustering is considered to be very important in Data Analysis and Data Mining. A good clustering algorithm helps in identifying clusters irrespective of their shapes.

14. How can I improve my clustering performance?

    Firstly, we need to perform a visual check that the clusters look as expected and similar examples considered do appear in the same cluster. Then we need to check the following metrics:
    1. Cluster Cardinality
    2. Cluster Magnitude
    3. Performance of downstream system.