Assignment 2-Mchine Learning

Ans.1) Clustering

Ans.2) Regression, classification and reinforcement learning

Ans.3) True

Ans.4) Capping and flooring of variables

Ans.5) B option - 1

Ans.6) B option – No

Ans.7) A option – Yes

Ans. 8) D option – All of the above

Ans. 9) K-means clustering algorithm

Ans. 10) All of the above

Ans.11) All of the above

Ans.12) The mean is easily influenced by extreme values. K-medoids clustering is a variant of K-means that is more robust to noises and outliers. Therefore, K sensitive to outliners.

Let me elaborate it with the help of an example:

The data set point are 1 2 3 7 8 80

Now 80 is outlier.

K=2

C1=1 C2=7

After first iteration

C1=2 C2=31.67

As 80 data point which is outlier comes in cluster 2.

Cluster 2 centroid changes to accommodate 80.

Therefore K means is sensitive to outliers.

Ans.13) K-means clustering is a very famous and powerful unsupervised machine learning algorithm. Followings are the advantages of k-means which proves that K means better:

1. It is very simple to implement.

2. Guarantees convergence.

3. It is scalable to large data set and also faster to large datasets.

4. Easily adapts to new examples.

5. Can warm-start the positions of centroids.

6. Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

Ans.14) K means a deterministic algorithm because of its random selection of data points as initial centroids. Since it starts with a random set of data points as initial centroids, therefore it influences the quality of the resulting clusters.