# MACHINE LEARNING ASSIGNMENT-2

**Answer of Q1 - (B)** 1 and 2

**Answer of Q2 - (A)** 1 , 2 and 4

**Answer of Q3 - (A)** True

**Answer of Q4 - (A)** Capping and flooring of variables

**Answer of Q5 - (B)** 1

**Answer of Q6 - (B)** No

**Answer of Q7 - (A)** Yes

**Answer of Q8** - **(D)** All of the above

**Answer of Q9** - **(A)** K-means clustering algorithm

**Answer of Q10** - **(D)** All of the above

**Answer of Q11** - **(D)** All of the above

**Answer of Q12**

The *K*-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values. *K*-medoids clustering is a variant of *K*-means that is more robust to noises and outliers. In K-Means clustering outliers are found by distance based approach and cluster based approach. In case of hierarchical clustering, by using dendrogram outliers are found. The goal of the project is to detect the outlier and remove the outliers to make the clustering more reliable.

**Answer of Q13**

1. Relatively simple to implement.

2. Scales to large data sets.

3. Guarantees convergence.

4. Can warm-start the positions of centroids.

5. Easily adapts to new examples.

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

**Answer of Q14**

The basic k-means clustering is based on a non-deterministic algorithm. This means that running the algorithm several times on the same data, could give different results. The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids.