Machine Learning Assignment-2

A.1) Option a)- 2 Only - Clustering

A.2) Option d)- 1, 2 and 4 **–** Regression, Classification and Reinforcement

A.3) Option a)- True

A.4) Option a)- 1 Only - Capping and flooring of variables

A.5) Option b)- 1

A.6) Option b)- No

A.7) Option a)- Yes

A.8) Option d)- All of the above

A.9) Option a)- K-means clustering algorithm

A.10) Option d)- All of the above

A.11) Option d)- All of the above

A.12) 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.

A.13) Advantages of K means

- Relatively simple to implement
- Scales to large data sets
- Guarantees convergence
- Can warm-start the positions of centroids
- Easily adapts to new examples
- Generalizes to clusters of different shapes and sizes
- Such as elliptical clusters.

A.14) K means is a non-deterministic algorithm. K-Means starts with a random set of data points as initial centroids. This random selection influences the quality of the resulting clusters. Besides, each run of the algorithm for the same dataset may yield a different output.