

# MACHINE LEARNING ASSIGNMENT-2

**Ans1) Clustering {(a) 2 only}**

**Ans2) Regression, Classification and Reinforcement {(d)1,2,4}**

**Ans3) True (a)**

**Ans4) Capping and Flooring of variables {(a)1 only}**

**Ans5) 1 (b)**

**Ans6) No (b)**

**Ans7) Yes (a)**

**Ans8) All of the above (d)**

**Ans9) K-means clustering algorithm (a)**

**Ans10) All of the above (d)**

**Ans11) All of the above (d)**

**Ans12) Yes, k is sensitive to outliers because a mean is easily influenced by extreme values. Therefore, if we have any outlier then the mean changes drastically and affects the K-means algorithm altogether.**

**Ans13) The advantages of the K-means clustering algorithm are:**

- a) It is relatively easy to implement.**
- b) It scales to the large size datasets in them.**
- c) It guarantees converges to a global minimum.**
- d) It can warm start the position of centroids.**
- e) It generalizes to clusters of different shapes and sizes such as elliptical clusters.**
- f) It easily adapts to new examples.**

**Ans14) 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. A deterministic algorithm is simply an algorithm that has a predefined output. For instance, if you are sorting elements that are strictly ordered (no equal elements) the output is well defined and so the algorithm is deterministic. Most computers are deterministic. The non-deterministic nature of**

**K-Means is due to its random selection of data points as initial centroids. The key idea of the algorithm is to select data points that belong to dense regions and which are adequately separated in feature space as the initial centroids.**