NAME: Dipesh Ramesh Limaje

INTERNSHIP BATCH : 33

TOPIC: MACHINE LEARNING

SME : Mr. Shwetank Mishra

WORKSHEET NO : 2

Q1. a) 2 Only

Q2. d) 1, 2 and 4

Q3. a) True

Q4. a) 1 only (Capping and flooring of variables)

Q5. b) 1

Q6. b) No

Q7. a) Yes

Q8. d) All of the above

Q9. a) K-means clustering algorithm

Q10. d) All of the above

Q11.d) All of the above

Q12. Is K sensitive to outliers?

Ans. 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. Instead of using the mean point as the center of a cluster, *K*-medoids uses an actual point in the cluster to represent it. Medoid is the most centrally located object of the cluster, with minimum sum of distances to other points.

Q13. Why is K means better?

ANS

- 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.

Q14. Is K means a deterministic algorithm?

ANS Clustering algorithms with steps involving randomness usually give different results on different executions for the same dataset. This non-deterministic nature of algorithms such as the K-Means clustering algorithm limits their applicability in areas such as cancer subtype prediction using gene expression data. It is hard to sensibly compare the results of such algorithms with those of other algorithms. The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids.