

Flip Robo  
**Machine Learning**  
**Assignment – 2**

Ans 1: Option b - > 1&2

Ans 2: Option d -> 1,2 & 4

Ans 3: Option a -> True

Ans 4: Option a -> 1 Only – Capping and flooring of variables

Ans 5: Option b -> 1

Ans 6: Option b -> No

Ans 7: Option a -> Yes

Ans 8: Option d -> All of the above

Ans 9: Option a -> K-means clustering algorithm

Ans 10: Option f -> All of the above

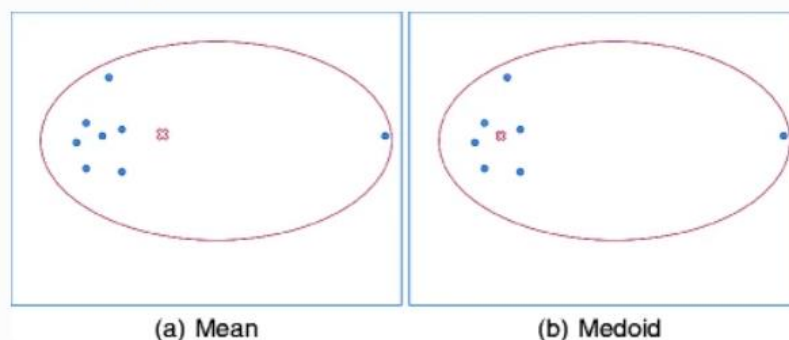
Ans 11: Option d-> All of the above

## Q 12. 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. Figure 1 shows the difference between mean and medoid in a 2-D example. The group of points in the right form a cluster, while the rightmost point is an outlier. Mean is greatly influenced by the outlier and thus cannot represent the correct cluster center, while medoid is robust to the outlier and correctly represents the cluster center.

**K-Medoids Clustering. Figure 1**



Mean vs. medoid in 2-D space. In both figures (a) and (b), the group of points in the right form a cluster and the rightmost point is an outlier. The red point represents the center found by mean...

**Q 14. Why is K means better?**

**Ans:** k-means is one of the simplest algorithm which uses unsupervised learning method to solve known clustering issues. It works really well with large datasets.

However, there are also drawbacks of K-Means which are:

- Strong sensitivity to outliers and noise
- Doesn't work well with non-circular cluster shape -- number of cluster and initial seed value need to be specified beforehand
- Low capability to pass the local optimum.

**Q 15. Is K means a deterministic algorithm ?**

**Ans:** The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids. Method: We propose an improved, density based version of K-Means, which involves a novel and systematic method for selecting initial centroids.

One of the significant drawbacks of K-Means is its **non-deterministic nature**. 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.