MACHINE LEARNING

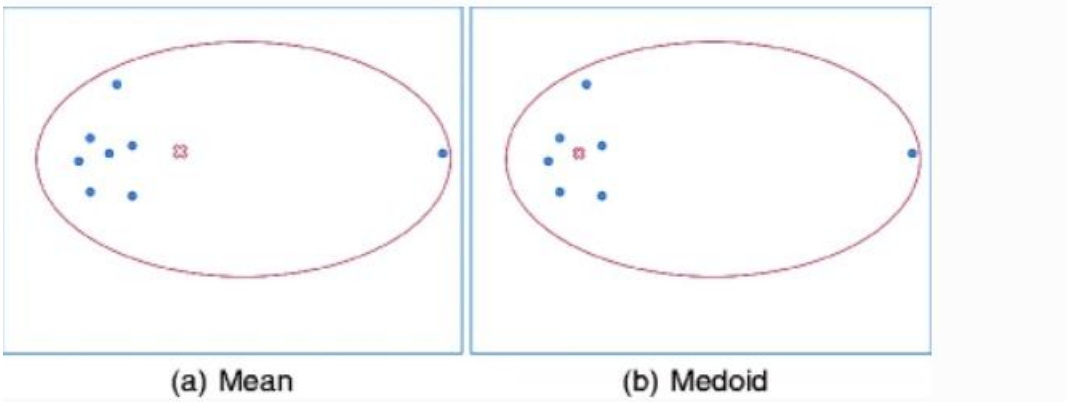
1. **A**
2. **D**
3. **A**
4. **A**
5. **B**
6. **B**
7. **A**
8. **D**
9. **A**
10. **D**
11. **D**

12. Is K sensitive to outliers?

Answer: 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

From: *K*-Medoids Clustering



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 or medoid.

13. Why is K means better?

Answer:

1. It is very simple to implement.
2. It is scalable to a huge data set and also faster to large datasets.
3. it adapts the new examples very frequently.
4. Generalization of clusters for different shapes and sizes.

14. Is K means a deterministic algorithm?

Answer: No, 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. Method: We propose an improved, density based version of K-Means, which involves a novel and systematic method for selecting initial centroids.