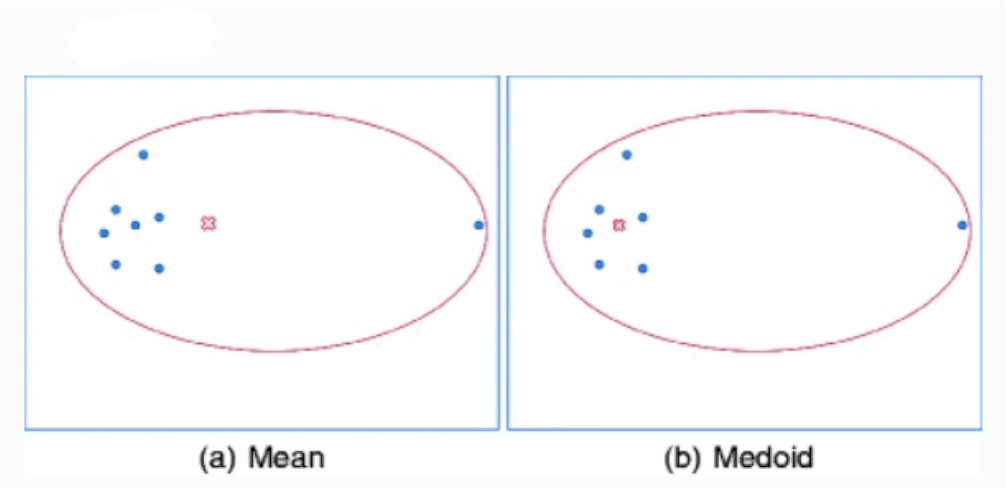
WORKSHEET 2 - MACHINE LEARNING

Submitted by Jwala R, (21/01/2023)

1) a) 2 only
2) d) 1, 2 and 4
3) a) True
4) a) 1 only
5) b) 1
6) b) No
7) a) Yes
8) d) All of the above
9) a) K means clustering algorithm
10) d) All of the above
11) d) All of the above

12) A mean is easily impacted by extreme values, hence the K-means clustering technique is sensitive to outliers. A K-means variation that is more resistant to noise and outliers is K-medoids clustering. K-medoids employs a real point in the cluster to represent the cluster centre rather than the mean point. Medoid is the item in the cluster that is closest to the centre and has the fewest total distances from other locations. In a 2-D example, The figure illustrates the distinction between mean and medoid. The rightmost point is an outlier, whereas the collection of points to its right form a cluster. 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.



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.

14) 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. In cluster analysis, the elbow method is a heuristic used in determining the number of clusters in a data set. The method consists of plotting the explained variation as a function of the number of clusters, and picking the elbow of the curve as the number of clusters to use.

    Also, the compiler cannot solve the problem in polynomial time and doesn't clearly know the next step. This is because some problems have a great degree of randomness to them. These algorithms usually have 2 steps — 1)Guessing step 2)Assignment step. On similar lines is the K-means algorithm. The K-Means algorithm divides the data space into K clusters such that the total variance of all data points with respect to the cluster mean is minimized.