MACHINE LEANING- WORKSHEET – 2

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

12. The K-means algorithm is sensitive to the outliers. In these papers, we propose a robust two stages k- means clustering algorithm is sensitive to the outliers. In this paper, we paper, we propose a robust two stages k-means clustering algorithm based on the observation point mechanism, which can accurately discover the cluster centers without the disturbance of outliers.

The K-means clustering algorithm is sensitive to outliers because a mean is easily influences by extreme values.

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 forms 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.

13.

1. Relatively simple to implement.

2.Scales to large data set.

3. Guarantees convergence.

4.Can warm-start the position of centriods.

5.Easily adapts to new examples.

6.Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

7. Choosing manually.

8. being dependent on initial values.

9.Easy to implement.

10. With a large number of variables, K means may be computationally faster than hierarchical clustering.

11. K means may produce higher cluster than Hierarchical clustering.

12. An instance can change cluster when the centroids are recomputed.

13. Moreover it has linear space complexity and it is generally fast.

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.

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.

K-Means is a non-deterministic algorithm. This means that a 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.