

## Machine Learning Assignment-2

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. 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.
  
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. Cluster analysis aims to group a set of objects/events in such a way that objects/events in the same group] are more similar to each other than to those in other groups. *k*-means is a partitioning-based clustering algorithm. *k*-means method for clustering is an iterative process in which an initial partition of given *k* clusters is then improved by applying a search algorithm to the data. Simplifying, given a pre-defined number (*k*) of clusters, the algorithm:
  - begins with an initial set of *k* cluster centers (*i.e.* the centroids)
  - (re)assigns objects to the closest centroids
  - recalculates centroids according to new memberships of the data points.

- repeats the last two steps until a consistent result is found or until the maximum number of iterations is reached.

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.