Machine Learning Assignment-2

- **Q1 to Q11**

Question No.	Answer
1	A
2	D
3	A
4	A
5	B
6	B
7	A
8	D
9	A
10	D
11	D

Q12. Is K sensitive to outliers?

Ans: Yes. The K-means clustering algorithm is sensitive to outliers because extreme values easily influence a mean. 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 use an actual point in the cluster to represent it. Medoid is the most centrally located object of the cluster, with a 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.

Q13. Why is K means better?
Ans:
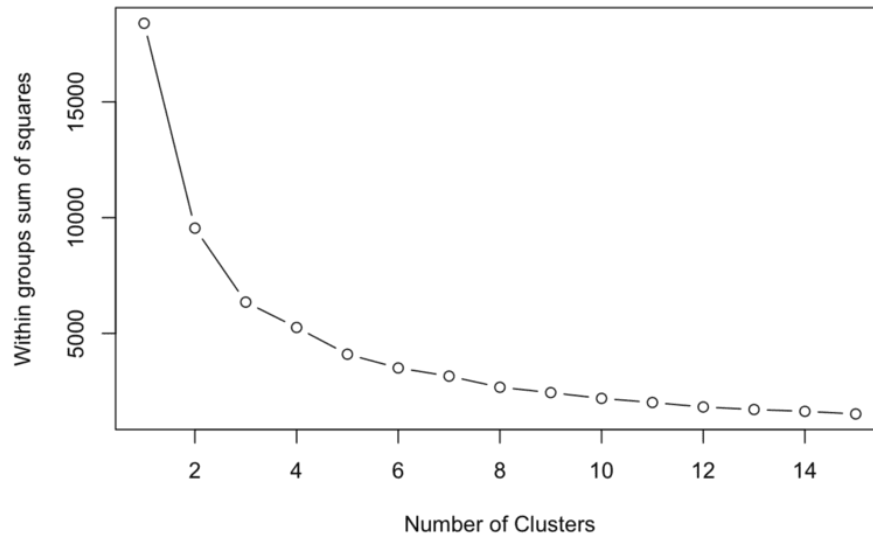
K-Means for Clustering is one of the popular algorithms for this approach. Where K means the number of clustering and means implies the statistics mean a problem. It is used to calculate code vectors (the centroids of different clusters). According to a tutorial, for any word/value/key that needs to be 'vector quantized', it is by calculating the distance from all the code vectors and assigning the index of the code vector with the minimum distance to this value. For example, clustering can be applied to MP3 files, cellular phones are the general areas that use this technique.

Q14. Is K means a deterministic algorithm?

Ans: K means is non-deterministic in nature and this is one of the significant drawbacks of K-Means. 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.