K-means Clustering

K-means Clustering is an unsupervised ML algorithm used for classification problems. It segregates the unlabeled data into clusters based on similar features.

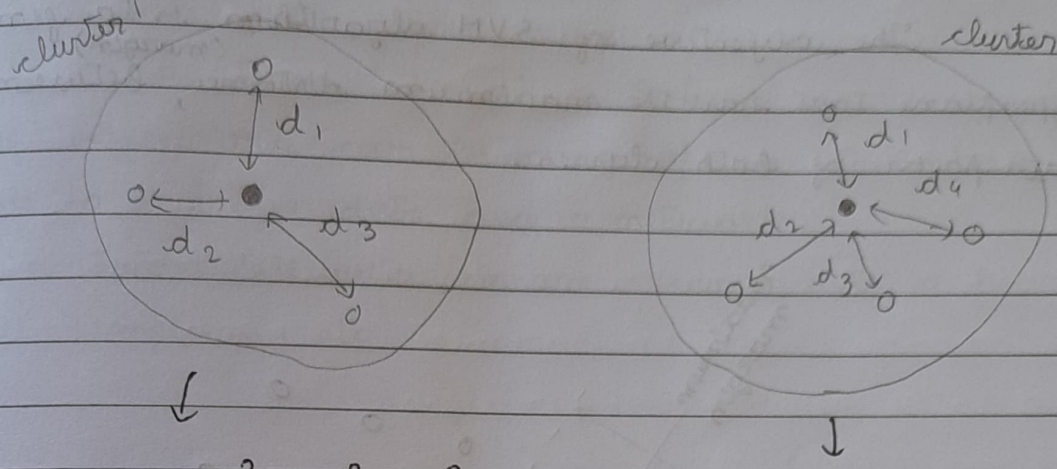
Pseudocode for K-means :-

1. Find the optimal value for K (no. of clusters)
2. Randomly select K ~~data~~ points which will act as centroids.
3. Calculate the distance of each data point with each of the centroids.
4. Assign each data point to the ~~the~~ centroid with which it has minimum distance.
5. Calculate the new centroids by finding the mean of each cluster.
6. Repeat Step 3 - 5 until old centroids := new centroids.

Value of K :-

The optimal value for K is chosen by 'Elbow method'. Here, WCSS (within cluster sum of squares) is calculated for different values of K and the value of K which has minimum WCSS is chosen.

WCSS is the sum of squares of distances of the data points in each and every cluster from its centroid.



$$WCSS_1 = d_1^2 + d_2^2 + d_3^2$$

$$WCSS_2 = d_1^2 + d_2^2 + d_3^2 + d_4^2$$

$$WCSS = WCSS_1 + WCSS_2$$