53. K-Means Clustering

- K-Means Clustering is an unsupervised learning algorithm, which groups the unlabelled dataset into different clusters.
- K defines the number of pre-defined clusters that need to be created in the process.

K-Means algo:

- First decide the centriod, center in the dataset
- take two data point, and draw a line b/w them
- pass another line from middle of the line
- take neighbouring data points from the decided central data point

How K-Means work:

1. Take random sample point
2. Create groups
3. Search nearest point
4. Calculate mean (Move points)

 No description has been provided for this image

Elbow Method:

- The Elbow method is one of the most popular ways to find the optimal number of clusters
- This method uses the concept of WCSS value. WCSS stands forWithin-Cluster Sum of Squares, which defines the total variations within a cluster.
- The formuls of **WCSS** is:

$$\mathrm{WCSS} = \sum_{i=1}^{K} \sum_{x \in C_i} \|x - \mu_i\|^2$$

where:

- (K) = Number of clusters
- (C_i) = (i)-th cluster
- (x) = A data point in cluster (C_i)
- (\mu_i) = Centroid of cluster (C_i)
- (| x - \mu_i |) = Euclidean distance between data point (x) and centroid (\mu_i)

How does WCSS is calculated:

- Caclulate the distance from decided central data point and its neighbouring data points (x - u)
- Take square of the distance
- Sum of the distances from central point and all neighbouring data points

 No description has been provided for this image

K-Means ++: To have best clustering in the data. It takes 2 decided points away from each other.

In [ ]: