Unsupervised Learning

- It requires only input variable (x).

K-means:-

- K-means comes, under unsupervised Learning which makes use of unlabelled data.
- K-means is a clustering algorithm which cluster unlabelled data into different cluster based on similarity.

Ex:- Tinder, customer segmentation, fraud detection.

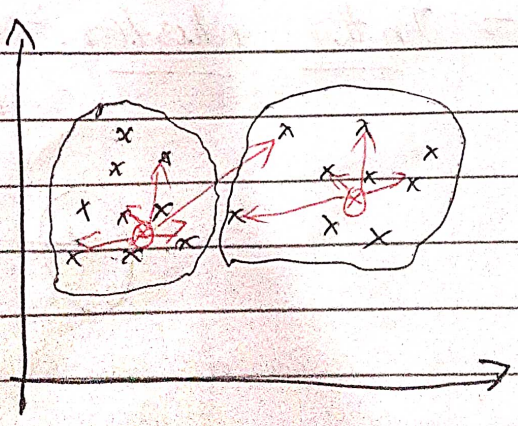
Labelled data ⇒ Both x & y
Unlabelled data ⇒ only x.
similarity ⇒ Nearest distance.
K ⇒ no of cluster.

- K-mean is distance based algorithm
- It will make use of distance measures like Euclidean, manhatton, & Minkowski to find distance b/w any two observation

How K-mean works:-

(1) plot data

(2) Define no of clusters (group) Ex: K = 2



(3) It will create 2 random cluster.

(4) It will initialise the centroids

centroids : cluster centers, centroids are calculated by taking average of all the observation in the cluster.

No of cluster = No of centroids.

(5) find distance b/w centroid and all other observation.

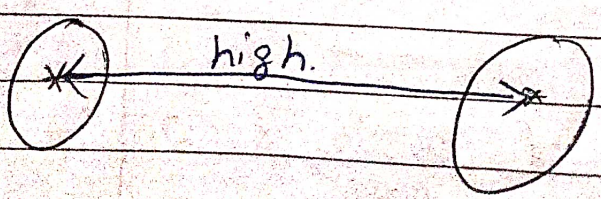
(6) Assign Each observation to the nearest centroid based on distance.

(7) It will repeat steps from 4th step till it gets clearer cluster / No overlapping

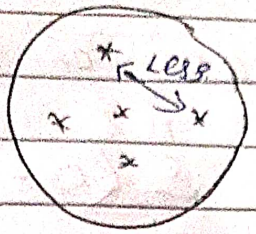
Goal / Aim:

- Inter cluster distance should be high.
- Intra cluster distance should be less

→ Inter cluster: distance b/w clusters should be high.



→ Intra cluster: distance within cluster should be less



⇒ How to evaluate k-mean?

silhovetle score = $\frac{b-a}{max(a,b)}$

a → Intra cluster distance
b → Inter cluster distance

Range = [-1, 1]
→ near 1 → clearer clusters
→ near 0 or 1 → cluster are overlapping

→ Imp: • scaling
• handling outliers

⇒ How to choose optimal value for k?

• Elbow Method

• It is an iterative process which works iteratively for different k-values and gives the optimal k.

K = [2, 3, 4, ... 11]

• Take different k-values.

- Implement K-mean.
- Calculate WCSS [with in cluster, sum of squares]

$$WCSS = \sum_{i=1}^{n} (c_i - x_i)^2$$

- WCSS is distance b/w centroid & observation.
- plot graph of K v/s WCSS
- Choose the point after which you see break down in WCSS