Unsupervised learning

k-mean

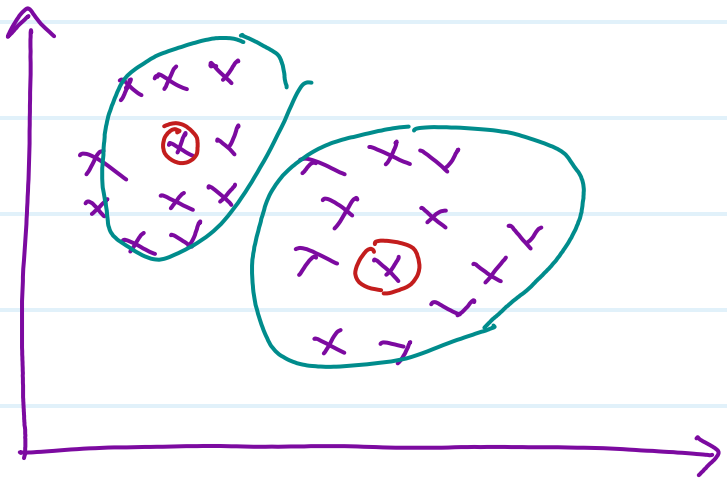
k- mean ++

hirarchical clustering

DB scan

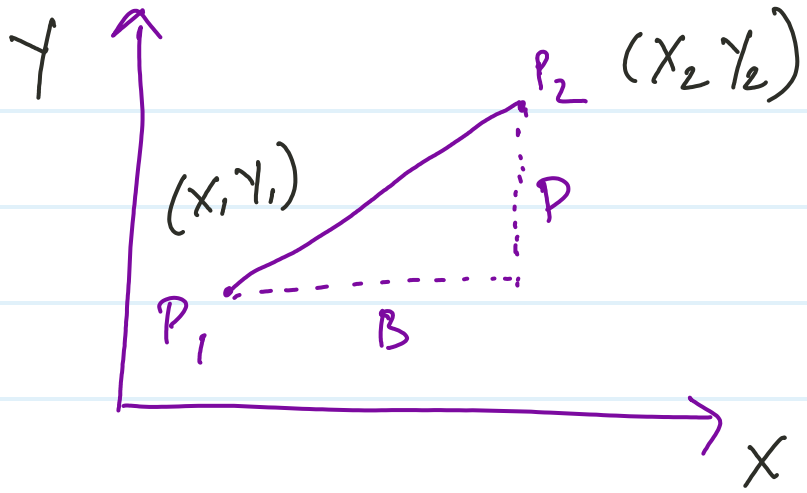
✱ K-mean clustering.

(Centroid based approch)



✱ PGT

H2 : $P^2 + B^2$



Example –

Height	weight	
185	72	– C_1
170	56	– C_2
168	60	
179	68	
182	72	
188	77	
180	71	
160	70	
183	84	
180	88	

✱ Random centroid

①	185	72
②	170	56

3rd point

168 — 60

c_1	c_2	3rd point
(185, 72)	(170, 56)	(168, 60)

$$D(c_1, 3) = \sqrt{(168-185)^2 + (60-72)^2}$$

$$= 20.80$$

$$D(c_2, 3) = \sqrt{(168-170)^2 + (60-56)^2}$$

$$= 4.4$$

Since the distance of point 3 is less for c_2, so it below to c_2.

c_2 update

$$\frac{170+168}{2} = 169$$

$$\frac{56+60}{2} = 58$$

$C_2 = (169, 58)$

4th point

$(179, 68)$

$$D(C_1, 4) = \sqrt{(179-185)^2 + (68-72)^2}$$

$$= 7.2$$

$$D(C_2, 4) = \sqrt{(179-169)^2 + (68-58)^2}$$

$$= 14.\ldots$$

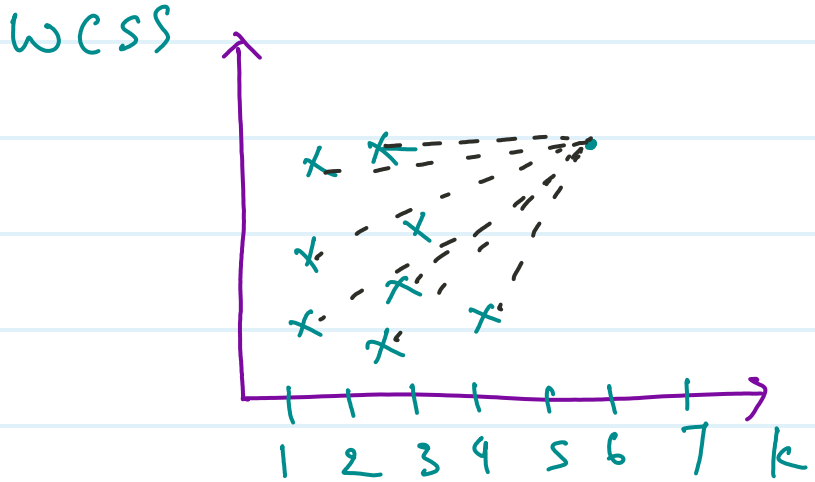
'C_1 update

$$\frac{185 + 179}{2} = 182$$

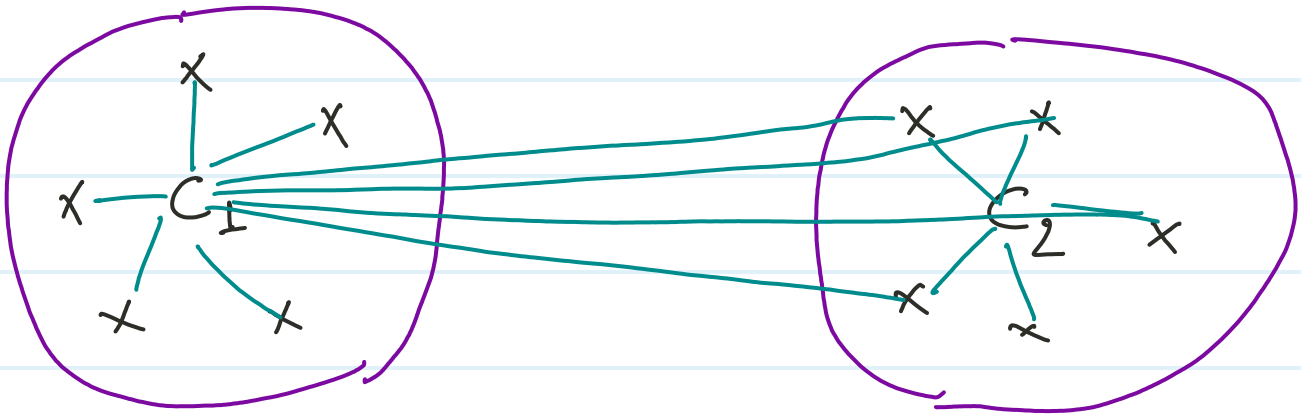
$$\frac{72 + 68}{2} = 70$$

$C_1 = (182, 70)$

How we can select k value means clusters how many clusters we need to select by elbow method.



wcss (withing clusters sum of square)

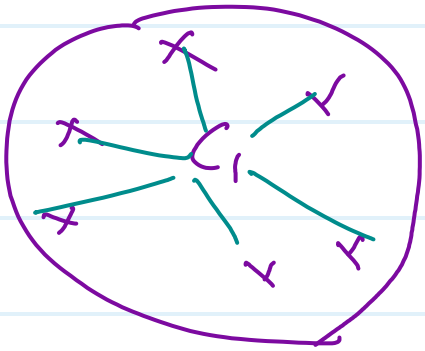


Intra cluster

Inter cluster.

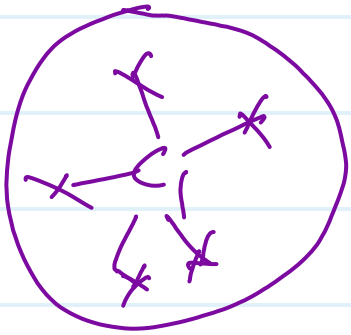
$$\Rightarrow \sum_{i=0}^{k} (\text{Diff. b/w centroid point})^2$$

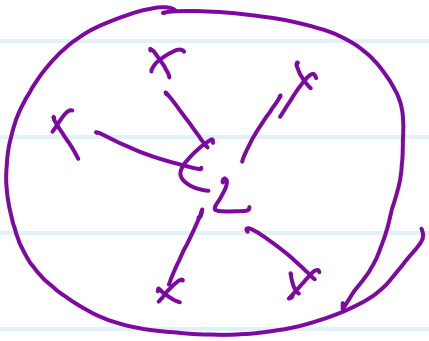
if $k = 1$



$wcss = 5$

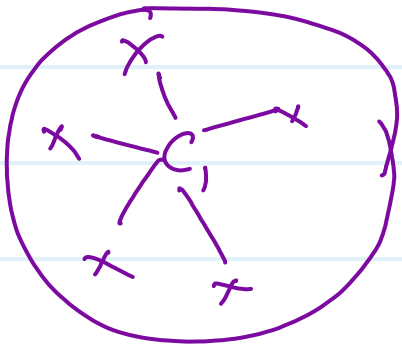
if $k = 2$

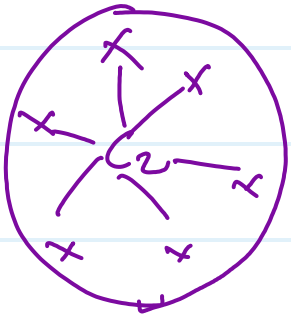


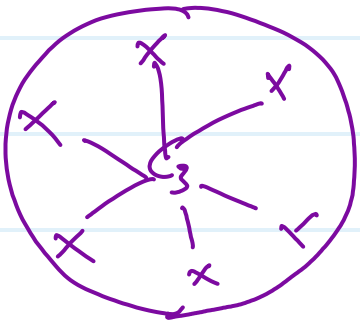


$wcss_1 > wcss_2$

If $k = 3$

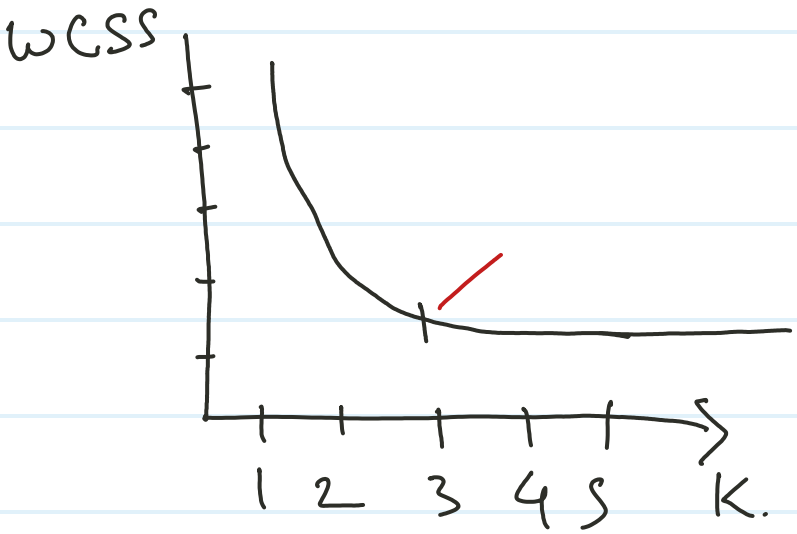




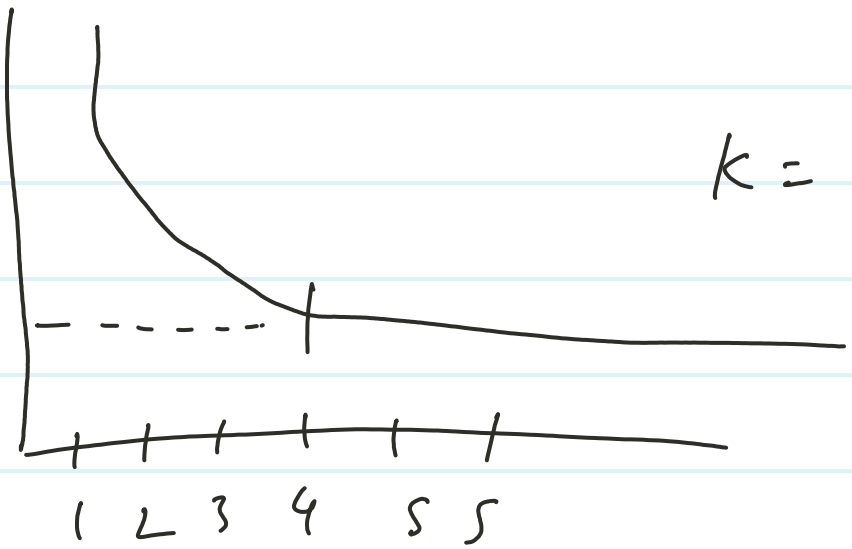


$wcss_1 > wcss_2 > wcss_3$

Elbow method



$k = 3$



$k = 4$

Performance Metrix

① Dunn Indexing
② silhout score

① Dunn Indexing

$$= \frac{\max \text{dist}(x_i, x_j)}{\max \text{dist}(y_i, y_j)}$$

② silhout score

$$= \frac{b_i - a_i}{\max(b_i - a_i)}$$

withing same clustes (intra) = a_i
inter clustes (inter) = b_i

Range of evaluation : −1 to +1

−1 worst

+1 best