✡ Unsupervised Learning

1. K-mean
2. DB scan
3. Hirarchical

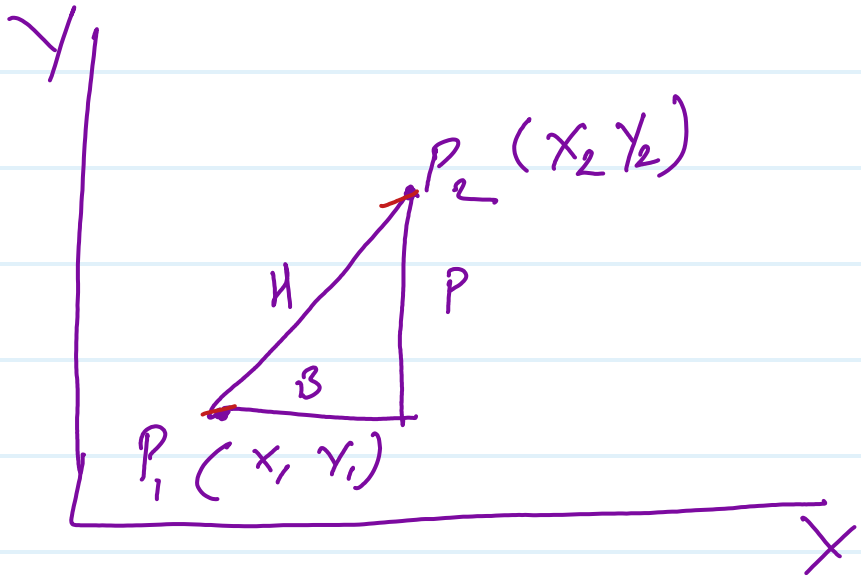
K-mean clusterjng

	Height	weight	cluster
1	185.	72 .	
1	170.	56 .	
	168	60	
	179	68	
	182	72	
	188	77	
	180	71	
	160	70	
	183	84	
	180	88	

Centroid Based approach

$$H^2 = P^2 + B^2$$

$$H = \sqrt{P^2 + B^2}$$



$$D(P_1, P_2) = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

- (i) Centroid
- (ii) Distance
- (iii) mean

To find centroid we two different method

(i) Elbow
(ii) WCSS

Evaluation metrix

(i) Dann Index
(ii) silhote coeff.

For now we will take random centroid from the dataset

(i) 185 72 ✓
(ii) 170 56 ✓

1st point

168 - 60

$$C_1 = \sqrt{(168-185)^2 + (60-72)^2}$$

$$= \sqrt{(-17)^2 + (-12)^2}$$

$$= \sqrt{289 + 144}$$

$$\Rightarrow \sqrt{433}$$

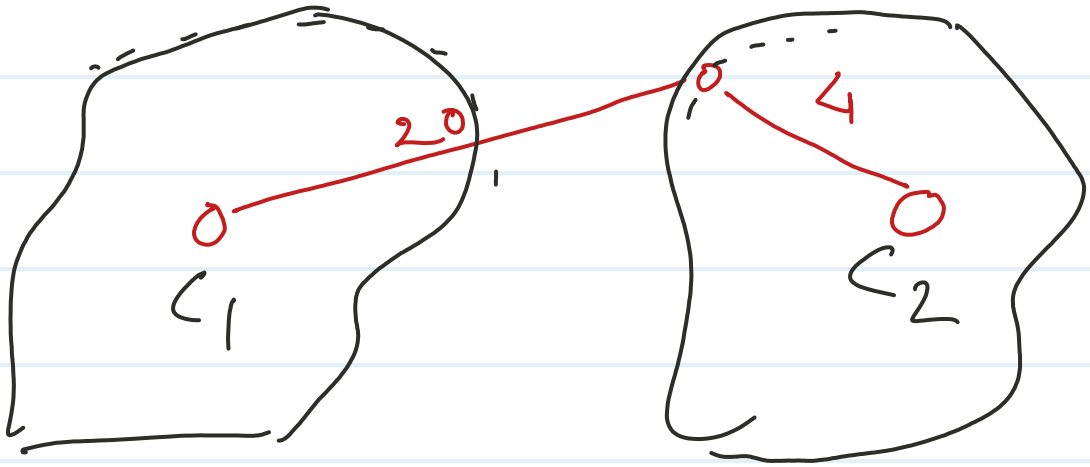
$$C_1 = 20.6$$

$$C_2 = \sqrt{(168-170)^2 + (60-56)^2}$$

$$\Rightarrow \sqrt{4 + 16}$$

$$\Rightarrow \sqrt{20}$$

$$C_2 = 4.4$$



For point 1 4.4 distance is from C_2 and 20.6 is from C_1, so the least distance is 4.4, so that the 1st point will be group with C_2.

Update value of C_2

$$\text{new } C_2 = \frac{168+170}{2}, \frac{56+60}{2}$$

$$\text{new } C_2 = (169, 58)$$

centroid

formula of WCSS

$$WCSS = \sum_{K=1}^{k} \sum_{i=1}^{n_K} \left[Distance(x_i, \mu_K) \right]^2$$

k = numbes of clustes

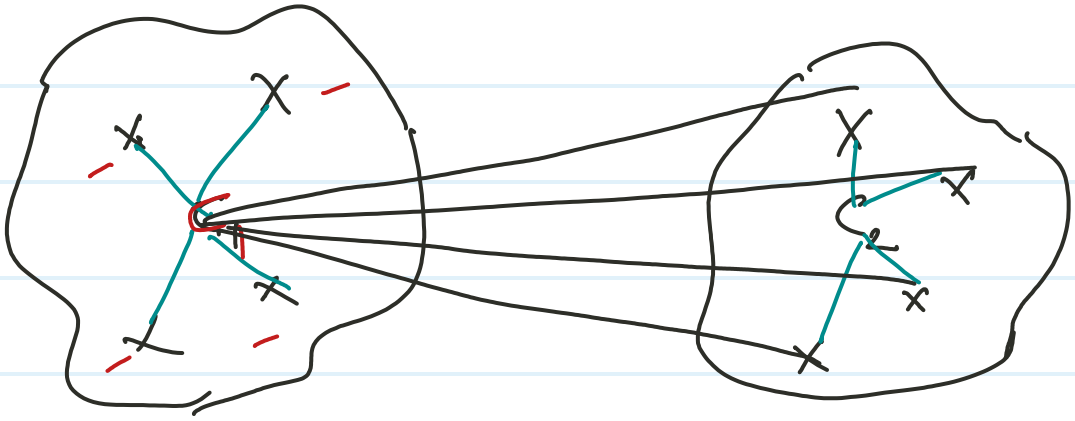
i = The datapoint in cluster k

μ_k = centroid of cluster k

n_k = The numbes of point in cluster k

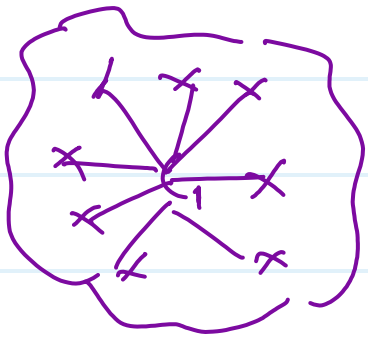
$Distance(x_i, \mu_k)$ = The Euclidient distance b/w data point x_i and assing cluster μ_k

WCSS (within cluster sum of square)



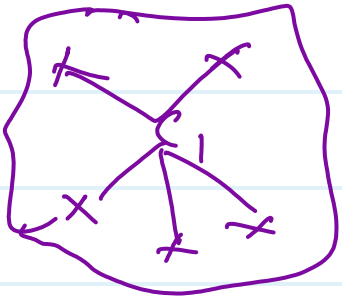
Intra cluster. Inter cluster.

$k = 1$

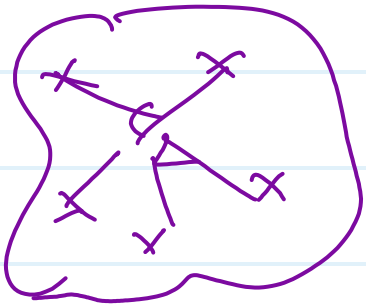


$wcss_1 = 5$

$k = 2$



$wcss_1$



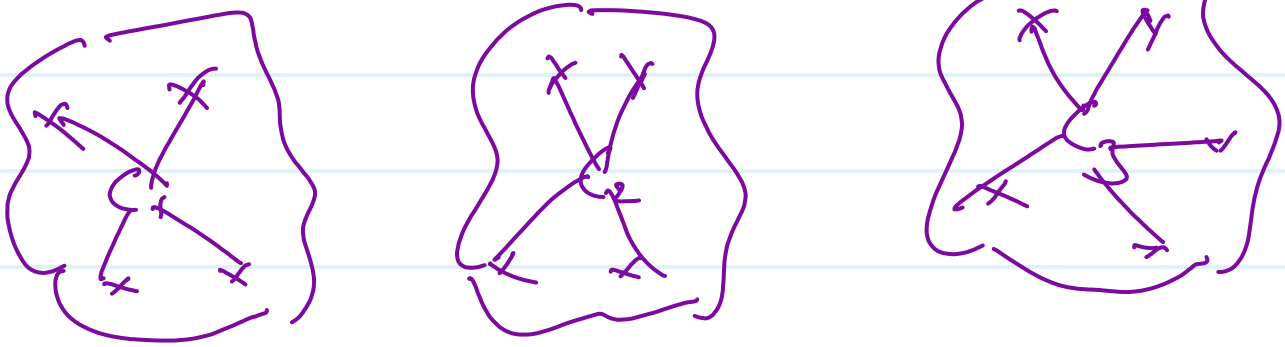
$wcss_2$

$WCSS_1 > WCSS_2$

4.4 2.5

$WCSS = 2.5$

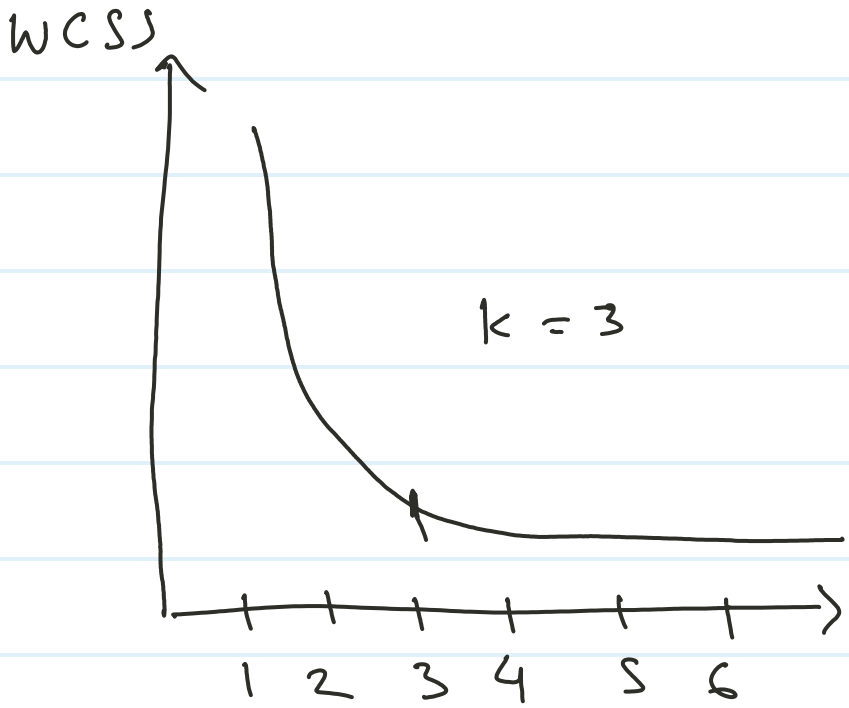
$K = 3$



w_1 w_2 w_3

$WCSS_1 > WCSS_2 > WCSS_3$

$WCSS_3 = 2$



✱ Dunn Indexing. —

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

✱ silhout coeffi.

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

a_i = intra clusty

b_i = inter cluster

intra dust

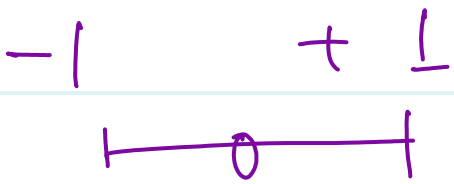
inter dust

Range of silhout score is

-1 to +1

-1 = worst

+1 = best



0.4

-0.7

* Agglomerative clustering (Hirerchical clustering)

A, B, C, D, E, F, G,