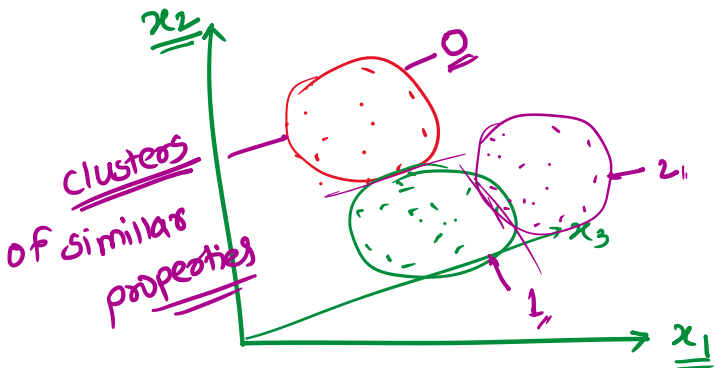
KMean

26 May 2023 07:25 AM

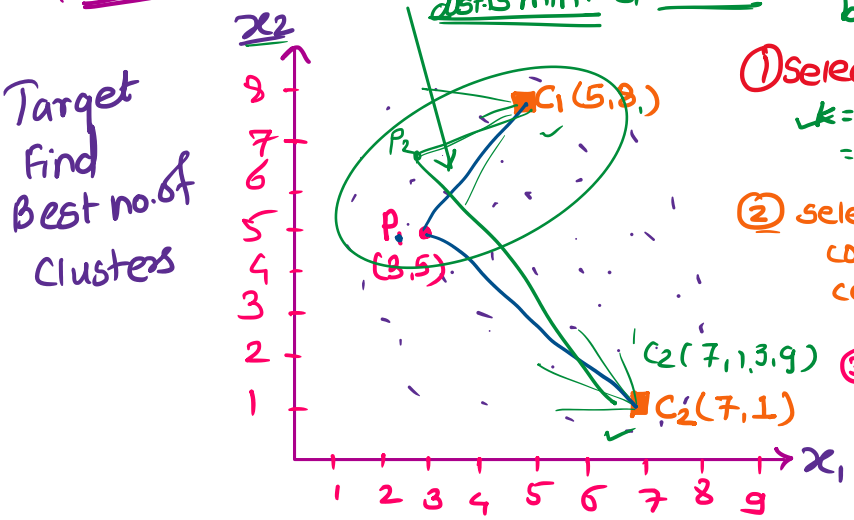


plotely

$x \rightarrow$ | Target

Supervised

$|x_1\ x_2|x_3\ x_4$



dist. is min. $C_1(5.8, 7.3)$

by default = 8

① Select value of K=2
k = no. of clusters
= no. of centroids

② select random co-ordinates of centroid k.

③ calculate dist. bet^n each centroid and all the dt. pts.

Euclidean Distance

$$= \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

C_1P

$= \sqrt{(5-3)^2+(8-5)^2+(\ )}$

$= \sqrt{4+9}$

$= \sqrt{13}$

C_2P

$= \sqrt{(3-7)^2+(1-5)^2}$

$= \sqrt{16+16}$

$= \sqrt{32}$

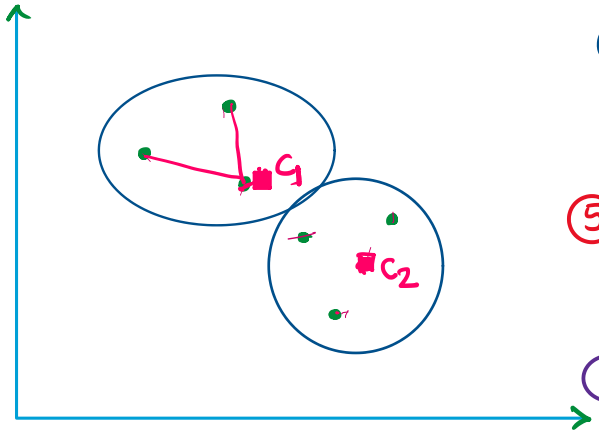
Smaller

randomly $C_1(7,3)$ $C_2(7,5)$

K=2

Data Points	x_1	x_2	C_1P	C_2P
P_1	5	2	2.23	6.40
P_2	2	3	3	5.2
P_3	3	5	6.6	7.6
P_4	5	8	8.5	4.8
P_5	7	7	2.3	1.2
P_6	6	4	3	

P5	7	7	2.3	1.2 ✓
P6	6	4	3 ✓	8



④ Assign every dt. pt. to the closest clust of centroid.

⑤ update value of centroid by taking mean of clustered dt. pts.

⑥ keep iterating until there is no change in centroid value

Cluster 1 datapoints

P1 (5,4)
P2 (3,5)
P3 (5,8)
P4 (2,3)

Cluster 2 datapoints

P5 (7,8)
P6 (6,4)

↓

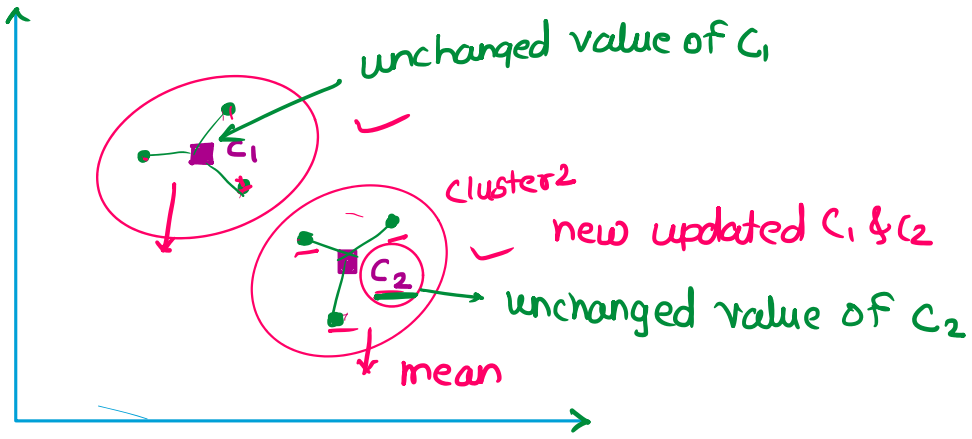
mean of all datapoints in a cluster to update centroid value.

updated $C_1 = \left(\frac{5+3+5+2}{4}, \frac{4+5+8+3}{4}\right)$

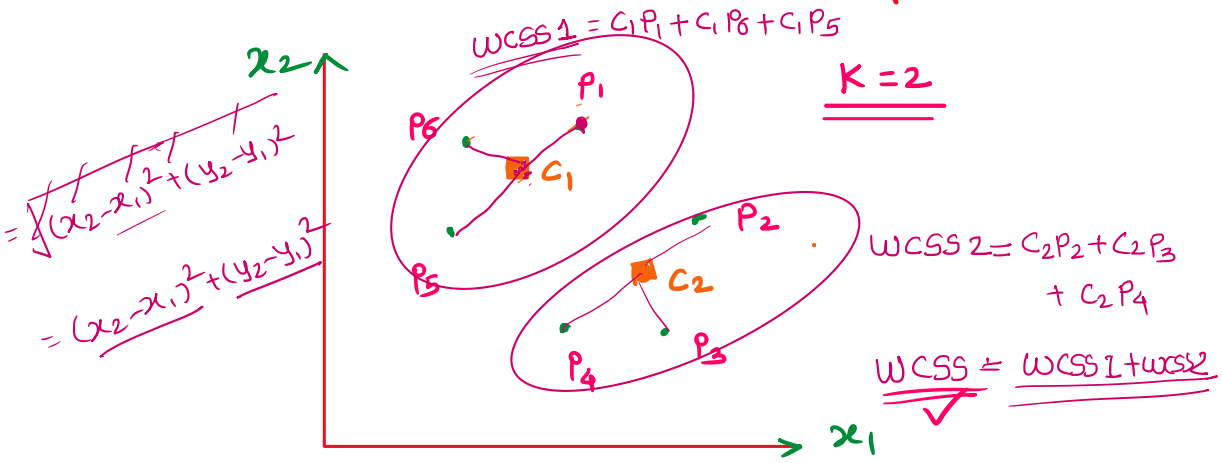
✓ $C_1 = (3.75, 5)$

updated $C_2 = \left(\frac{7+6}{2}, \frac{8+4}{2}\right)$

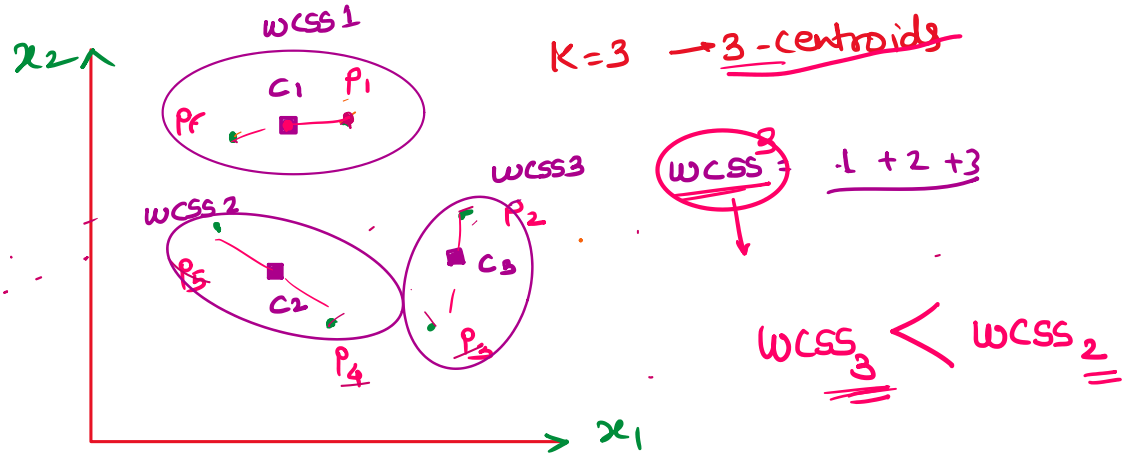
✓ $C_2 = (6.5, 6)$



WCSS distance → with in cluster sum of squares

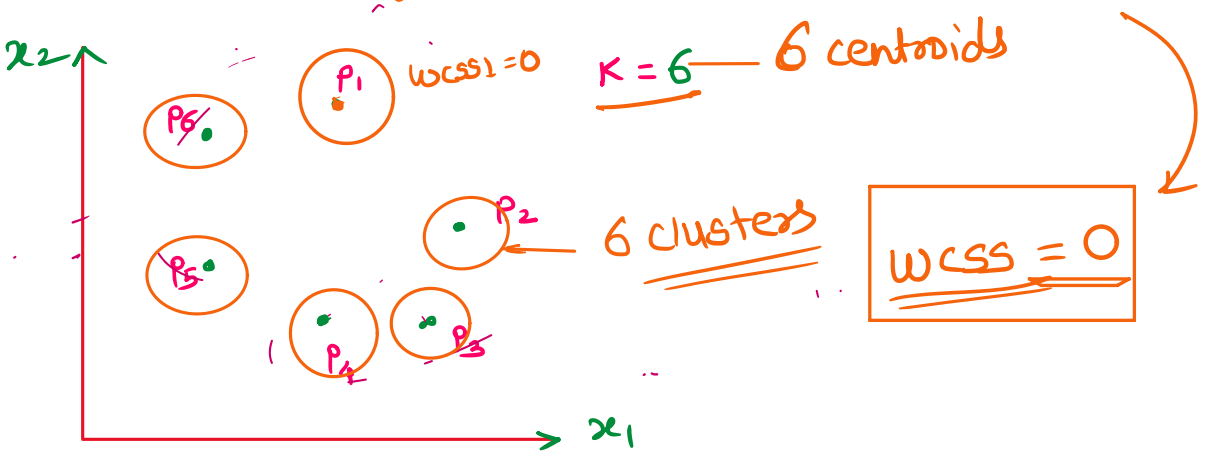


K = 2



K = 3 → 3-centroids

When no. of datapoints = no. of centroids



WCSS distance To find best value of K

Elbow Method

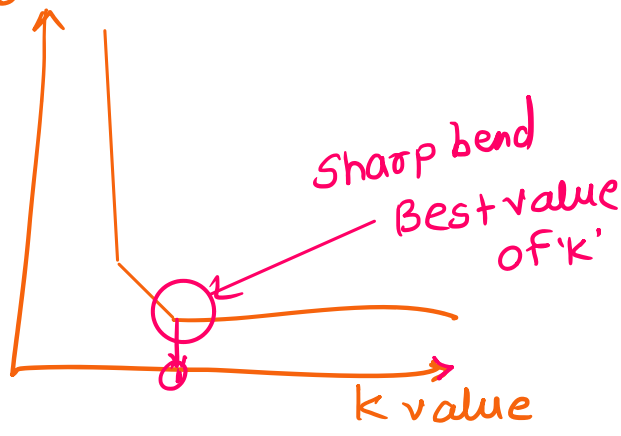
Elbow Method

No. of clusters

Elbow Method

k range (1 - 7) wcss

wcsss



Distance

scaling