

Linear Reg

↳ sup.
↳ Reg.

Log. Reg

↳ sup
↳ Classification

K-Means

* Unsupervised ML algorithm (I/p)

* Clustering (group) → Grouping of similar objects
in same group; dissimilar objects in other groups.

SML

Color	Size	Shape	Fruit
Red	-	-	Apple
Orange	-	-	Orange
Green	-	-	Gawra

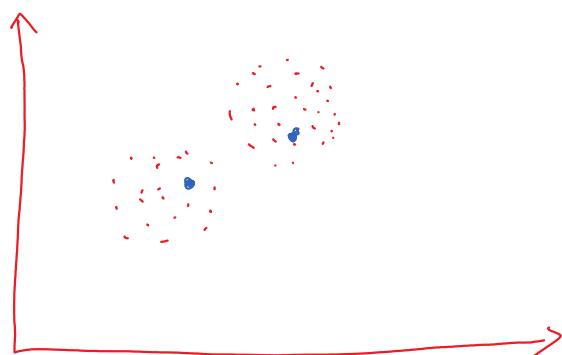
USML

Color	Size	Shape
-	-	-
-	-	-
-	-	-
-	-	-



How K-Means Clustering works?

K → Number of cluster.

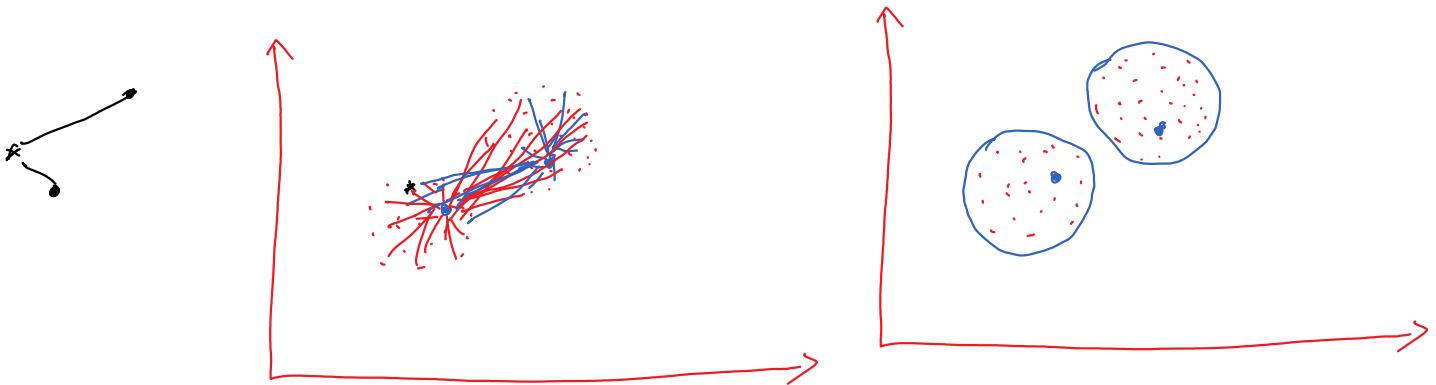


Step 1: Select number of clusters (K) (Say $K=2$)

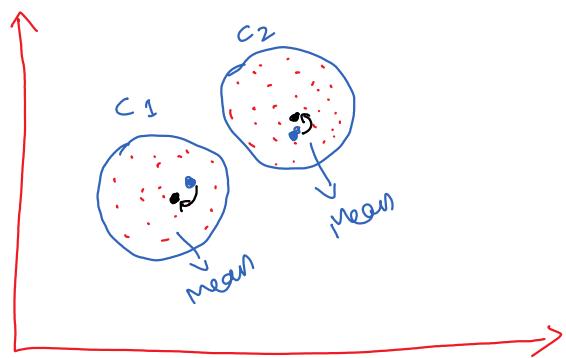
Step 2: Select 2 data points randomly & treat them as centroid (Middle point)

Step 3: calculate the distance b/w all the data pts with
→ the closest

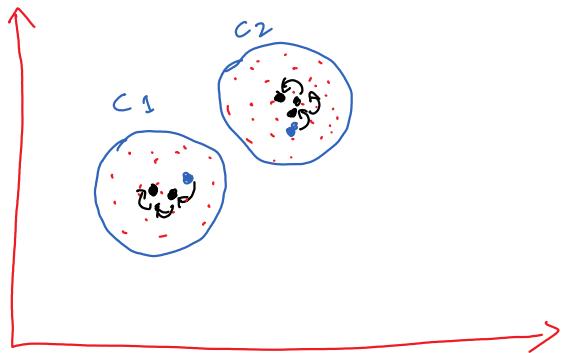
Step 3: ~~.....~~
centroid & assign the data pt to
centroid & make the cluster.



Step 4: Calculate mean value of the data pts for each
cluster & make the mean pt as new cluster.



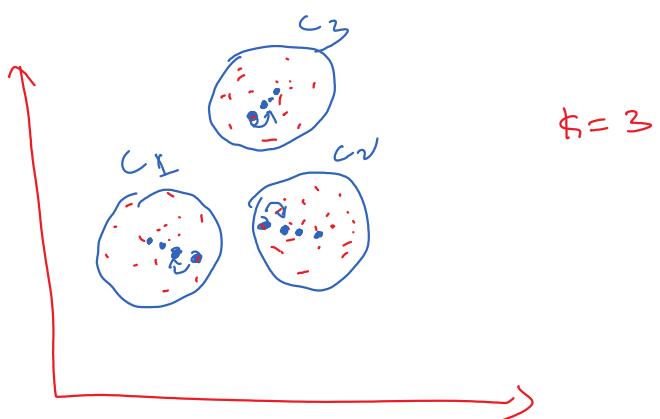
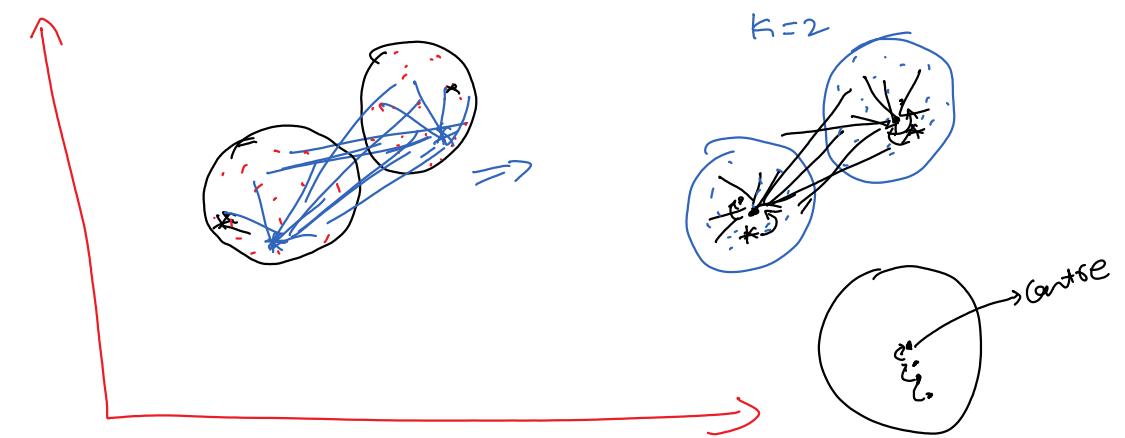
Step 5: Repeat step ③ & step ④ still there is no movement
of centroid



↑

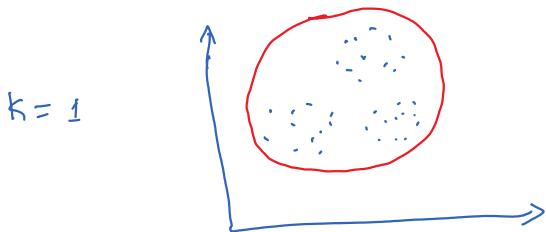
—

$k=2$ ↗

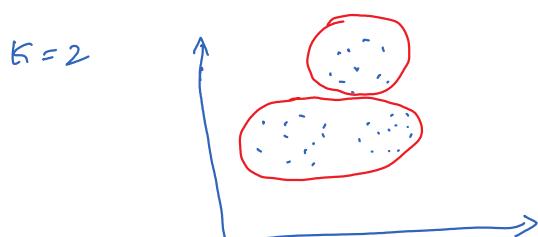


How to decide number of clusters?

* By Elbow Technique.

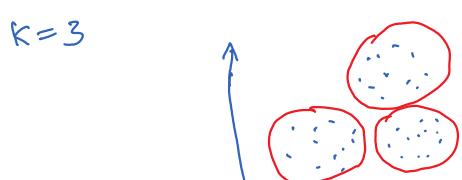


Error (WCSS)
 \downarrow
 WCSS
 Within Cluster
 square sum.



Error (WCSS)

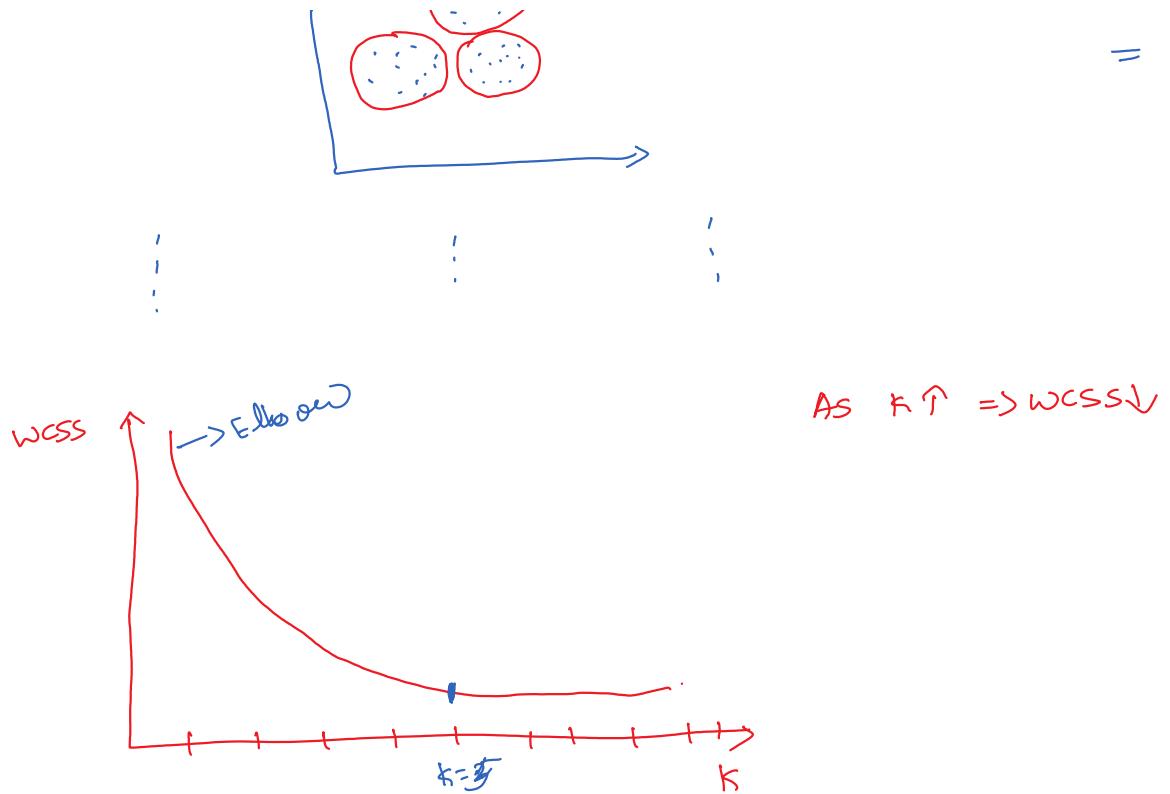
Inter cluster distance
 Intra cluster distance



Error (WCSS)

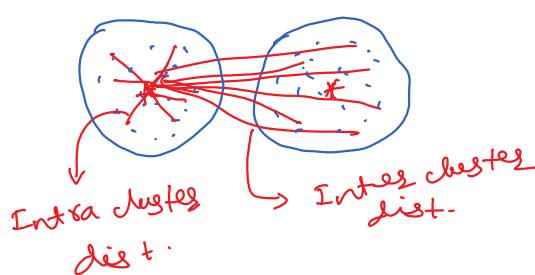
$$\text{WCSS} = \text{WCSS}_1 + \text{WCSS}_2$$

$$=$$



Evaluation Metric

Silhouette score (-1 to +1)



$$SS = \frac{\text{Inter} - \text{Intra}}{\max(b, a)}$$

$$\text{Intra cluster dist} < \text{Inter cluster dist} \Rightarrow SS > 0$$

$$\text{Intra cluster dist} > \text{Inter cluster dist} \Rightarrow SS < 0$$

$[-1, +1] \Rightarrow [0, 1]$ More accurate

E.g., $a = 0.2$ & $b = 0.8$

$$SS = \frac{0.8 - 0.2}{\max(0.8, 0.2)} = \frac{0.6}{0.8} = 0.75 \quad \checkmark$$

$a = 0.8$ & $b = 0.2$

$$SS = \frac{0.2 - 0.8}{\max(0.8, 0.2)} = \frac{-0.6}{0.8} = -0.75 \quad \times$$

—