

## K-Means

\* Unsupervised ML algorithm (I/P)

\* Clustering (group)  $\rightarrow$  Grouping of similar objects in same group & dissimilar objects in other groups.

Linear Reg

$\rightarrow$  Sup.  
 $\rightarrow$  Reg.

Log. Reg

$\rightarrow$  Sup  
 $\rightarrow$  Classification

SML

I/P			O/P
Color	Size	Shape	Fruit
Red	—	—	Apple
Orange	—	—	Orange
Green	—	—	Gauva

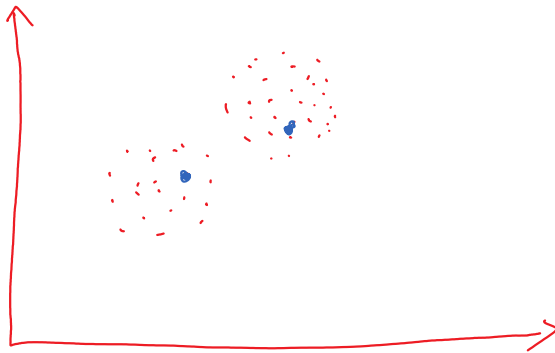
USML

Color	Size	Shape
—	—	—
—	—	—
—	—	—
—	—	—



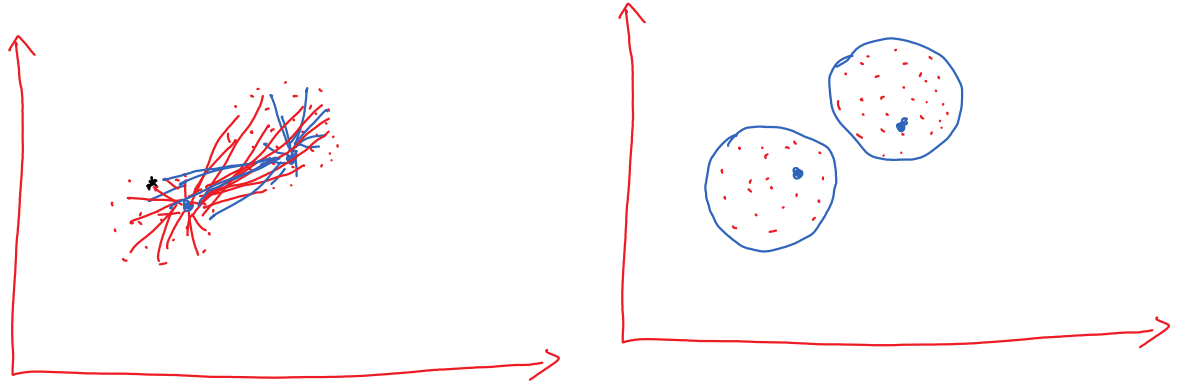
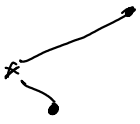
How K-Means Clustering works?

$K \rightarrow$  Number of clusters.

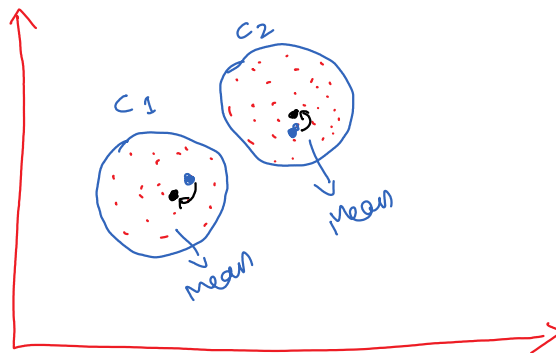


- Step 1: select number of clusters ( $K$ ) (say  $K=2$ )
- Step 2: select 2 data points randomly from them as centroid. (Middle points)
- Step 3: calculate the distance b/w all the data points with the centroid & assign them to the closest cluster.

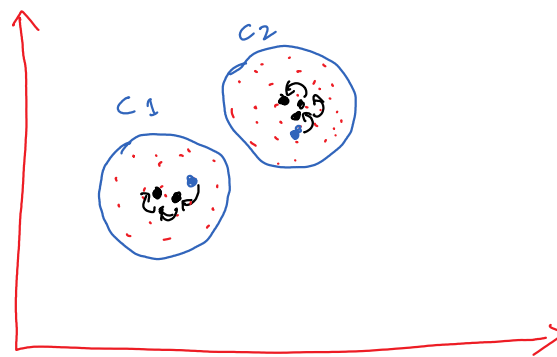
Step 1: ...  
 centroid & assign the datapoints to  
 centroid & make the cluster.



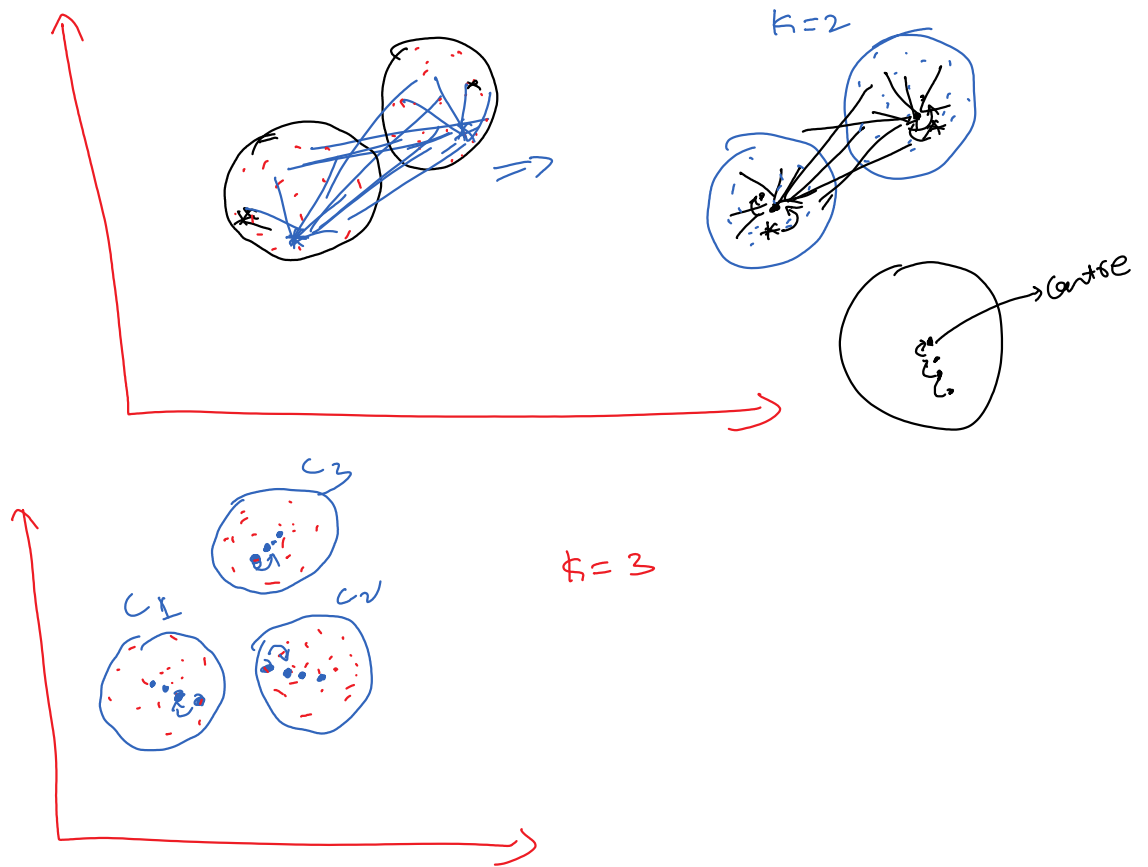
Step 2: Calculate mean value of the data pts for each  
 cluster & make the mean pt as new cluster.



Step 3 Repeat step 2 & step 4 still there is no movement  
 of centroid

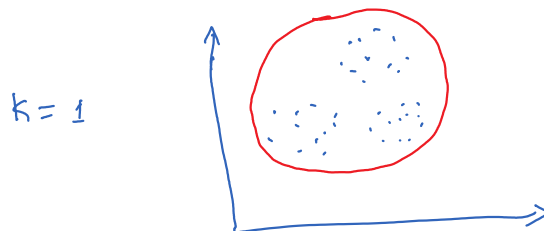


$K=2$



How to decide number of clusters?

\* By Elbow Technique.

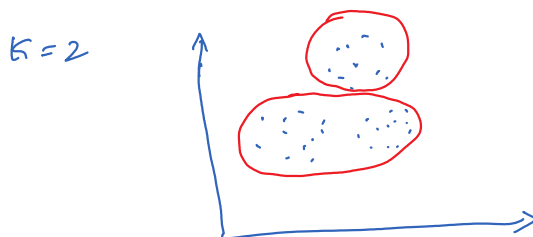


Error (WCSS)

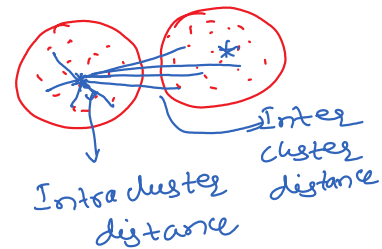
Error

↓  
WCSS

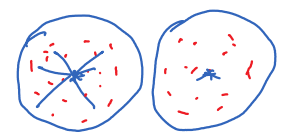
Within Cluster  
square sum.



Error (WCSS)

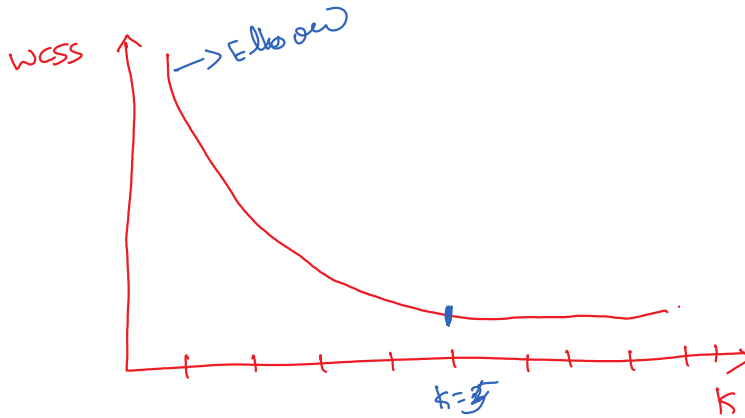
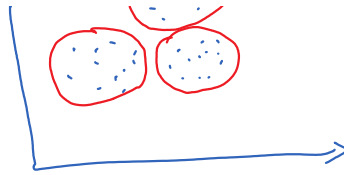


Error (WCSS)



$$WCSS = WCSS_1 + WCSS_2$$

$$=$$

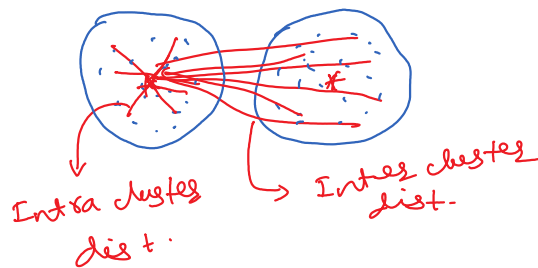


As  $k \uparrow \Rightarrow WCSS \downarrow$

## Evaluation Matrix

Silhouette score (-1 to +1)

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



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

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

$$[-1, +1] \Rightarrow [0, 1] \text{ more accurate}$$

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

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

$a = 0.8$  &  $b = 0.2$

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

—