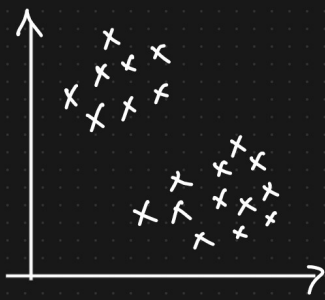
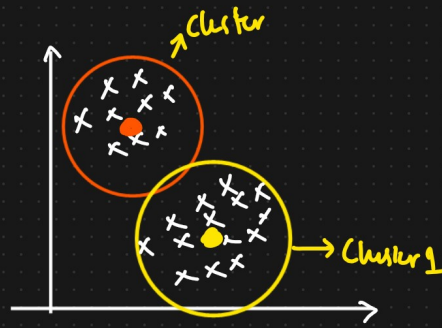


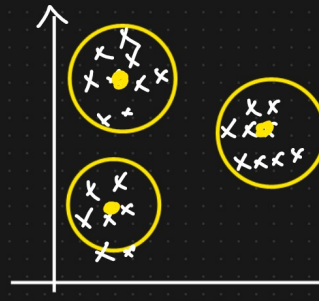
K Means Clustering Algorithm



K Means
 \Rightarrow



\Rightarrow



Eucledian Distance }
OR
MANHATTAN Distance }

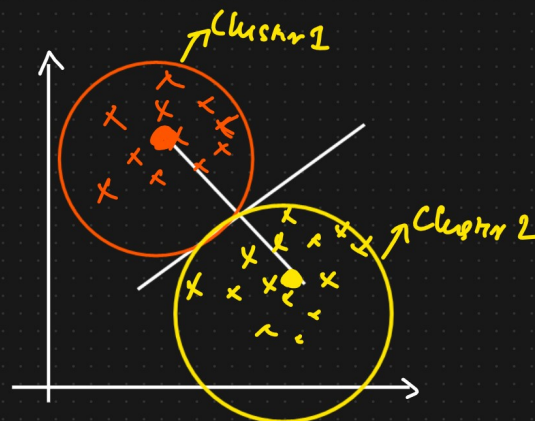
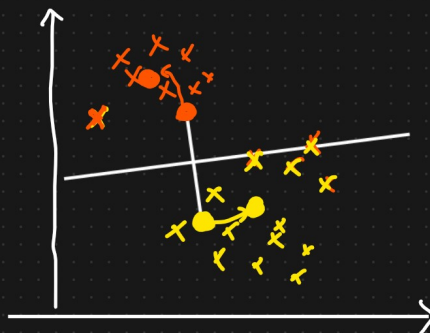
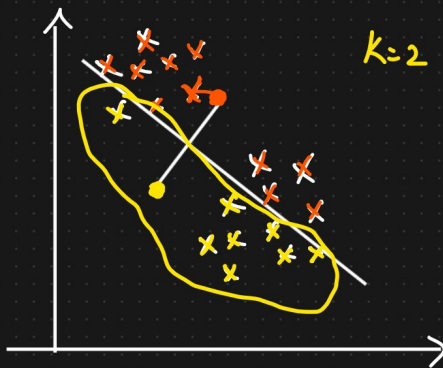
K=2

K Mean
 \equiv

① Initialize some K \rightarrow centroids

② Points that are nearest to the centroids \rightarrow Group

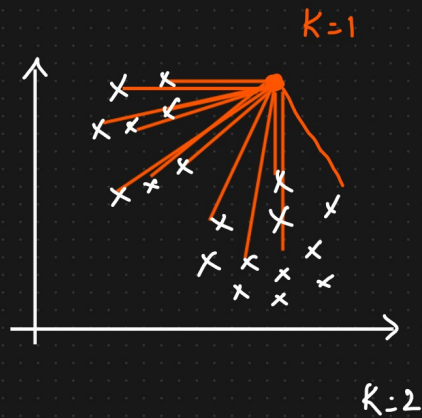
③ Move the centroids \rightarrow Average



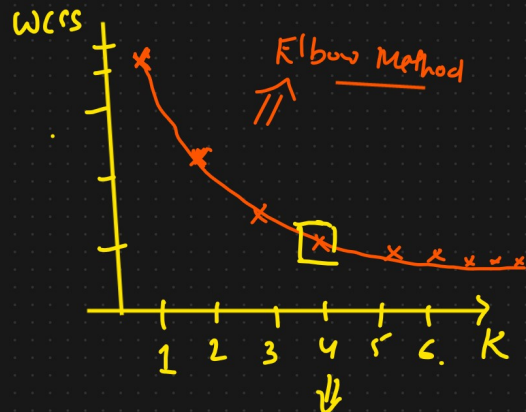
How do we select the K value?

WCSS = Within Cluster Sum of Squares

Initialize $K=1$ to 20

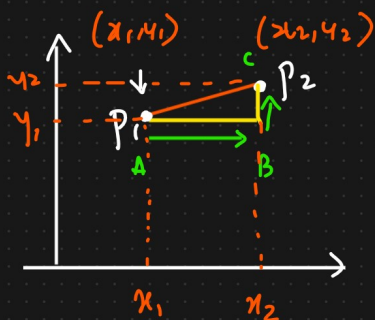


$$WCSS = \sum_{i=1}^n \left(\text{distance between points to nearest centroid} \right)^2$$



WCSS \downarrow

⑧ Euclidean Distance

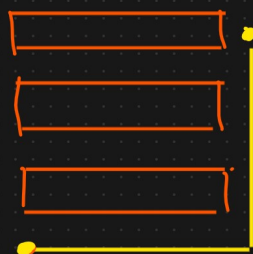


$$\text{Euclidean dist} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

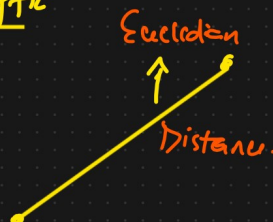
$$\text{Manhattan dist} = |x_2 - x_1| + |y_2 - y_1|$$

IRON MAN \rightarrow U.S

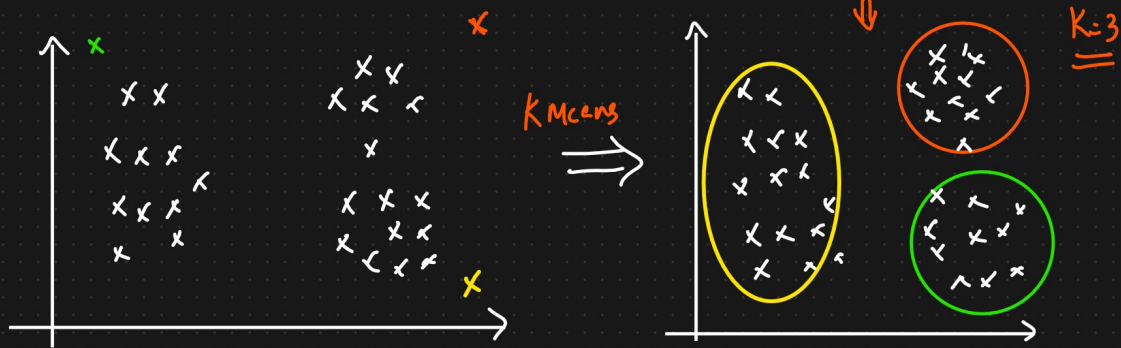
Air Traffic



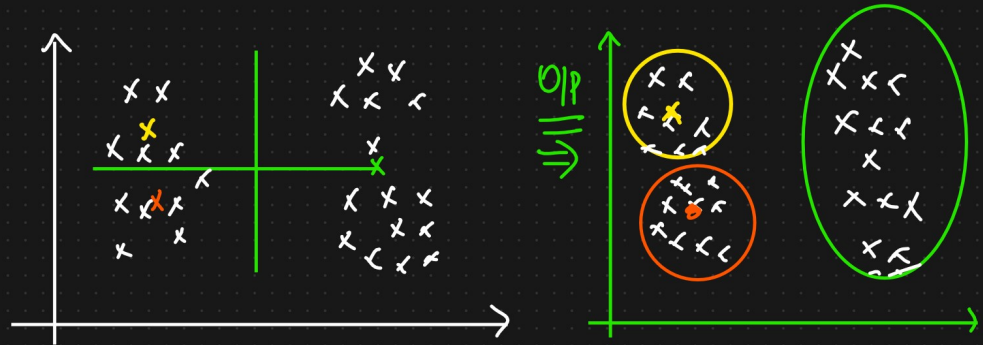
\Rightarrow Manhattan distance



Random Initialization TRAP (Kmeans++)



Random
Initialization



Kmeans++ Initialization Technique