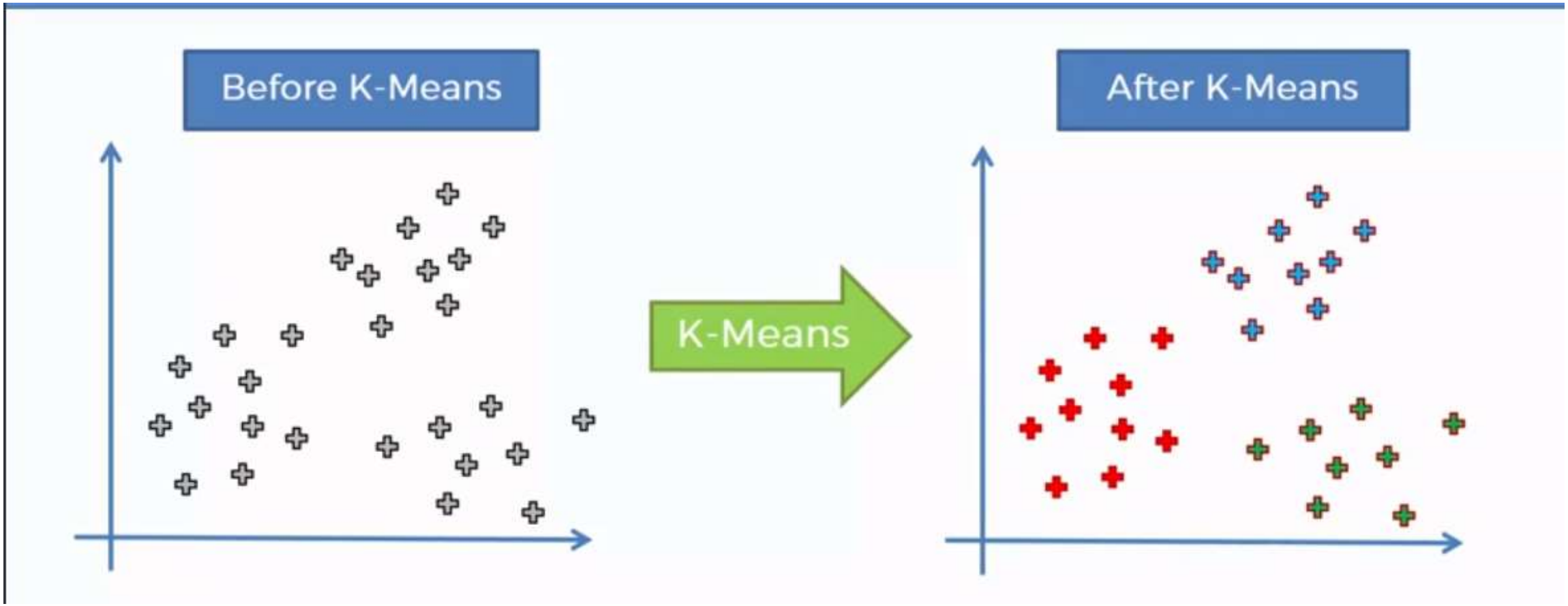
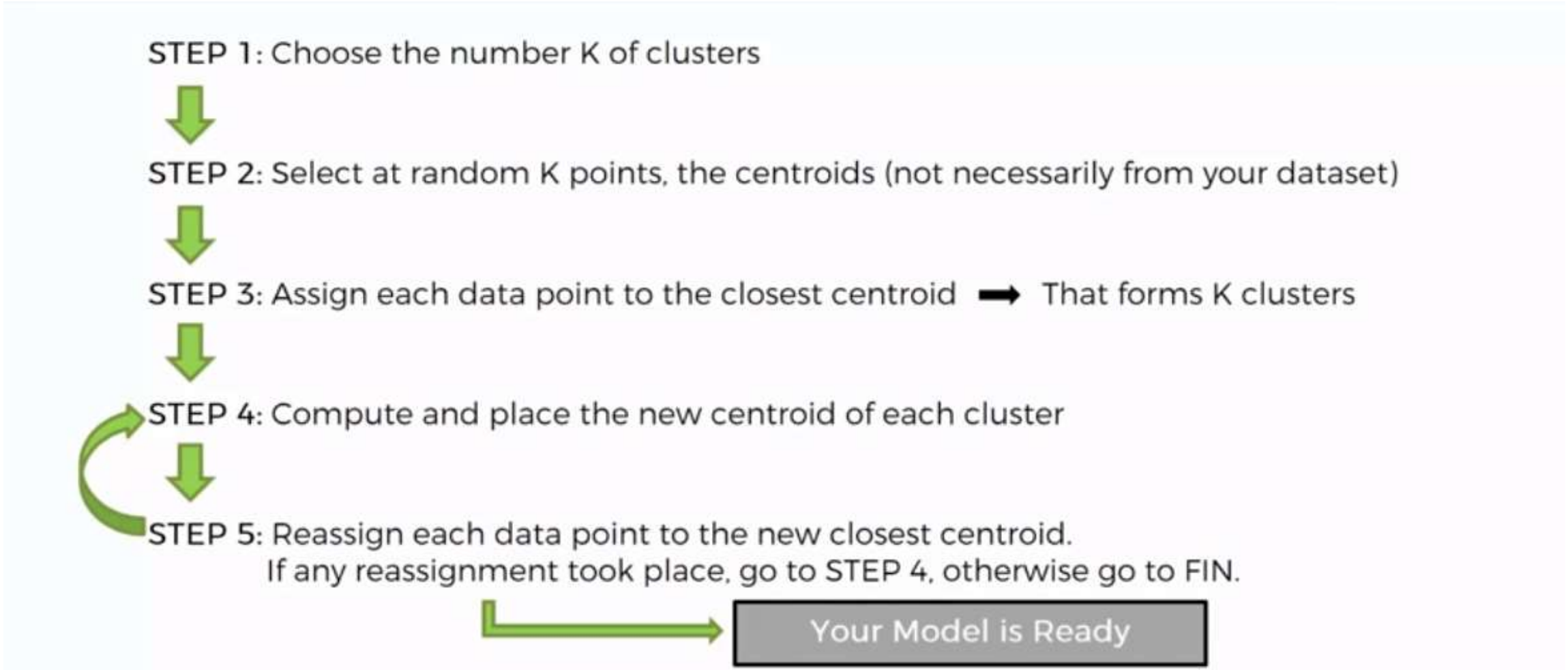
K-Means Clustering

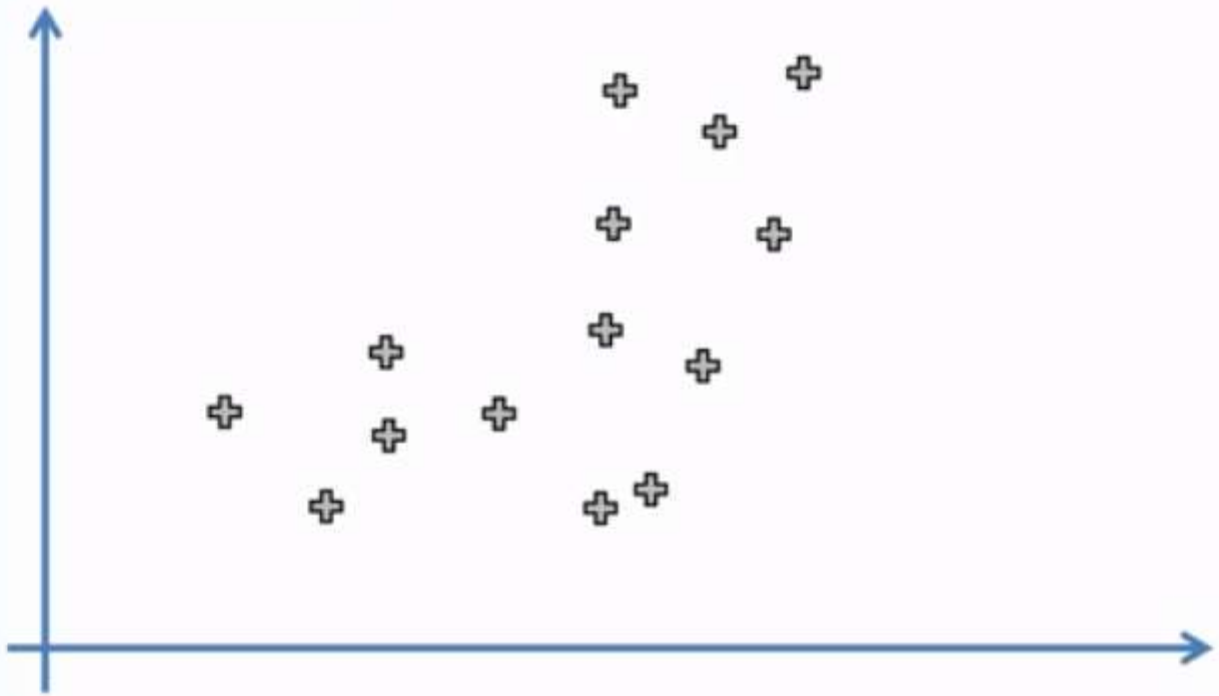


Sensitivity: Internal & Restricted



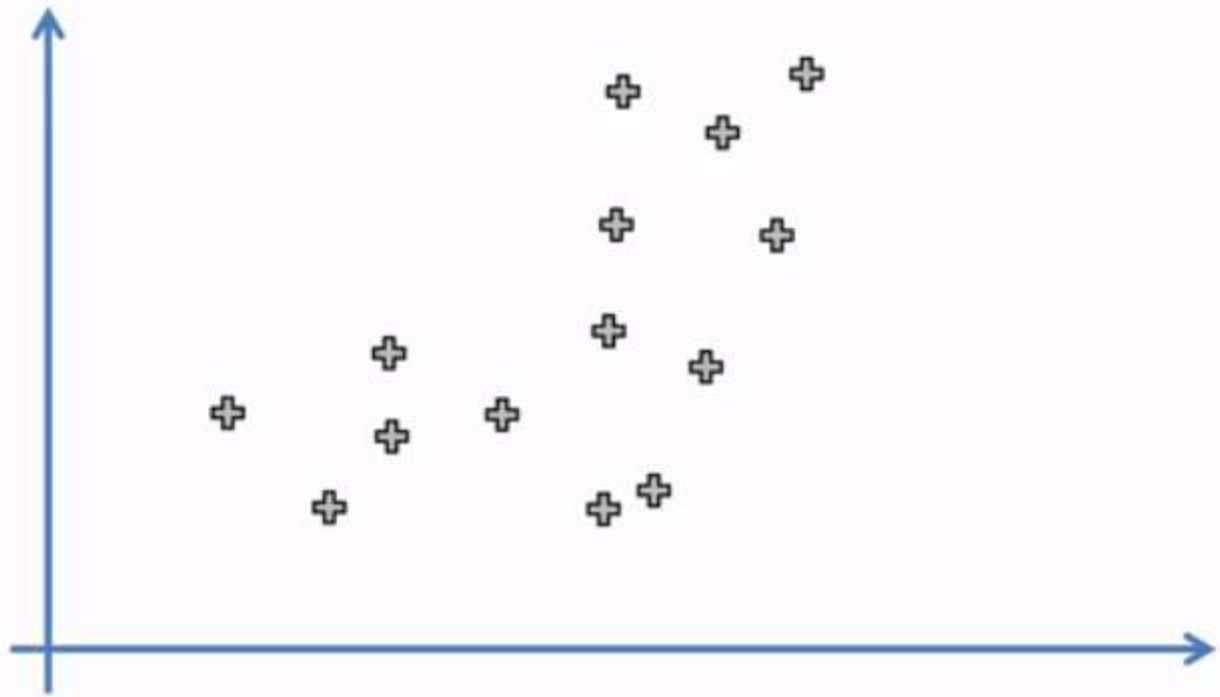
Sensitivity: Internal & Restricted

STEP 1: Choose the number K of clusters: K = 2



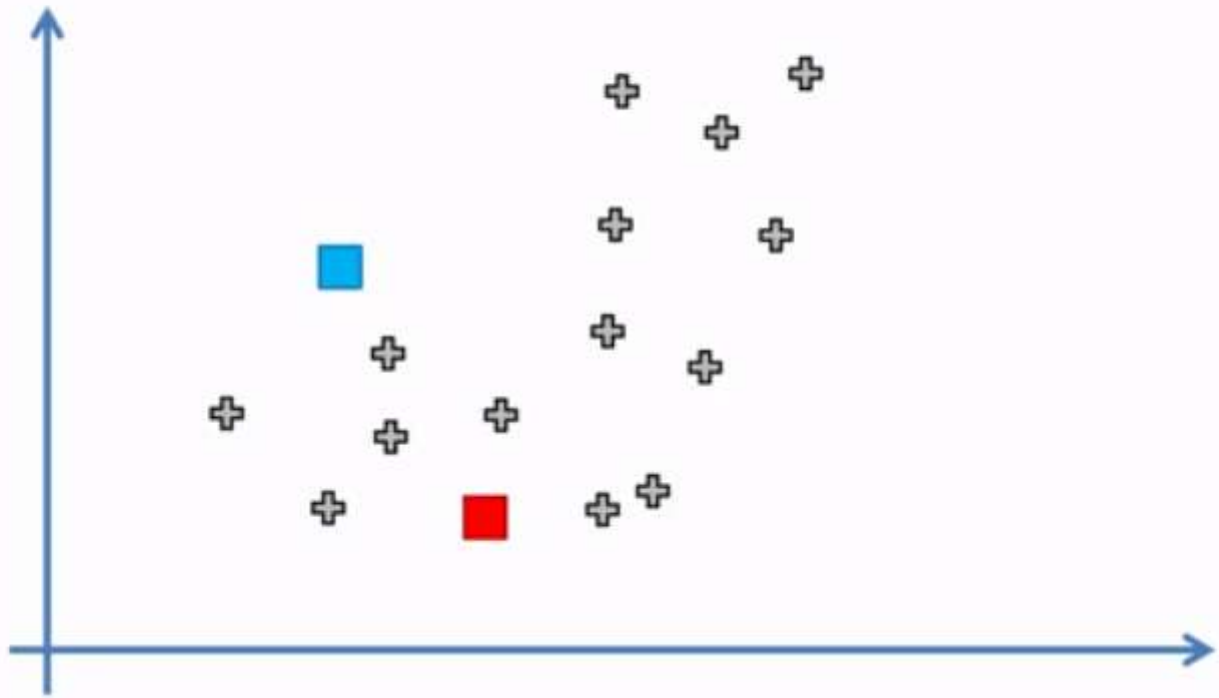
Sensitivity: Internal & Restricted

STEP 2: Select at random K points, the centroids (not necessarily from your dataset)



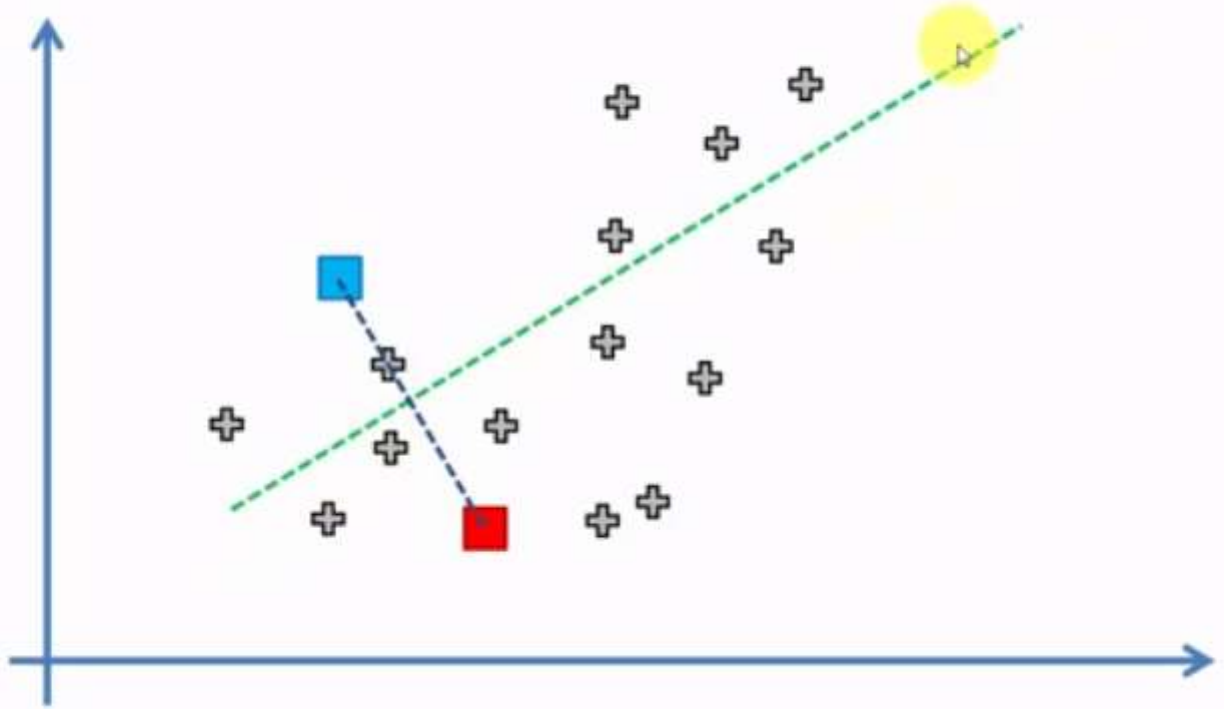
Sensitivity: Internal & Restricted

STEP 3: Assign each data point to the closest centroid ➡ That forms K clusters



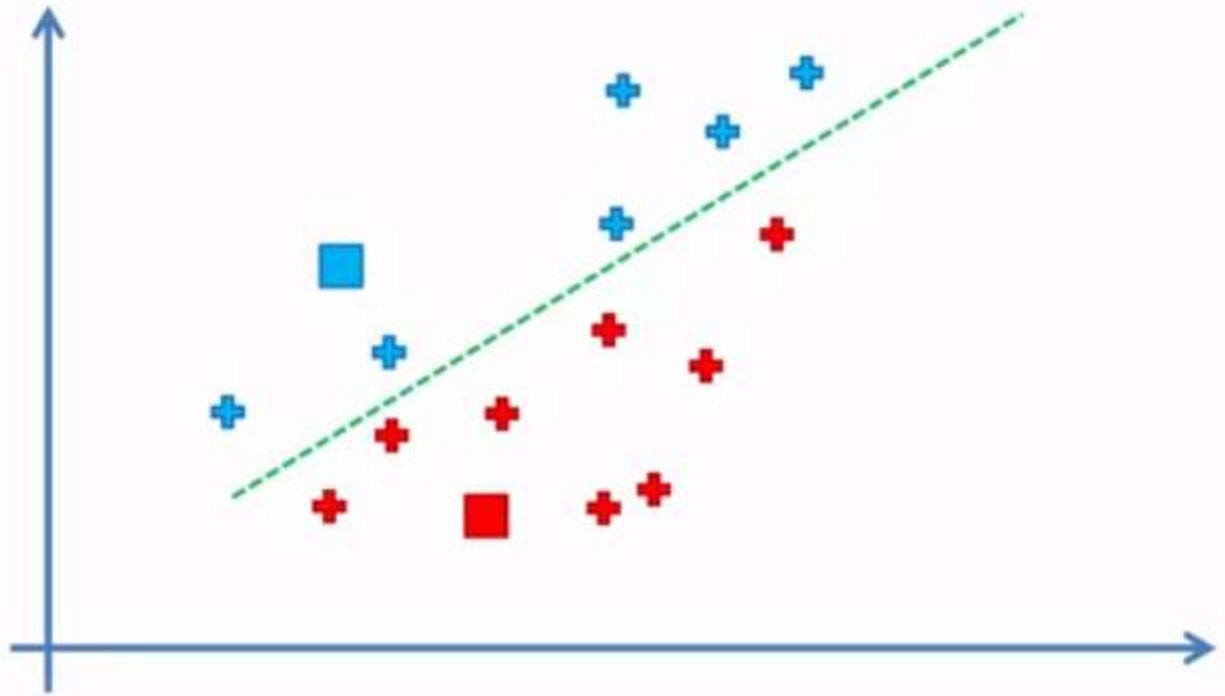
Sensitivity: Internal & Restricted

STEP 3: Assign each data point to the closest centroid ➡ That forms K clusters



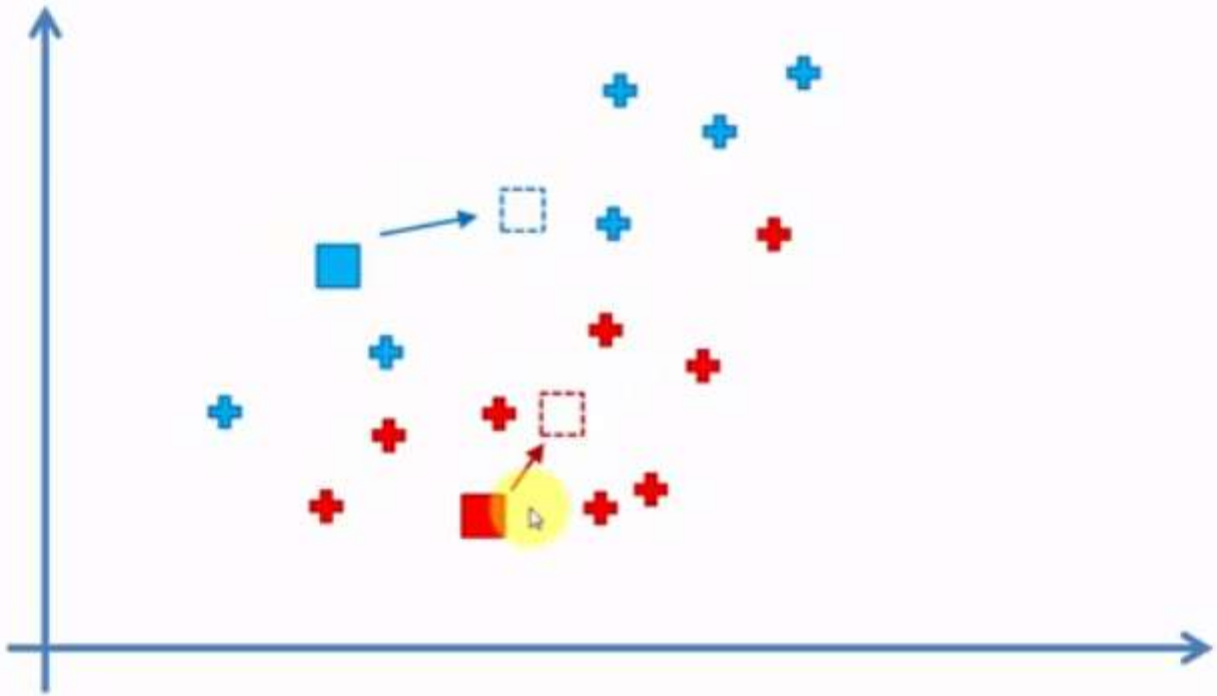
Sensitivity: Internal & Restricted

STEP 3: Assign each data point to the closest centroid ➡ That forms K clusters



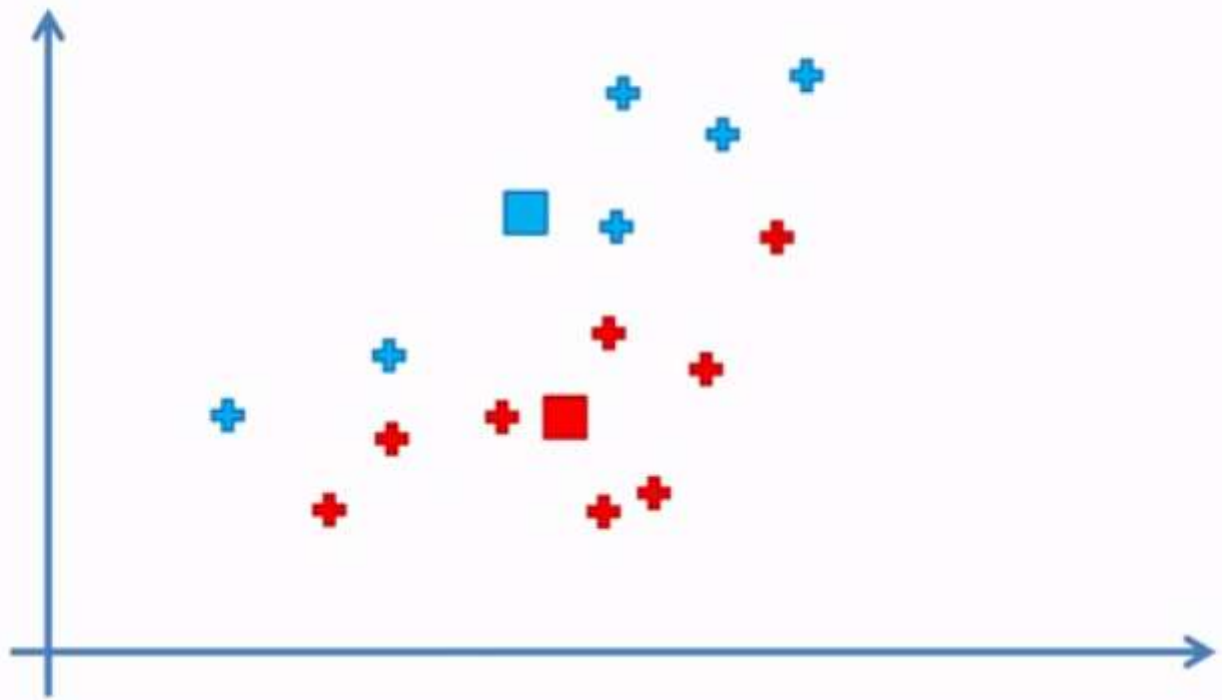
Sensitivity: Internal & Restricted

STEP 4: Compute and place the new centroid of each cluster



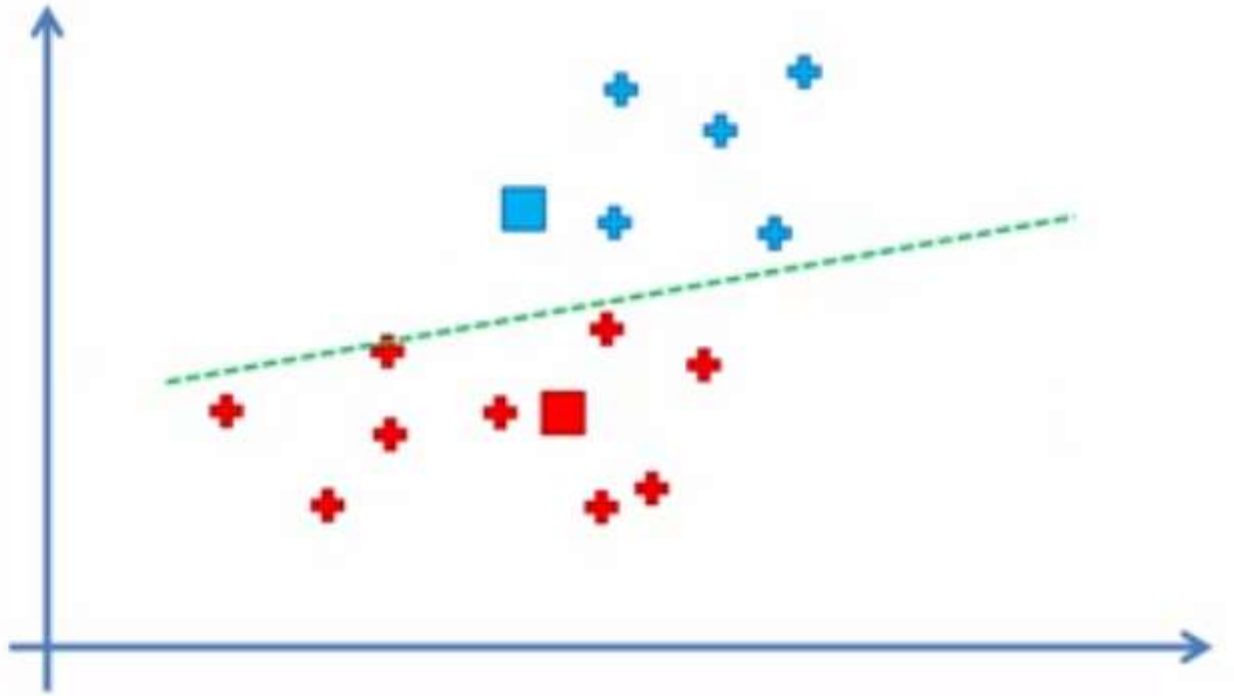
Sensitivity: Internal & Restricted

STEP 4: Compute and place the new centroid of each cluster



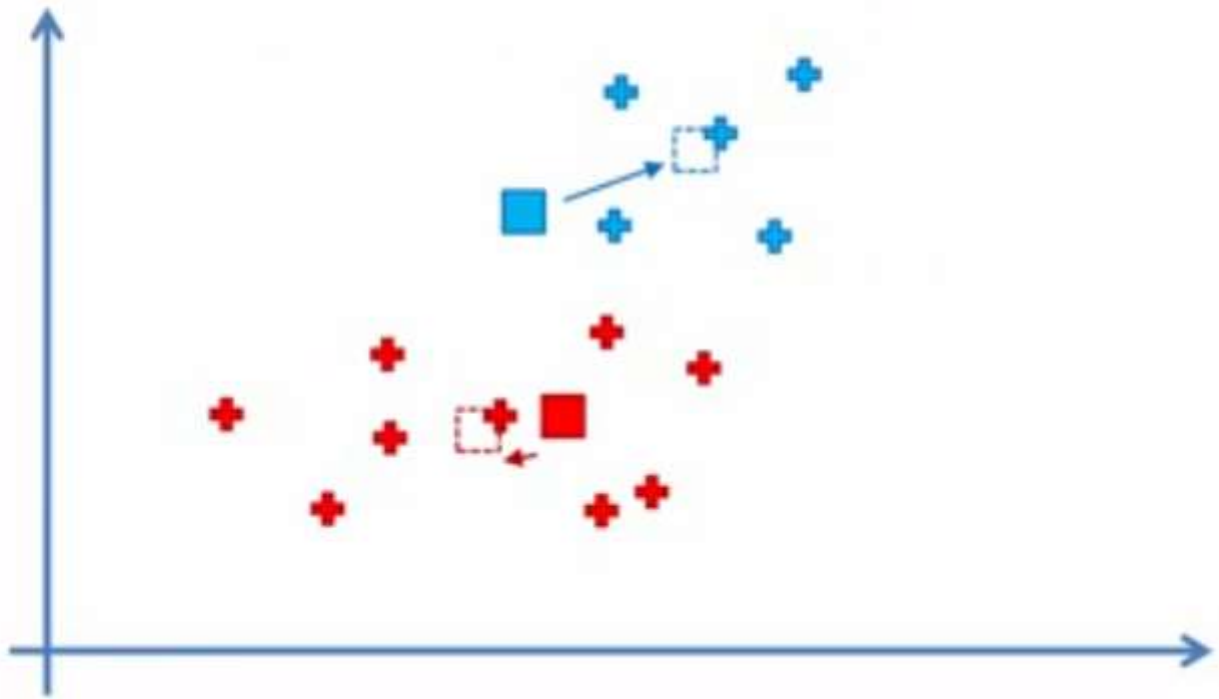
Sensitivity: Internal & Restricted

STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



Sensitivity: Internal & Restricted

STEP 4: Compute and place the new centroid of each cluster

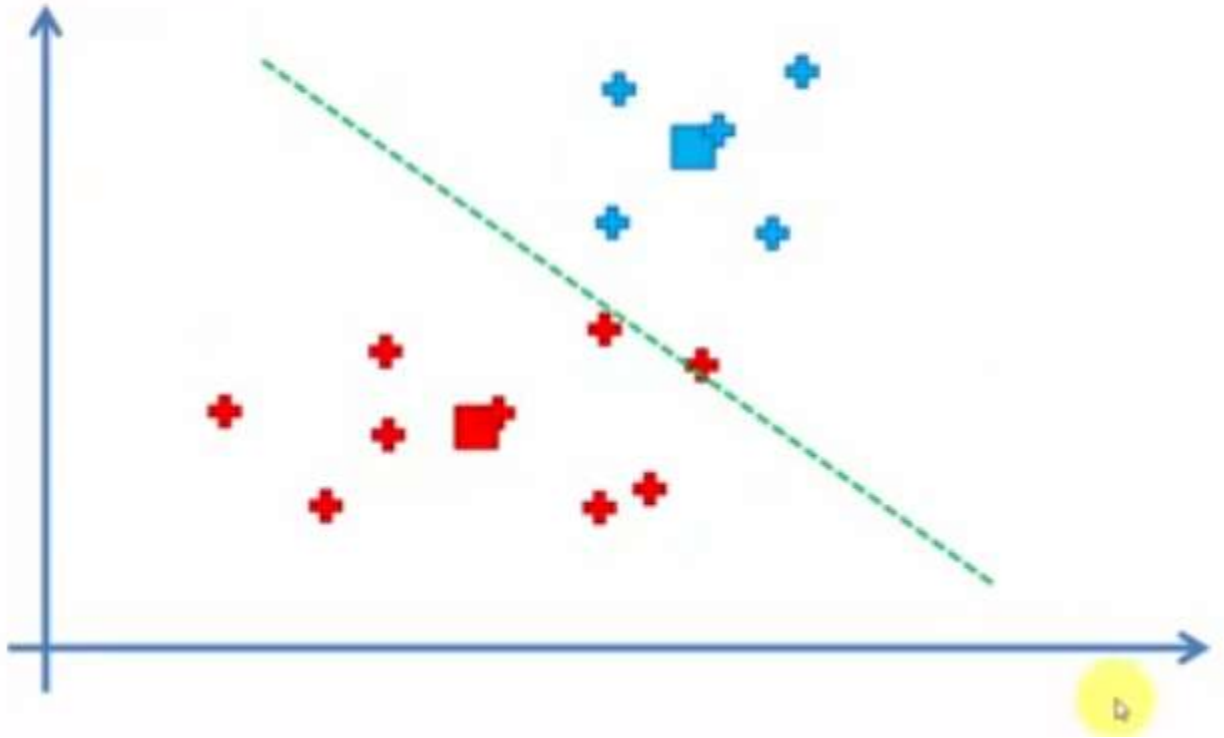


Sensitivity: Internal & Restricted

STEP 4: Compute and place the new centroid of each cluster

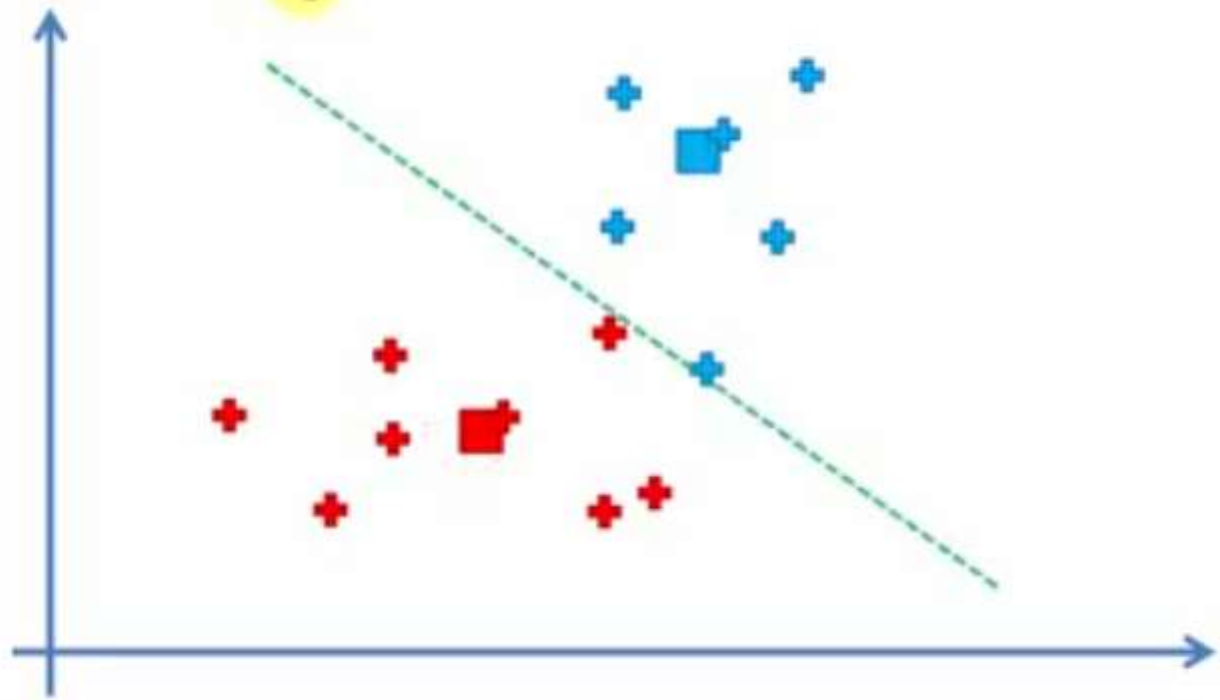
Sensitivity: Internal & Restricted

STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



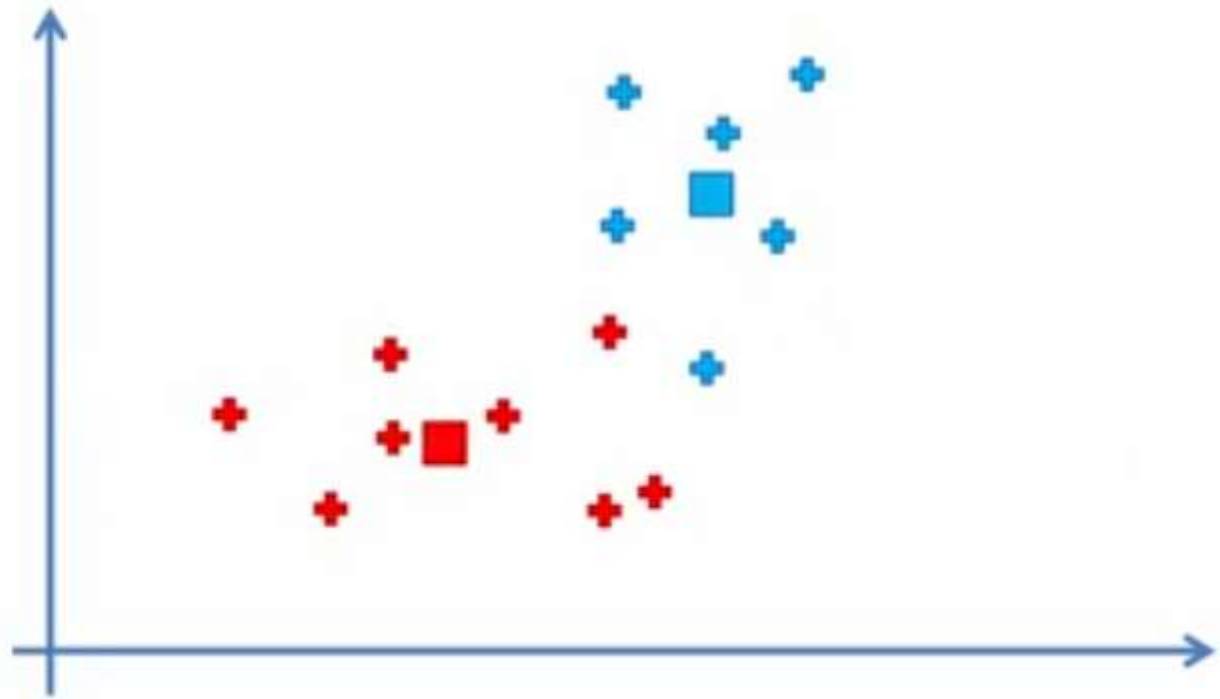
Sensitivity: Internal & Restricted

STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



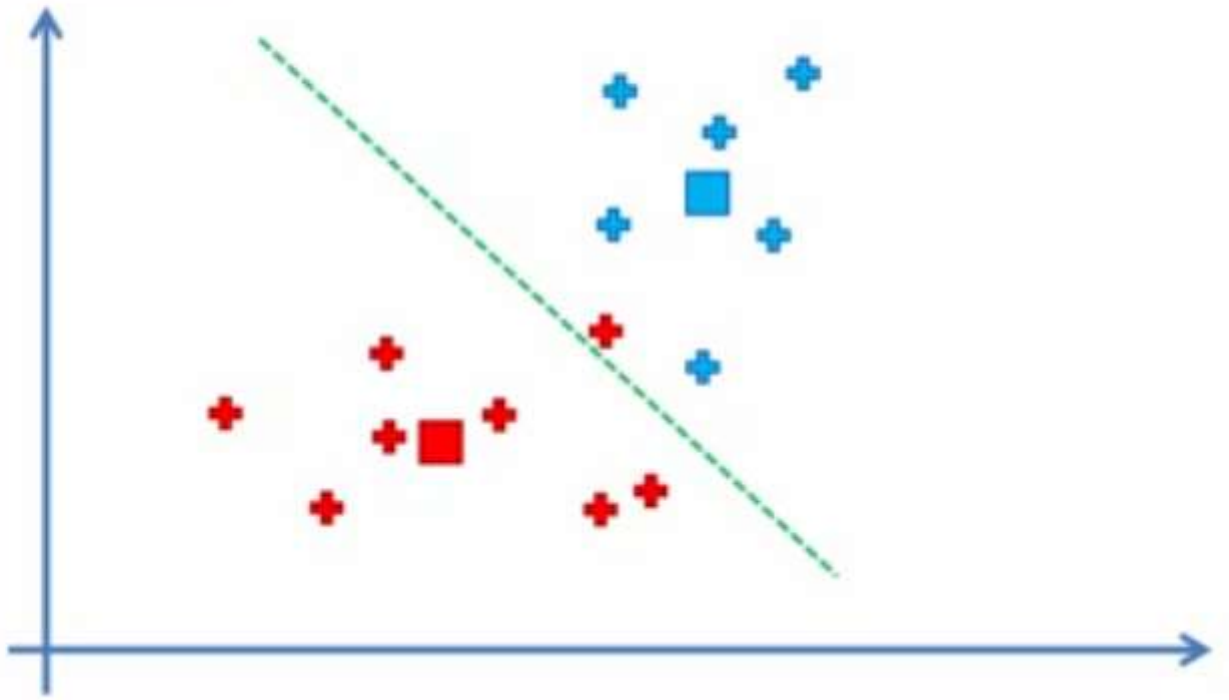
Sensitivity: Internal & Restricted

STEP 4: Compute and place the new centroid of each cluster



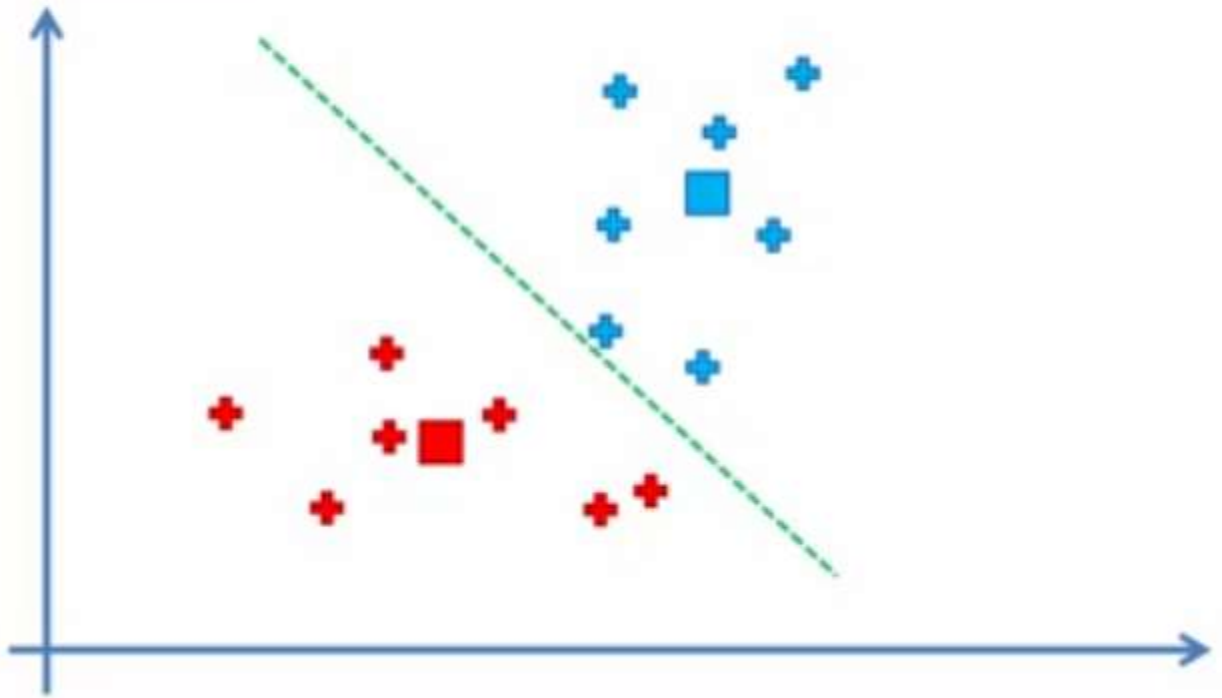
Sensitivity: Internal & Restricted

STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



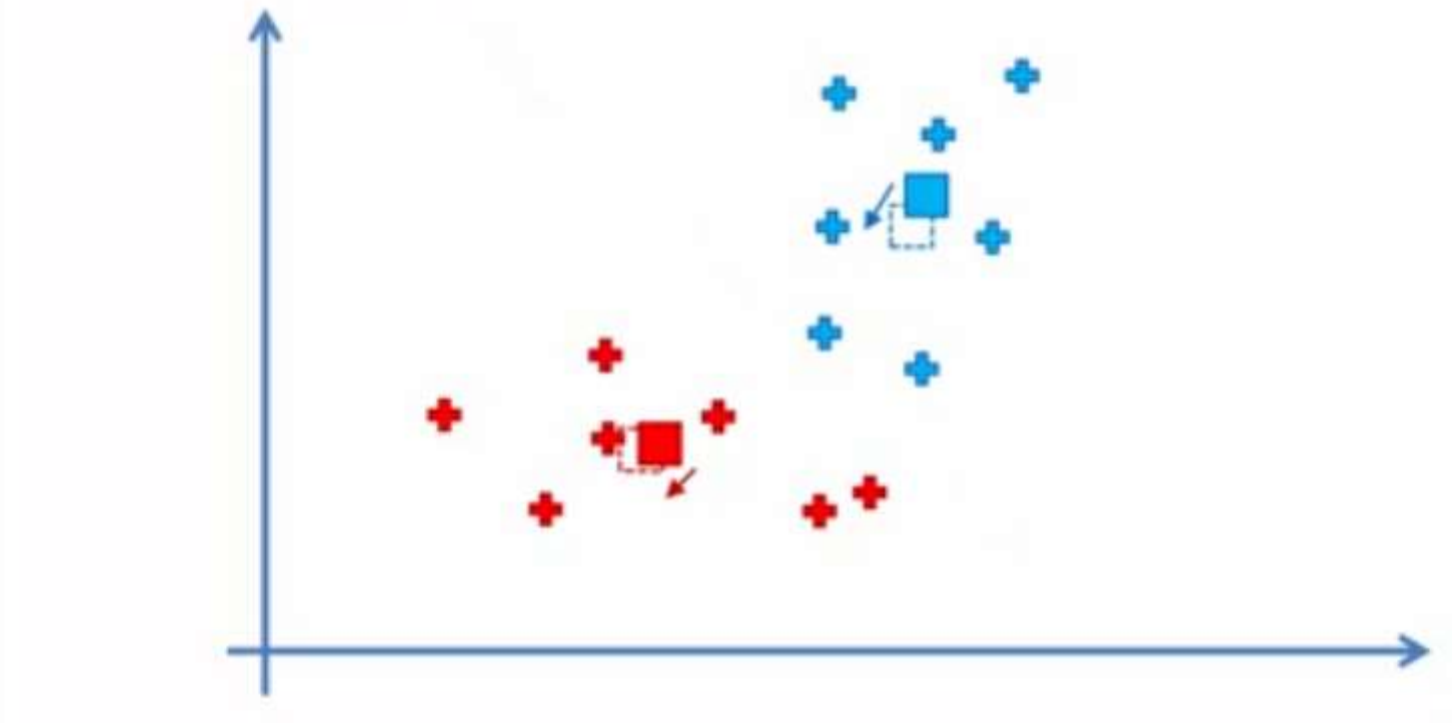
Sensitivity: Internal & Restricted

STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



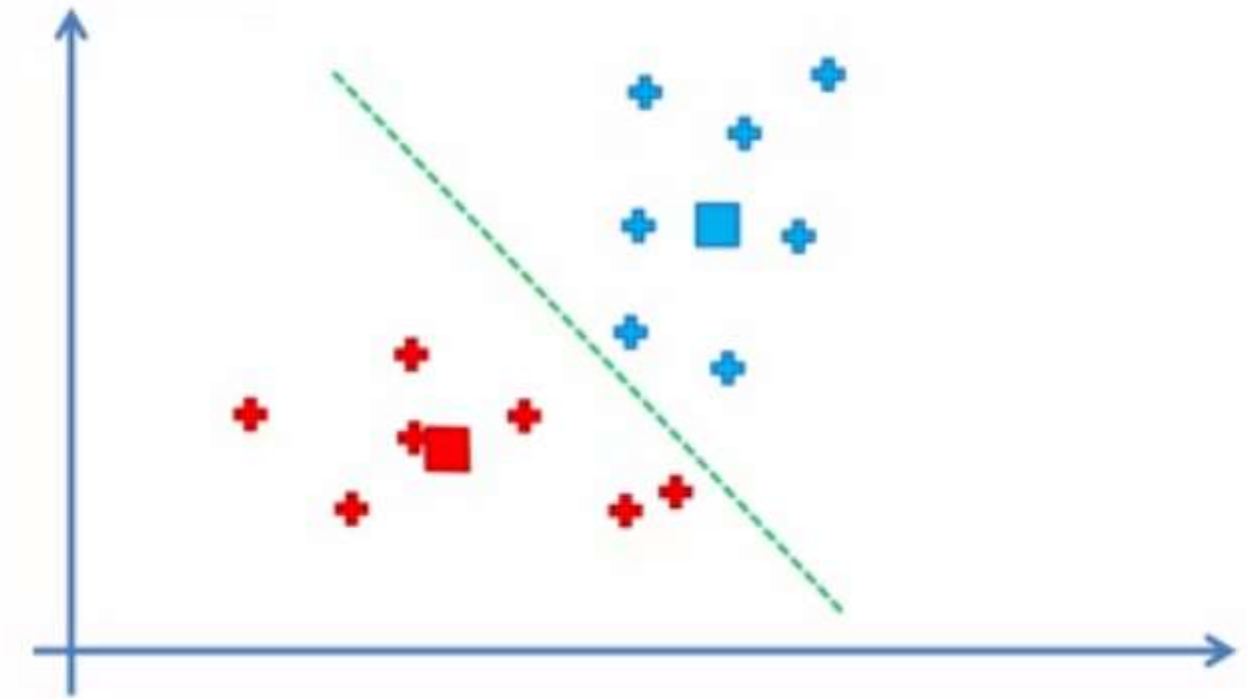
Sensitivity: Internal & Restricted

STEP 4: Compute and place the new centroid of each cluster



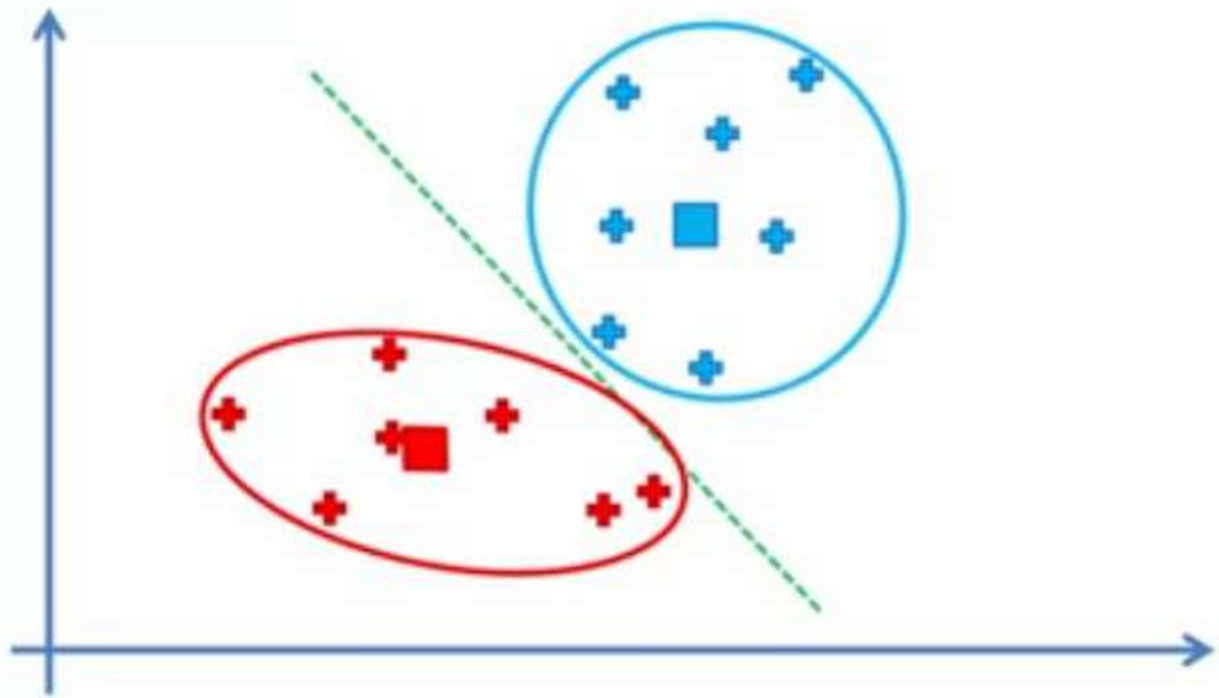
Sensitivity: Internal & Restricted

STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



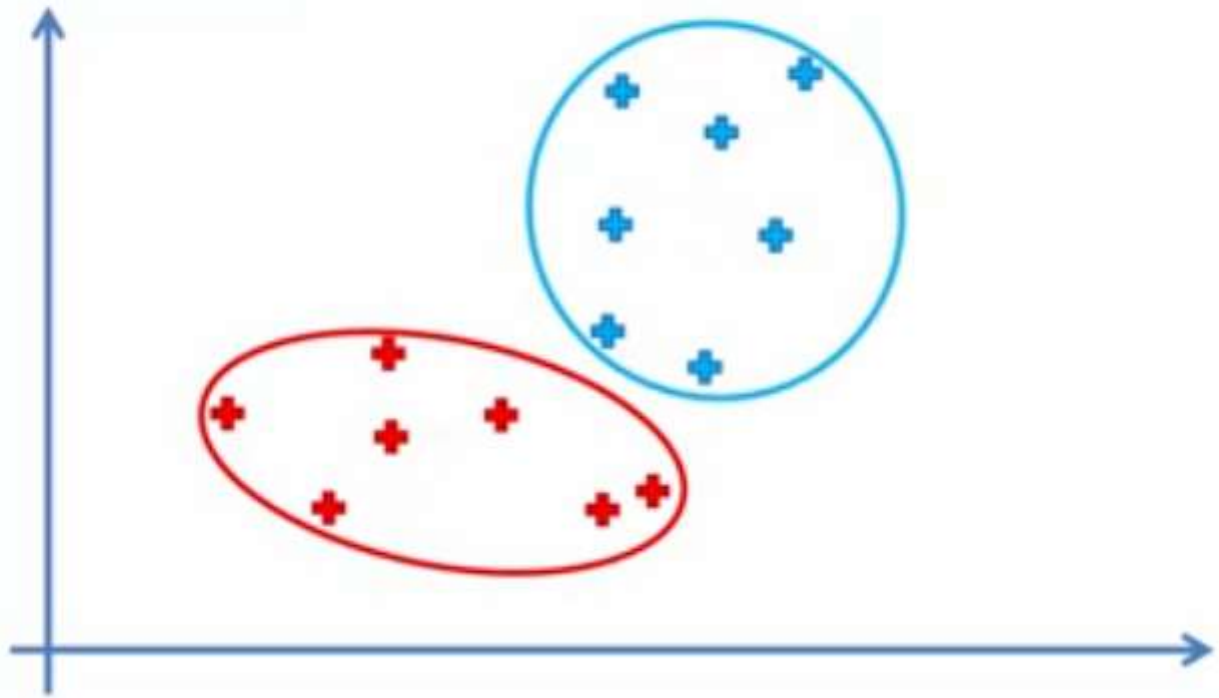
Sensitivity: Internal & Restricted

FIN: Your Model Is Ready



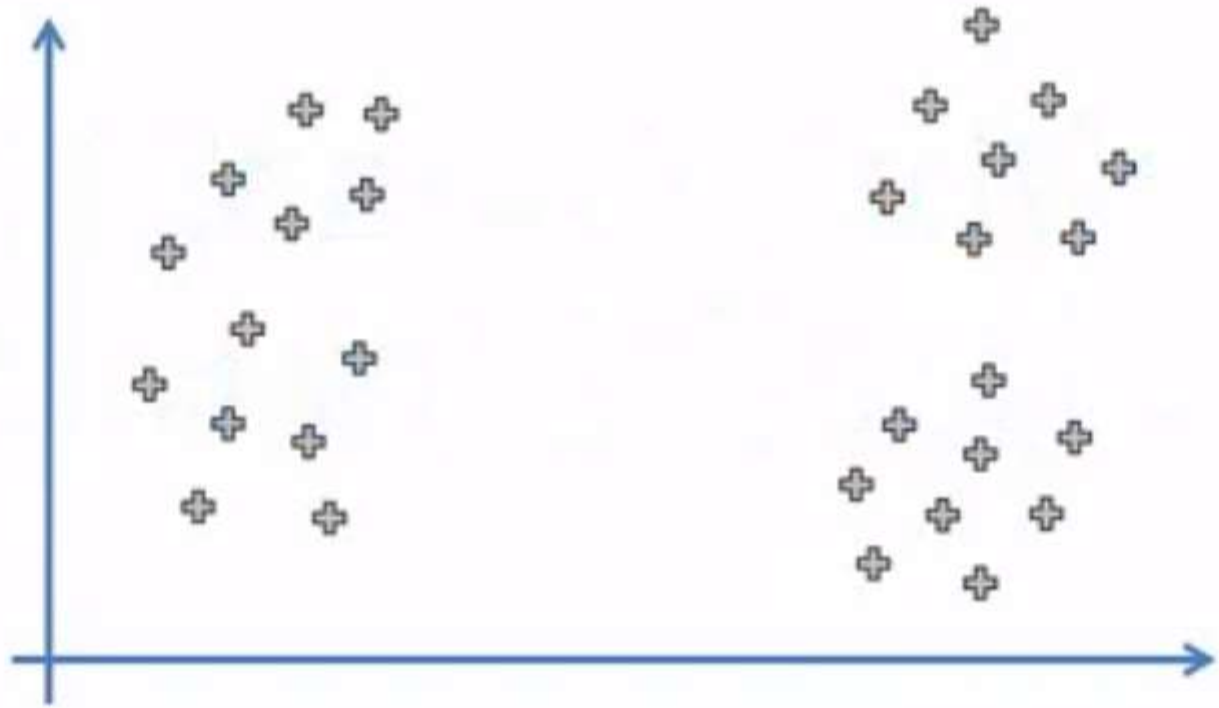
Sensitivity: Internal & Restricted

FIN: Your Model Is Ready



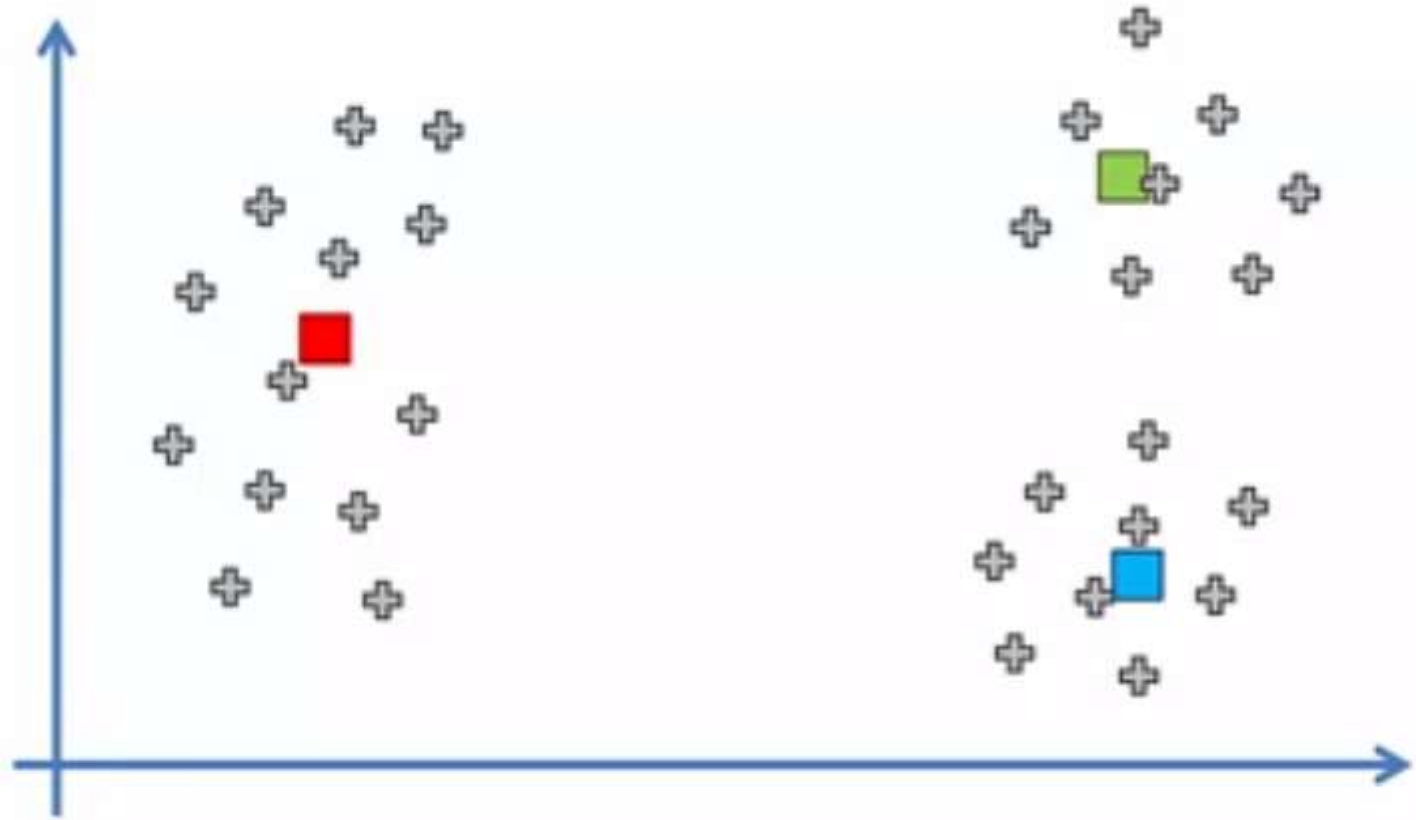
Sensitivity: Internal & Restricted

Random
Initialization Trap



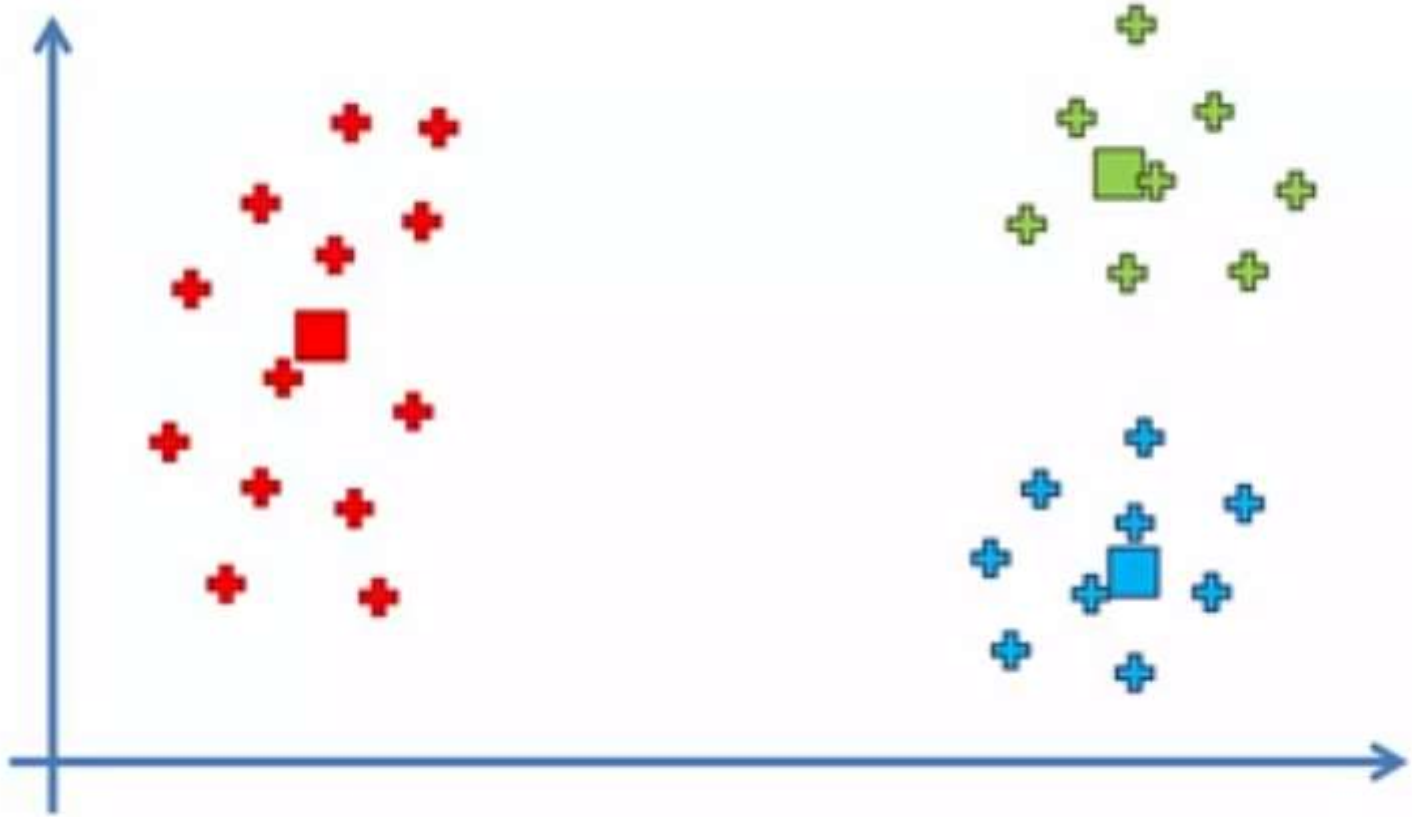
If we choose K = 3 clusters...

Sensitivity: Internal & Restricted



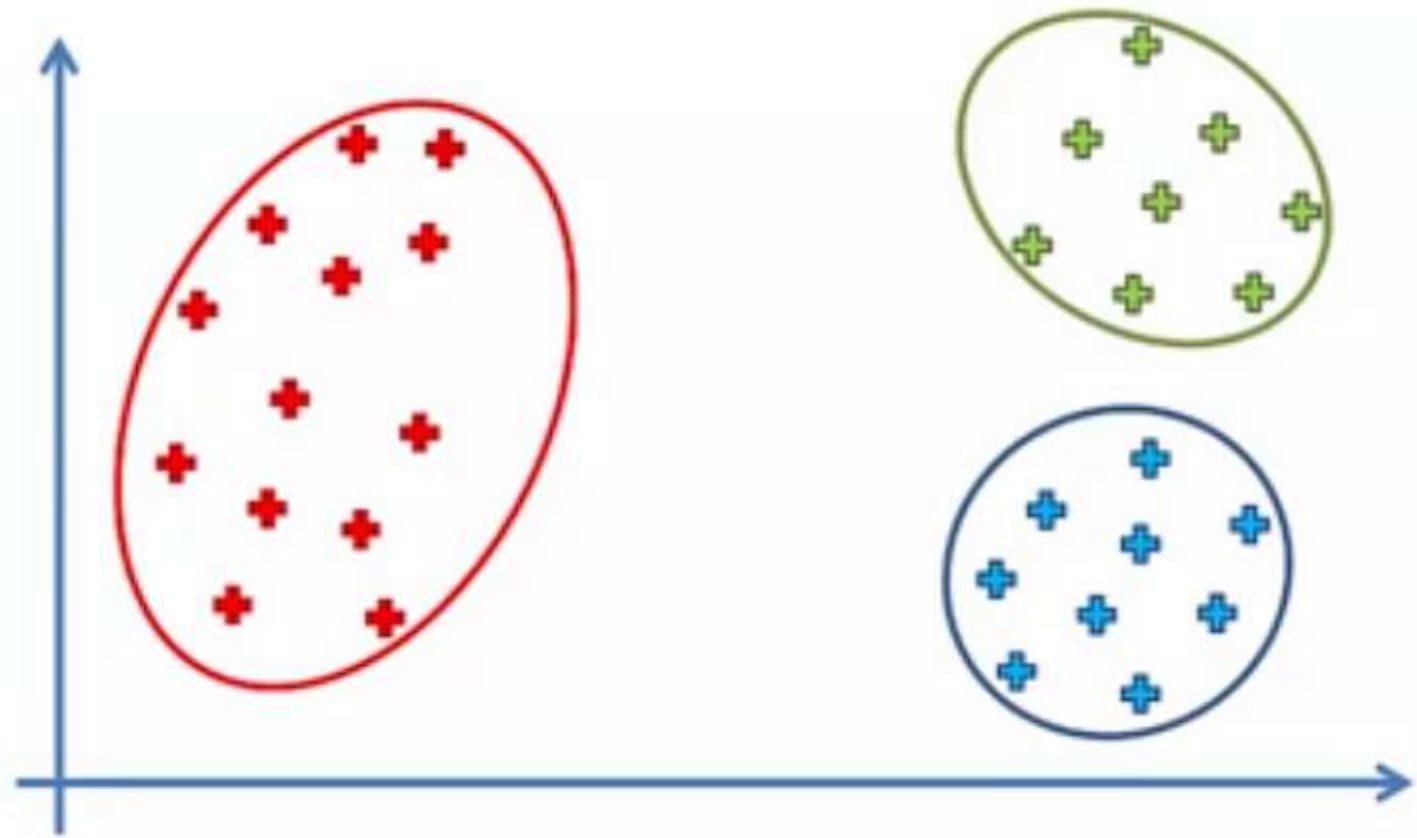
...this correct random initialisation would lead us to...

Sensitivity: Internal & Restricted



...the following three clusters

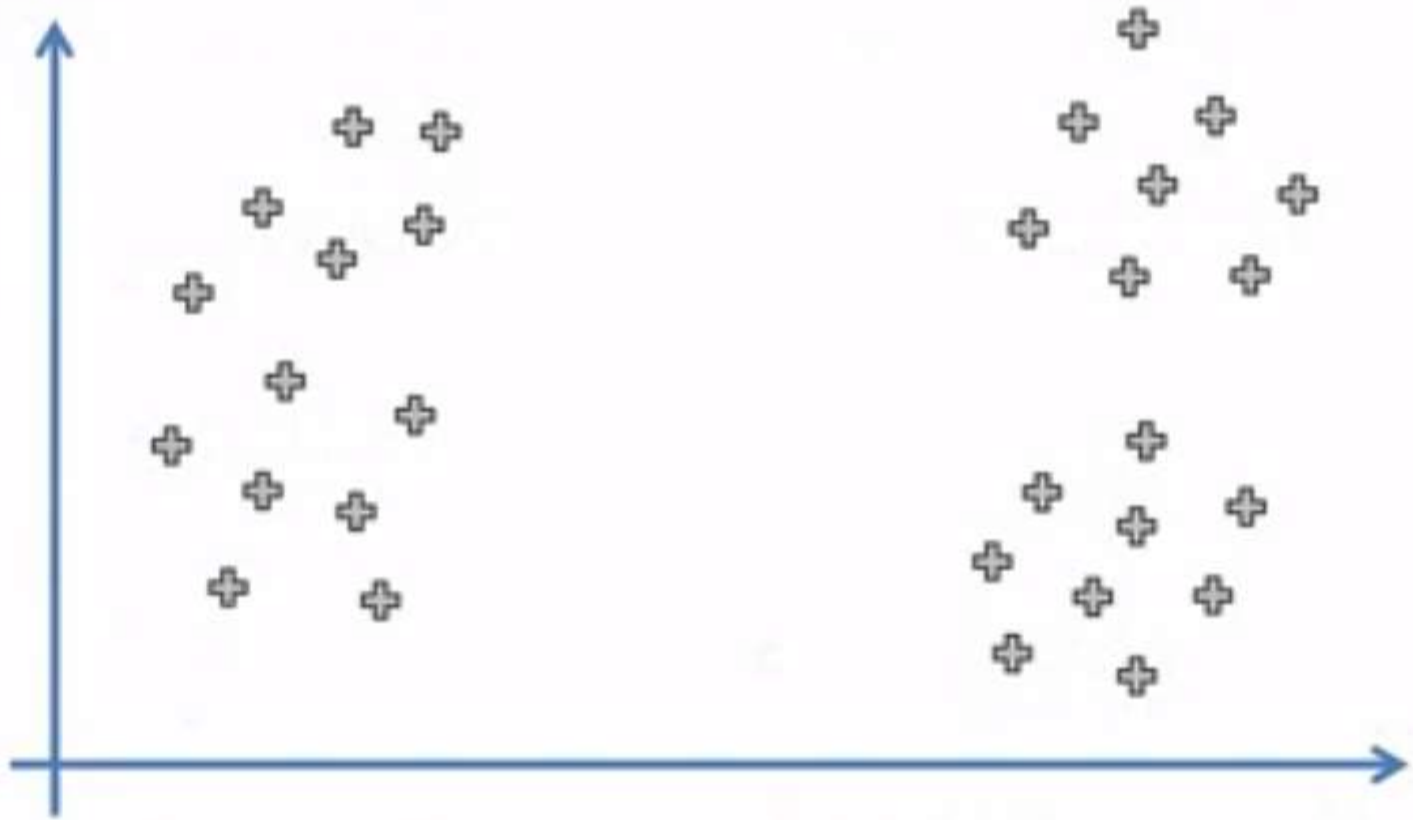
Sensitivity: Internal & Restricted



...the following three clusters

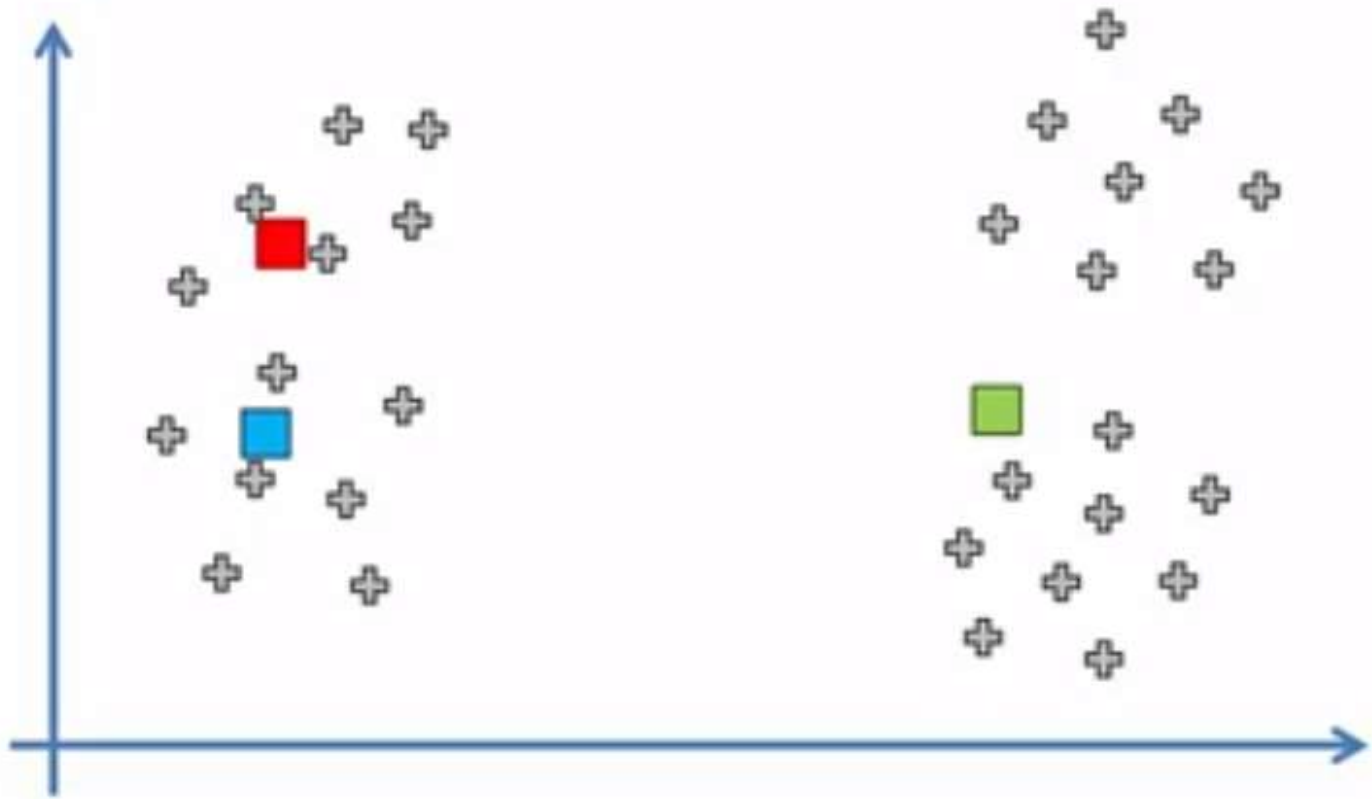
Sensitivity: Internal & Restricted

STEP 1: Choose the number K of clusters: K = 3



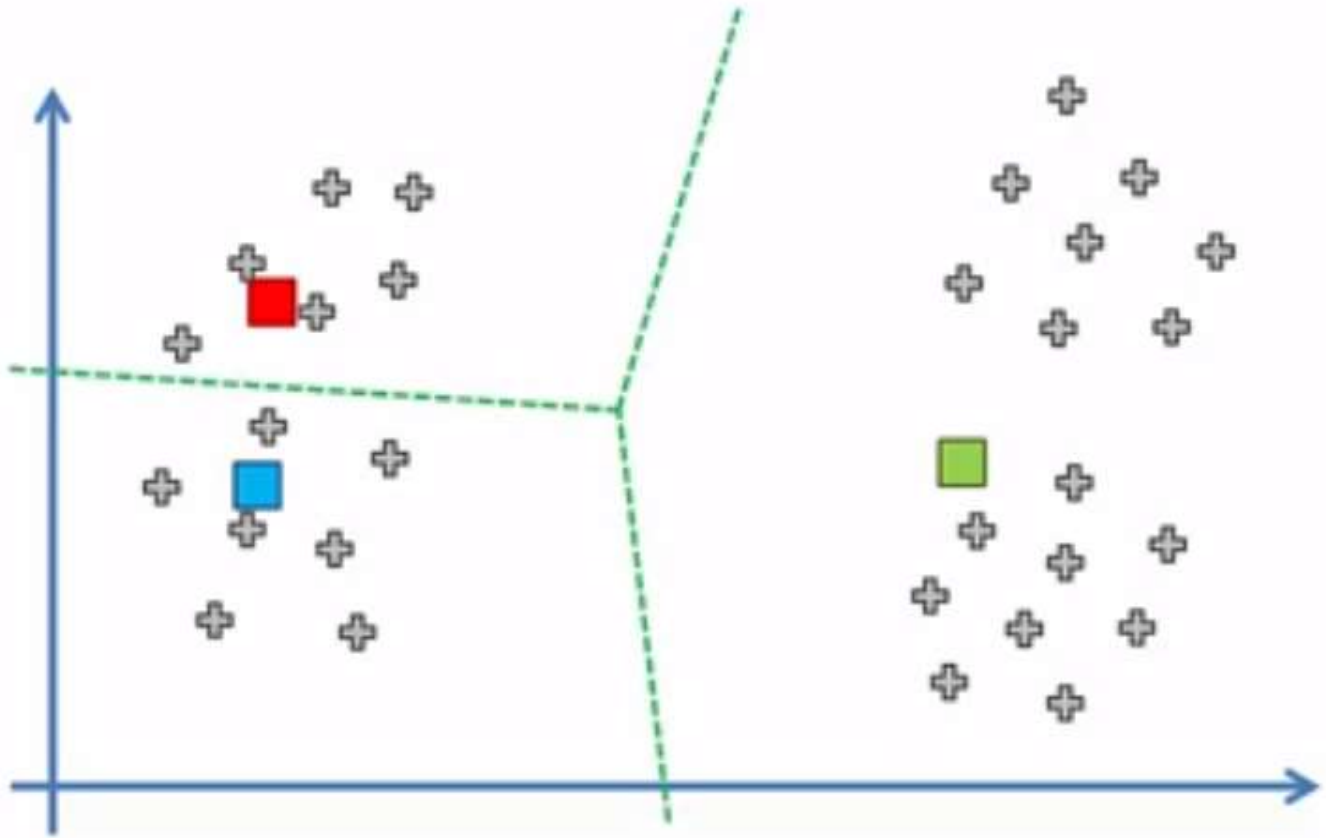
Sensitivity: Internal & Restricted

STEP 2: Select at random K points, the centroids (not necessarily from your dataset)



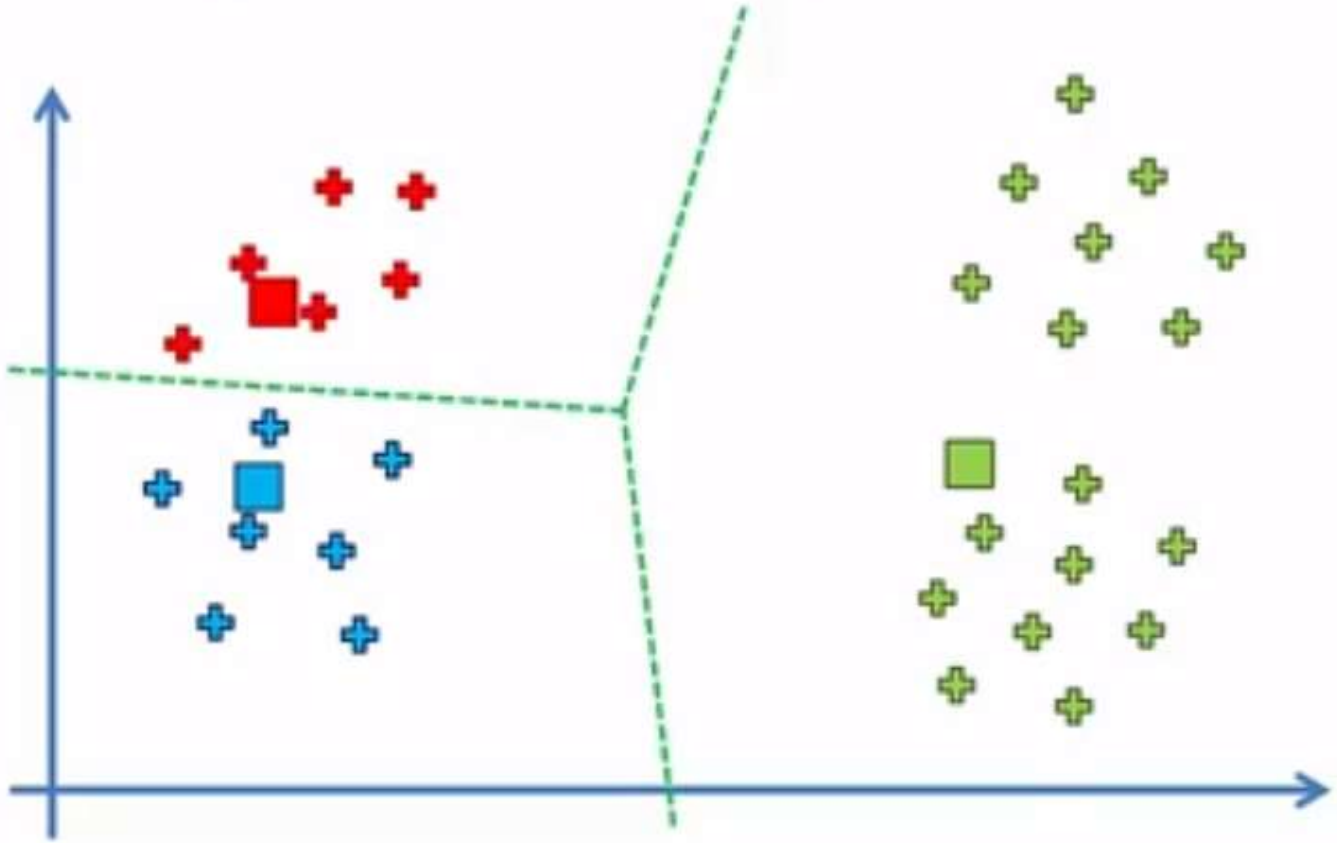
Sensitivity: Internal & Restricted

STEP 2: Select at random K points, the centroids (not necessarily from your dataset)



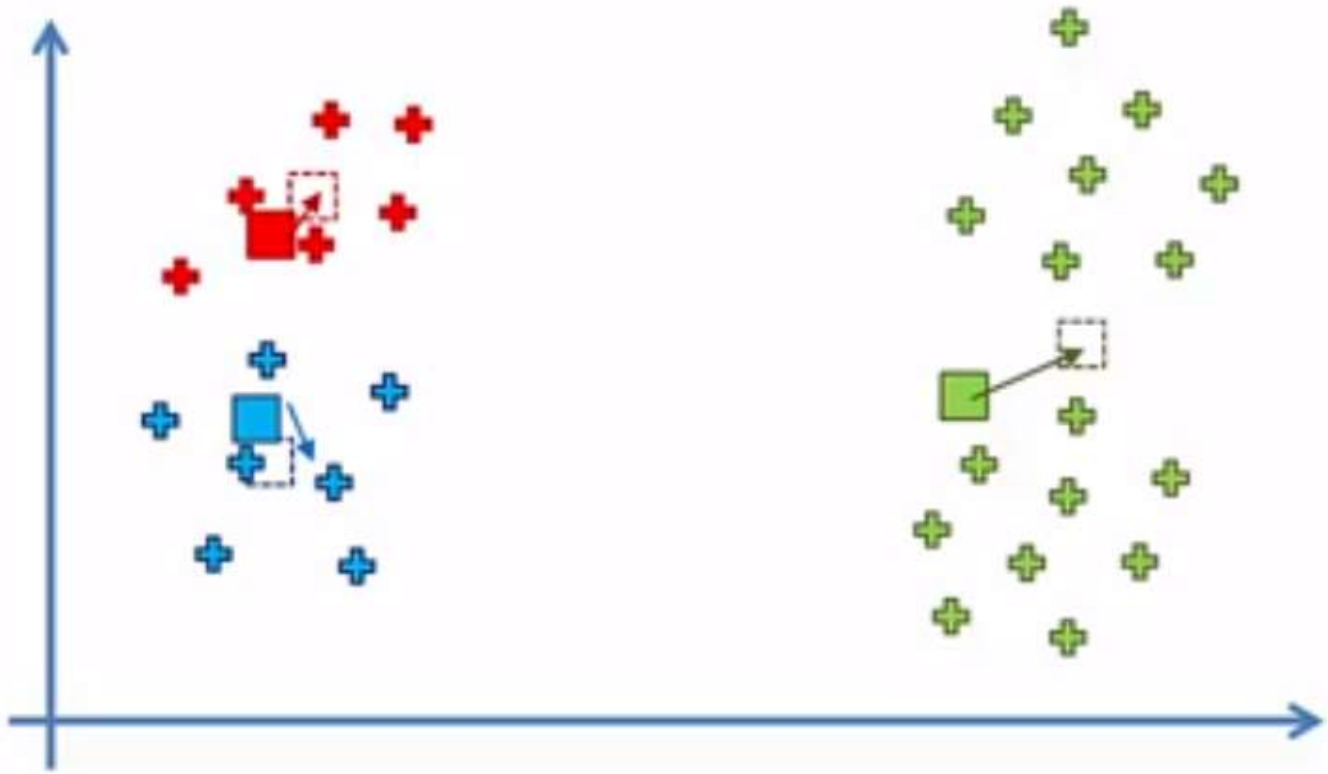
Sensitivity: Internal & Restricted

STEP 3: Assign each data point to the closest centroid ➡ That forms K clusters



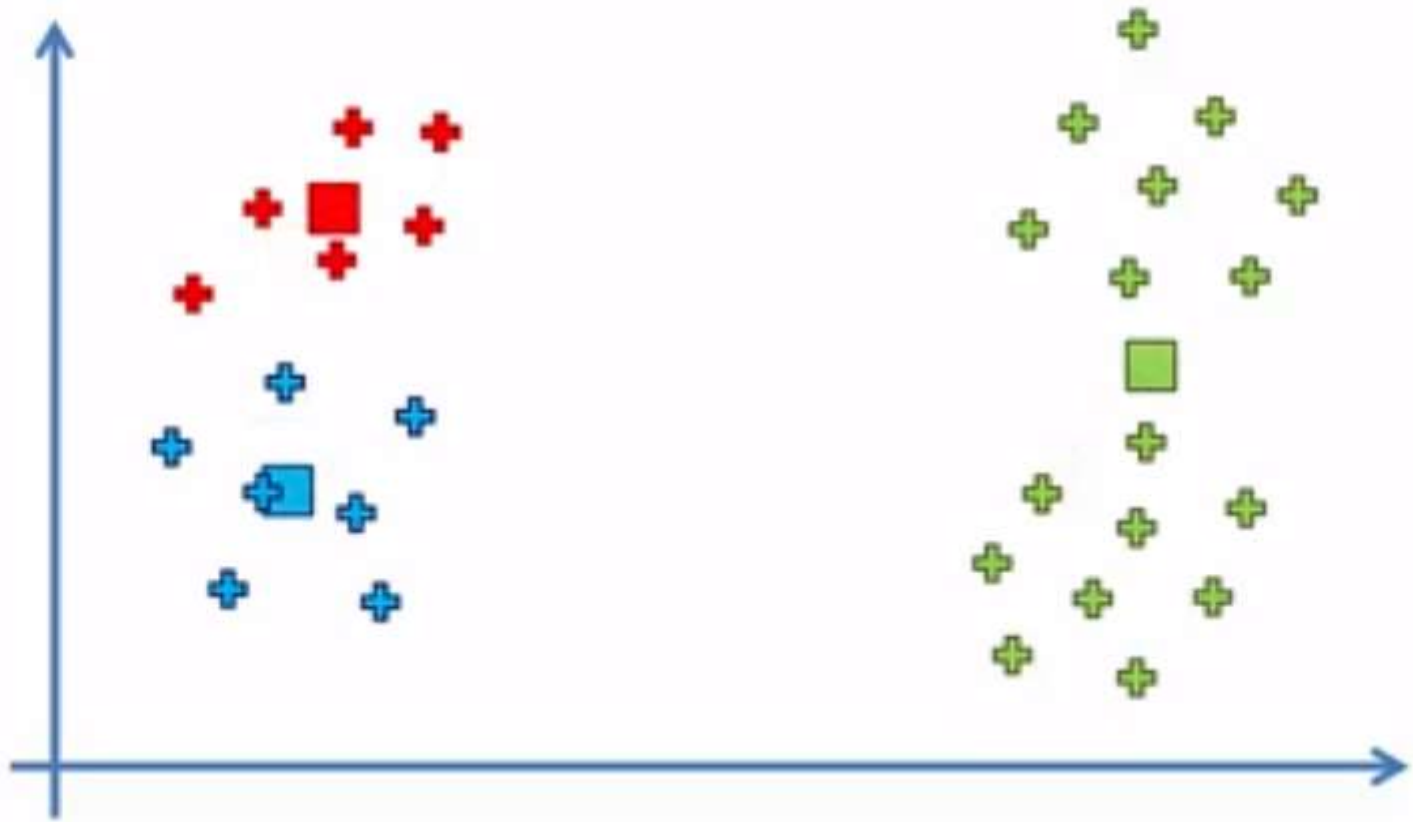
Sensitivity: Internal & Restricted

STEP 3: Assign each data point to the closest centroid ➡ That forms K clusters



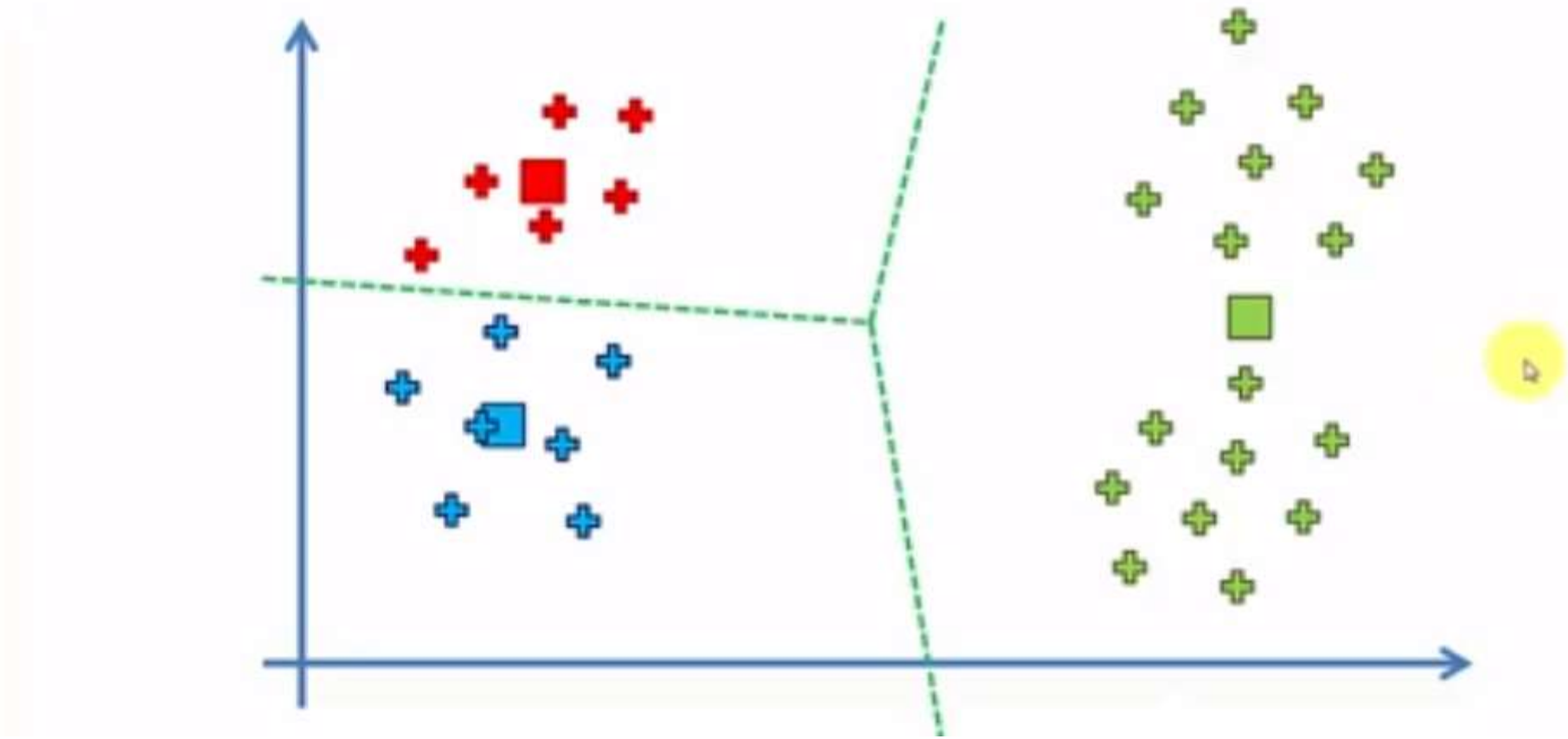
Sensitivity: Internal & Restricted

STEP 4: Compute and place the new centroid of each cluster



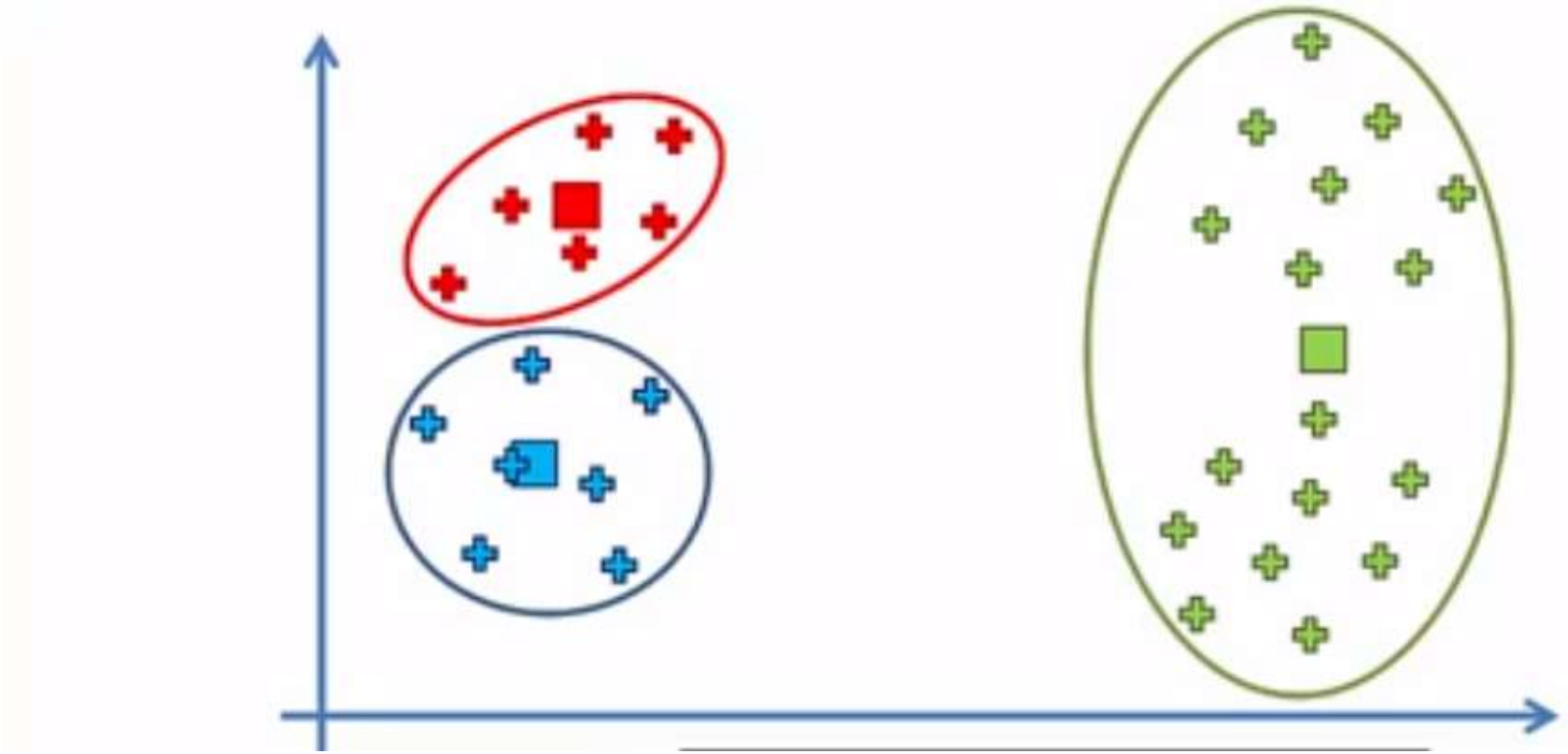
Sensitivity: Internal & Restricted

STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



Sensitivity: Internal & Restricted

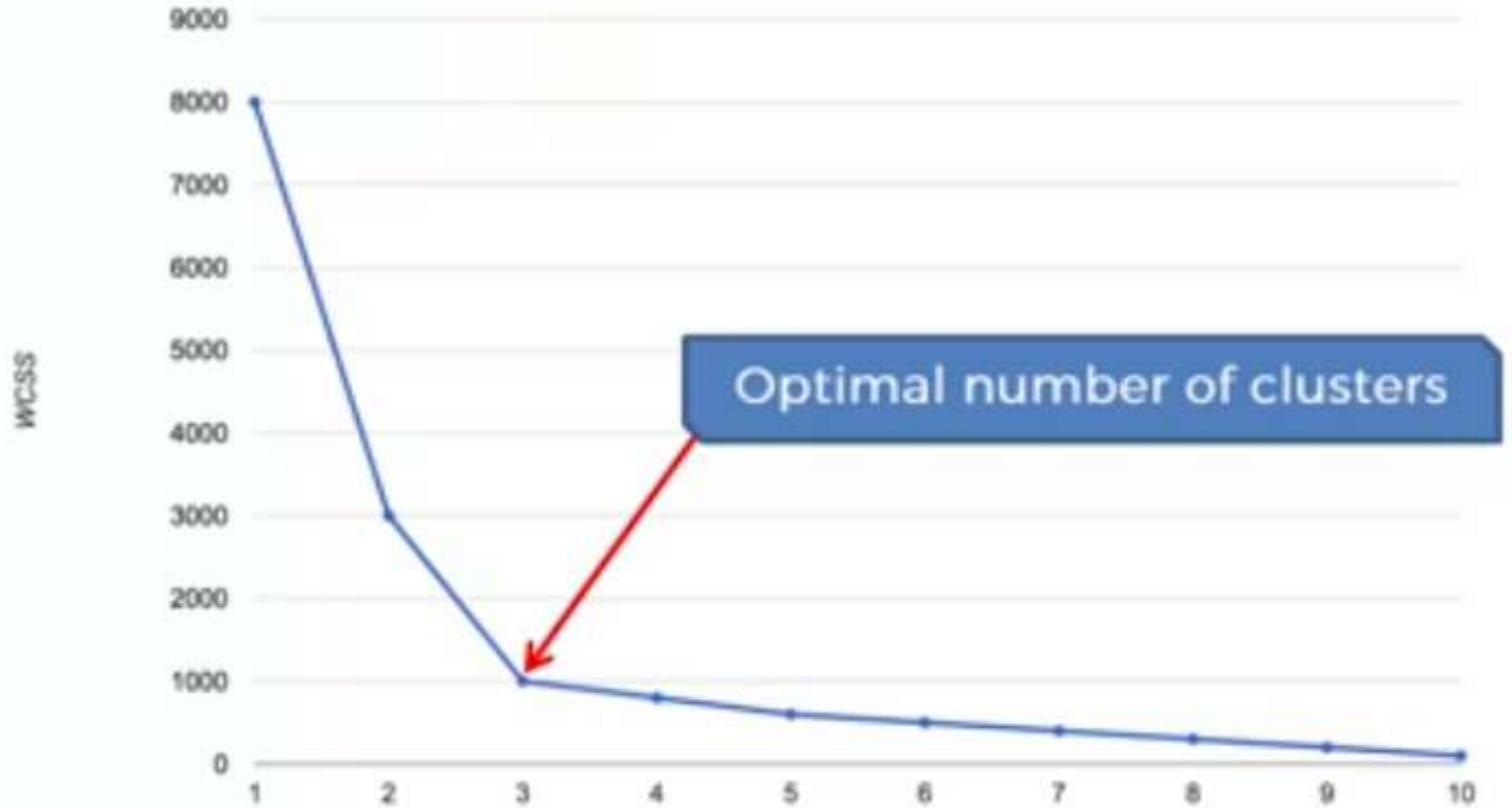
STEP 5: Reassign each data point to the new closest centroid.
If any reassignment took place, go to STEP 4, otherwise go to FIN.



Sensitivity: Internal & Restricted

The elbow method

The Elbow Method



Sensitivity: Internal & Restricted