

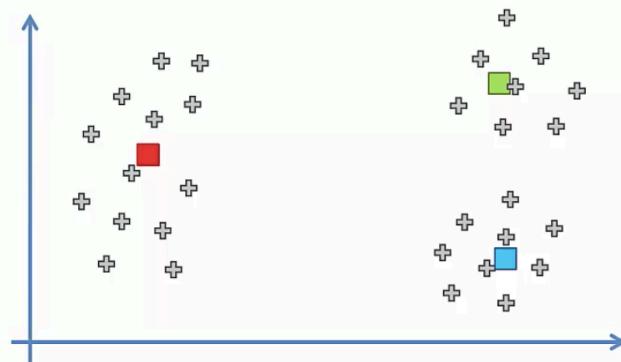
# K-Means Intuition: Random Initialization Trap

## Random Initialization Trap



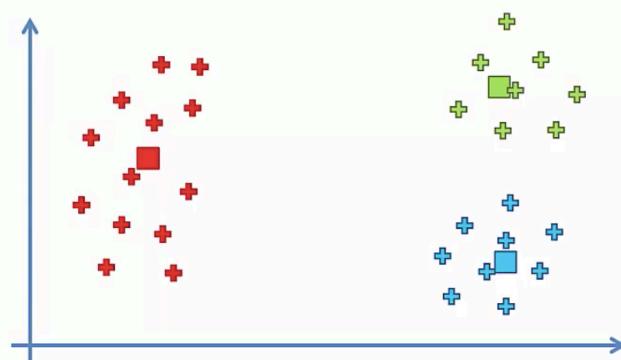
If we choose  $K = 3$  clusters...

# Random Initialization Trap



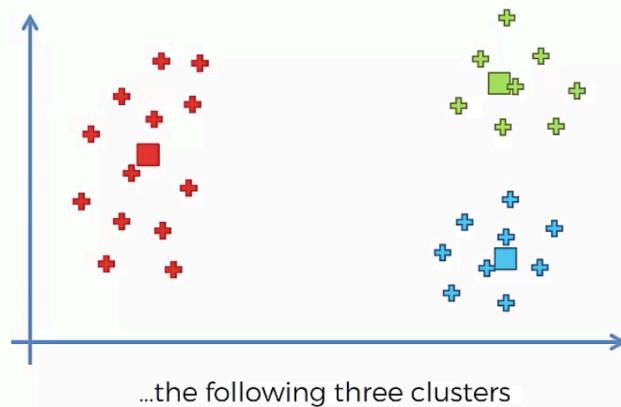
...this correct random initialisation would lead us to...

# Random Initialization Trap



...the following three clusters

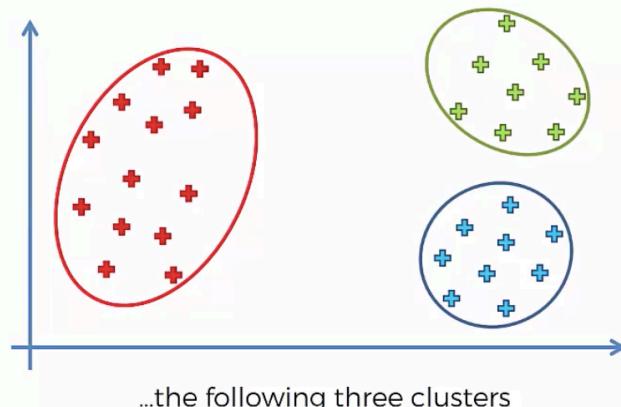
# Random Initialization Trap



Machine Learning A-Z

© SuperDataScience

# Random Initialization Trap



Machine Learning A-Z

© SuperDataScience

# Random Initialization Trap

But what would happen if we had a bad random initialisation ?

# Random Initialization Trap

STEP 1: Choose the number K of clusters



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



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



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



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



Your Model is Ready

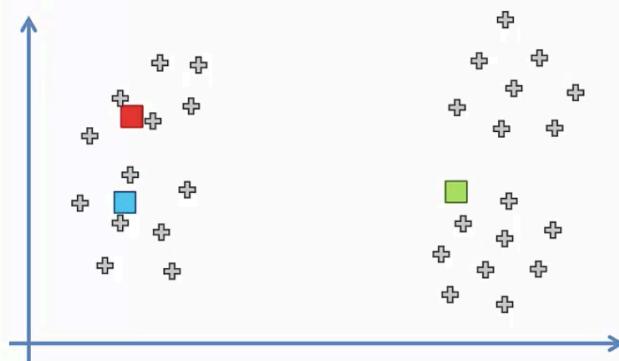
# Random Initialization Trap

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



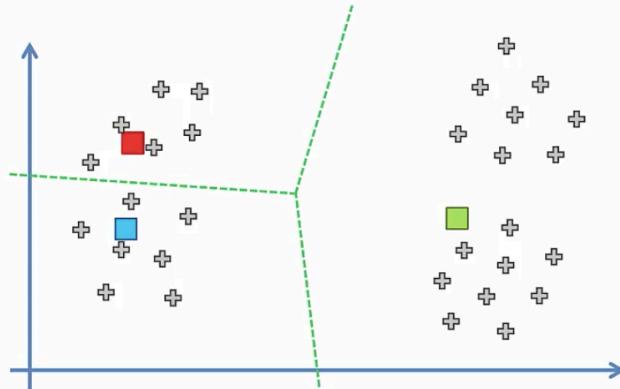
# Random Initialization Trap

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



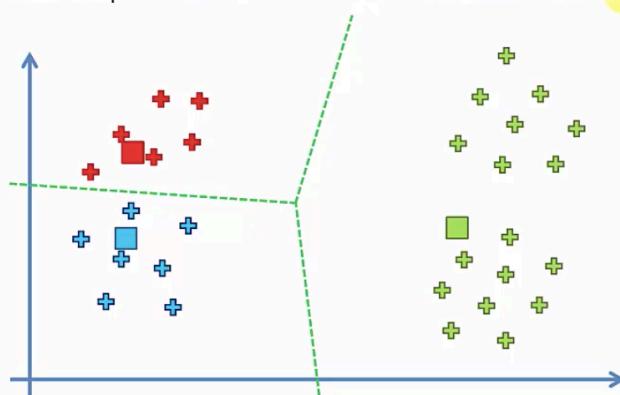
# Random Initialization Trap

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



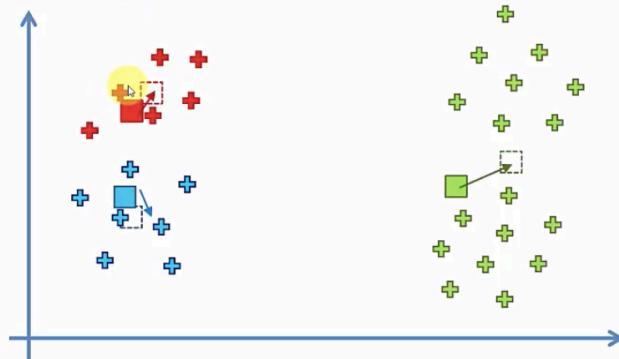
# Random Initialization Trap

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



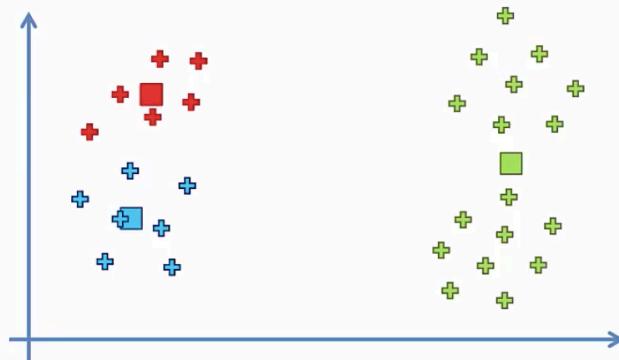
# Random Initialization Trap

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



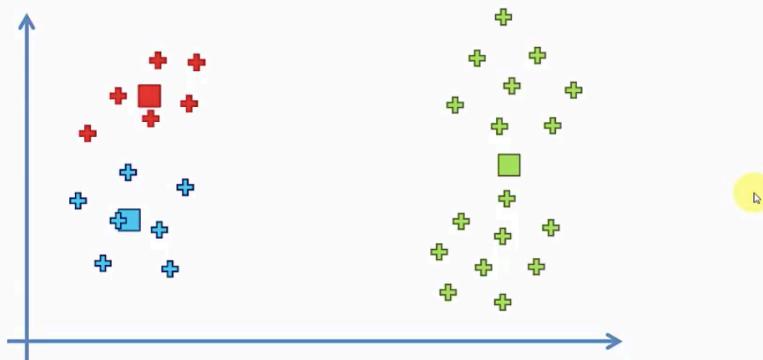
# Random Initialization Trap

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



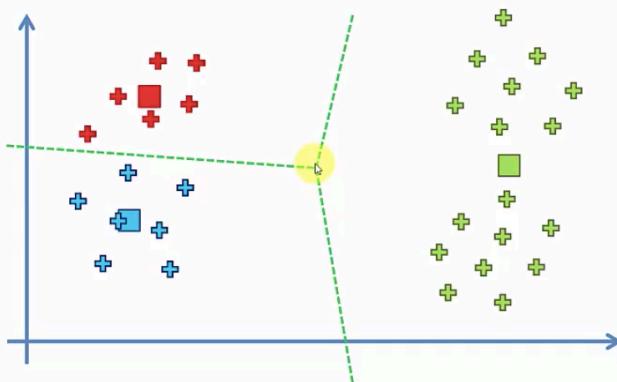
# Random Initialization Trap

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



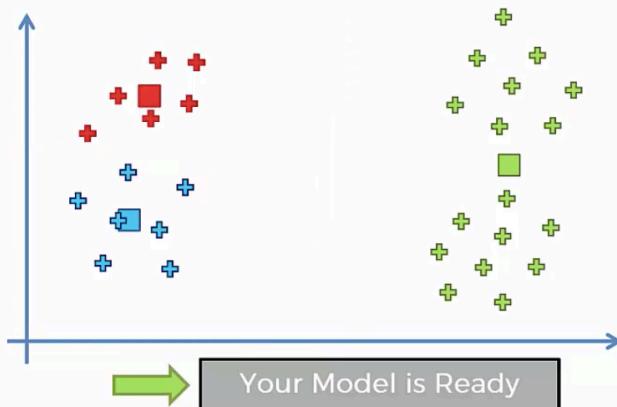
# Random Initialization Trap

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



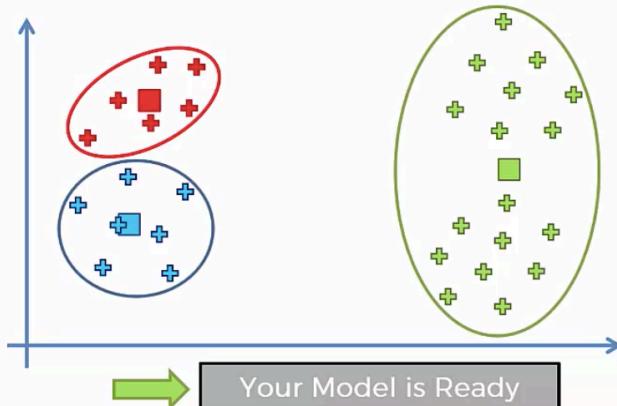
# Random Initialization Trap

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



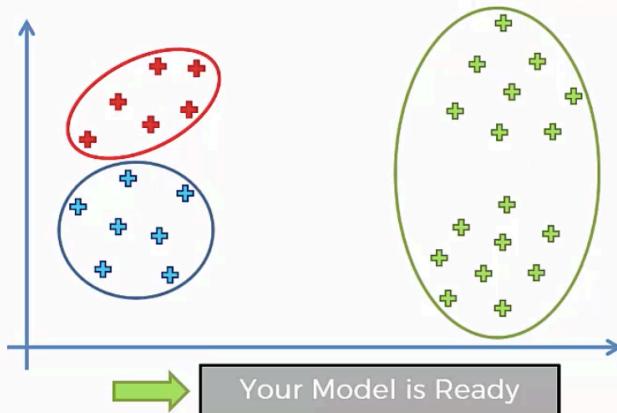
# Random Initialization Trap

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

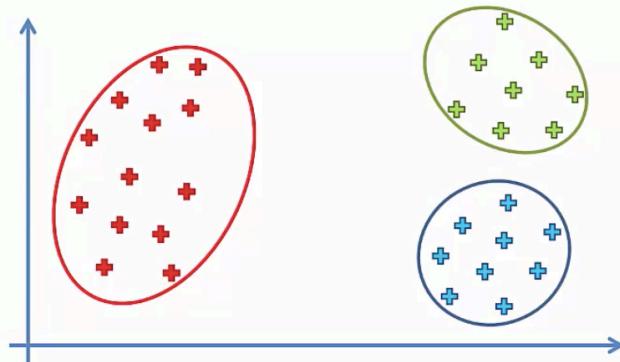


# Random Initialization Trap

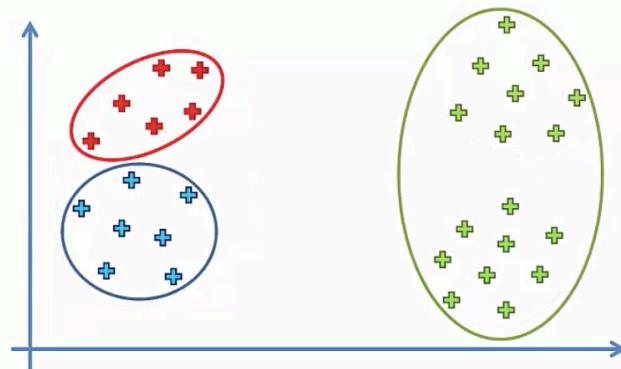
STEP 5: Reassign each data point to the new closest centroid.  
If any reassignment took place, go to STEP 4, otherwise go to FIN.



# Random Initialization Trap



# Random Initialization Trap



Machine Learning A-Z

© SuperDataScience

# Random Initialization Trap

Solution



K-Means++

Machine Learning A-Z

© SuperDataScience