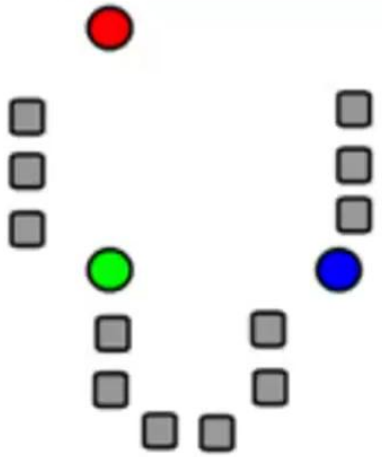
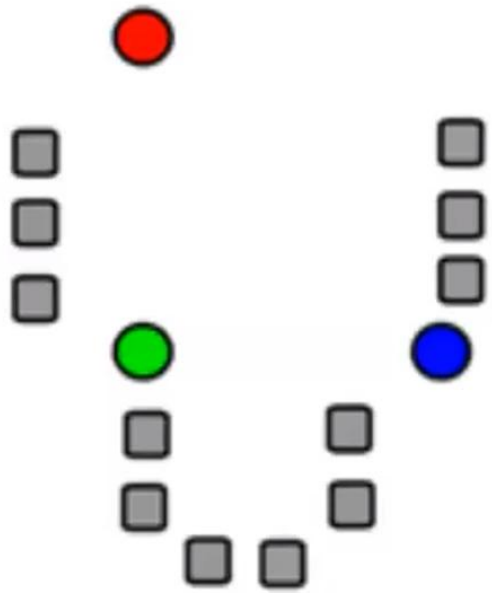
How K-means Clustering Works?

K-means Clustering takes an iterative approach to perform the clustering task. The working steps of this algorithm are as follows-

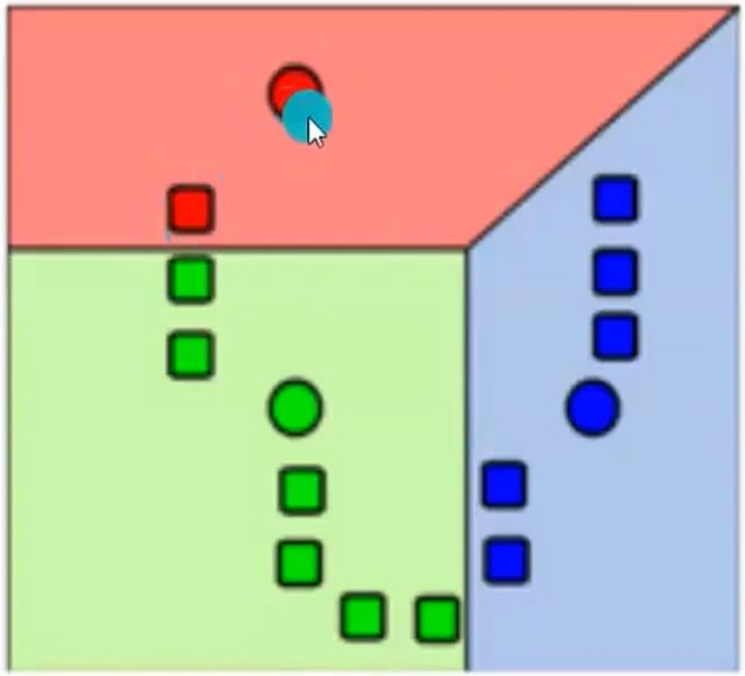
Step 1: Choose the number K (in this case k =3) of clusters.



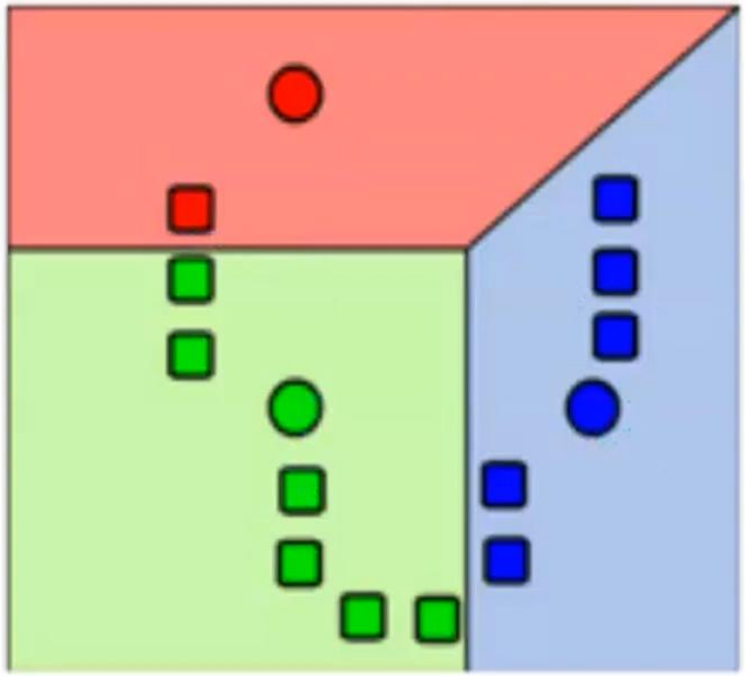
Step 2: Select at random K points, the centroids(not necessarily from our dataset).

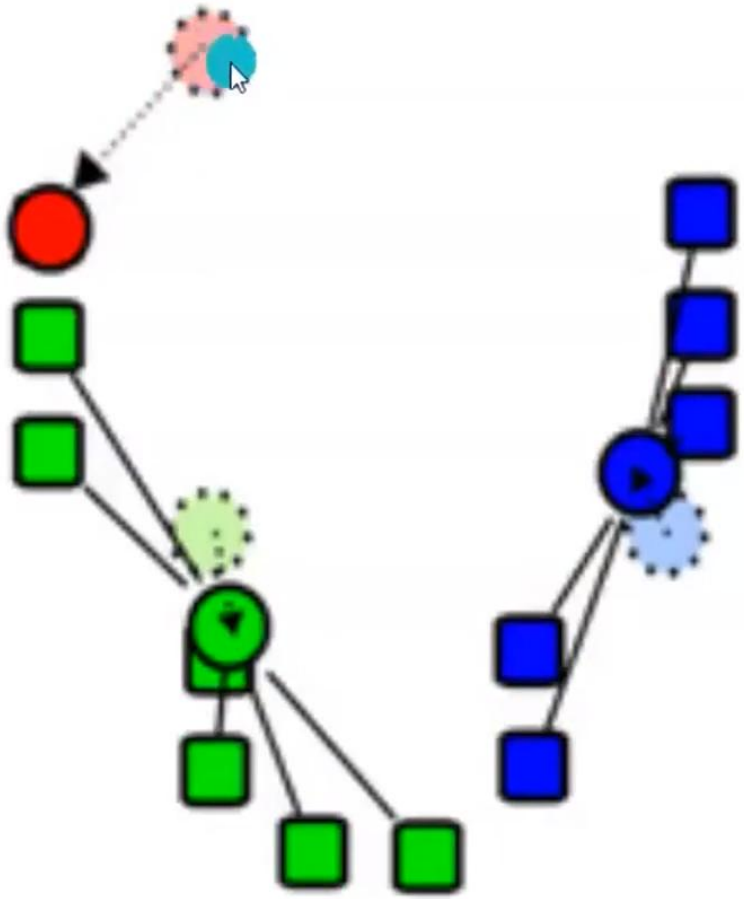


Step 3: Assign each data point to the closest centroid based on euclidian or manhattan distance. That forms K clusters.

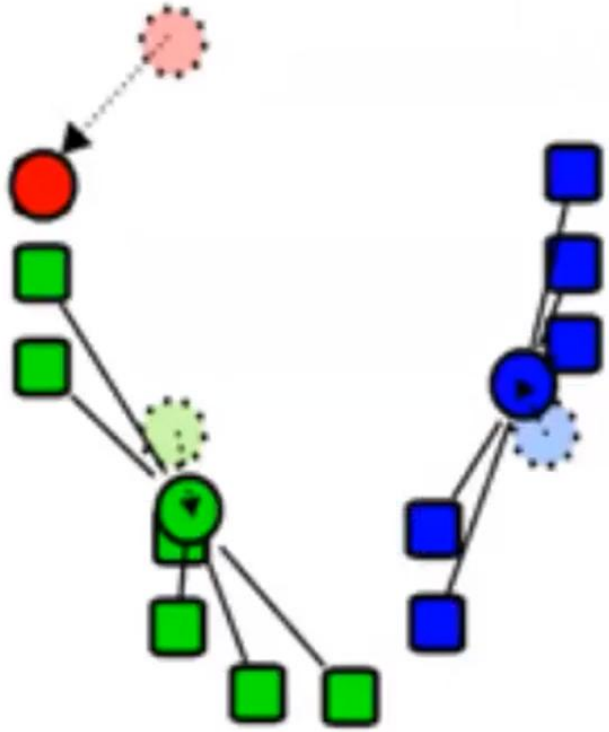


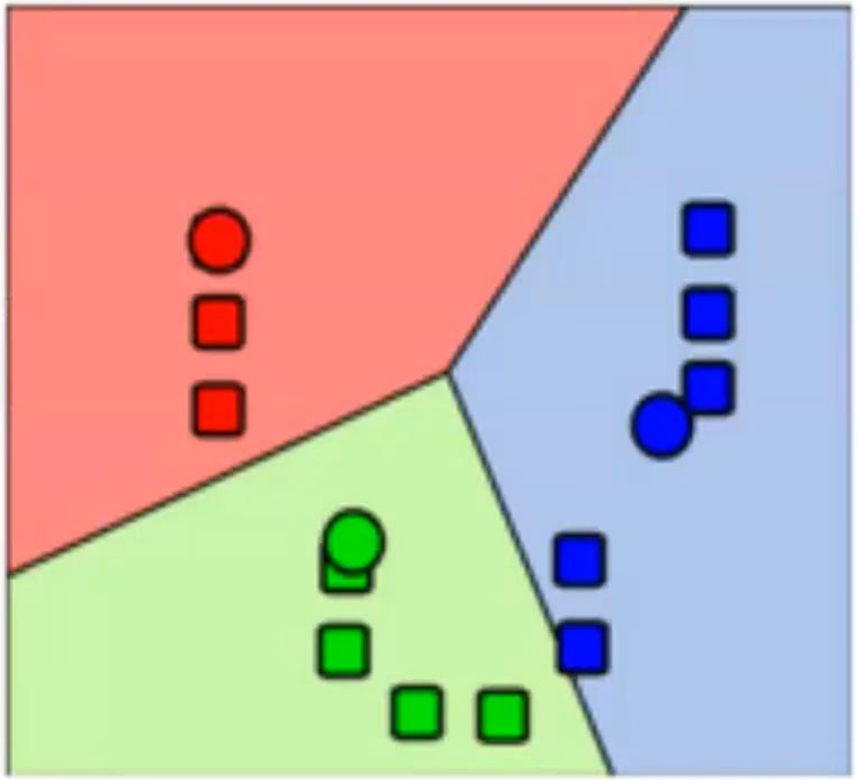
Step 4: Compute and place the new centroid of each cluster. The centroid of each of the k clusters becomes the new mean.





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





@nerchu_ko

Nerchuko

Choosing The Optimal Number of Clusters

The value of k is very crucial for optimal outcomes from the algorithm. There are several techniques to choose the optimal value for k including-

- Cross-Validation,
- Silhouette Method
- G-means Algorithm
- Elbow Method

Here we will implement the elbow method to find the optimal value for k.