**Clustering**

Clustering is similar to classification, but the basis is different. In Clustering you don’t know what you are looking for, and you are trying to identify some segments or clusters in your data. When you use clustering algorithms on your dataset, unexpected things can suddenly pop up like structures, clusters and groupings you would have never thought of otherwise.

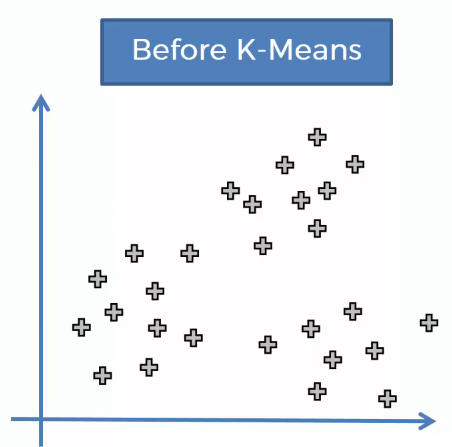
In this part, you will understand and learn how to implement the following Machine Learning Clustering models:

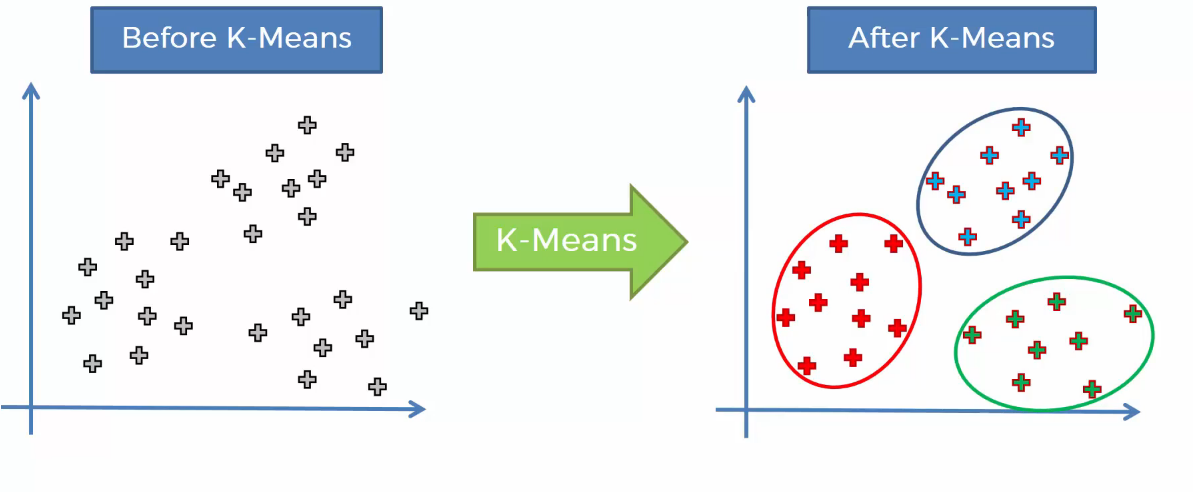
1. K-Means Clustering
2. Hierarchical Clustering

Enjoy Machine Learning!

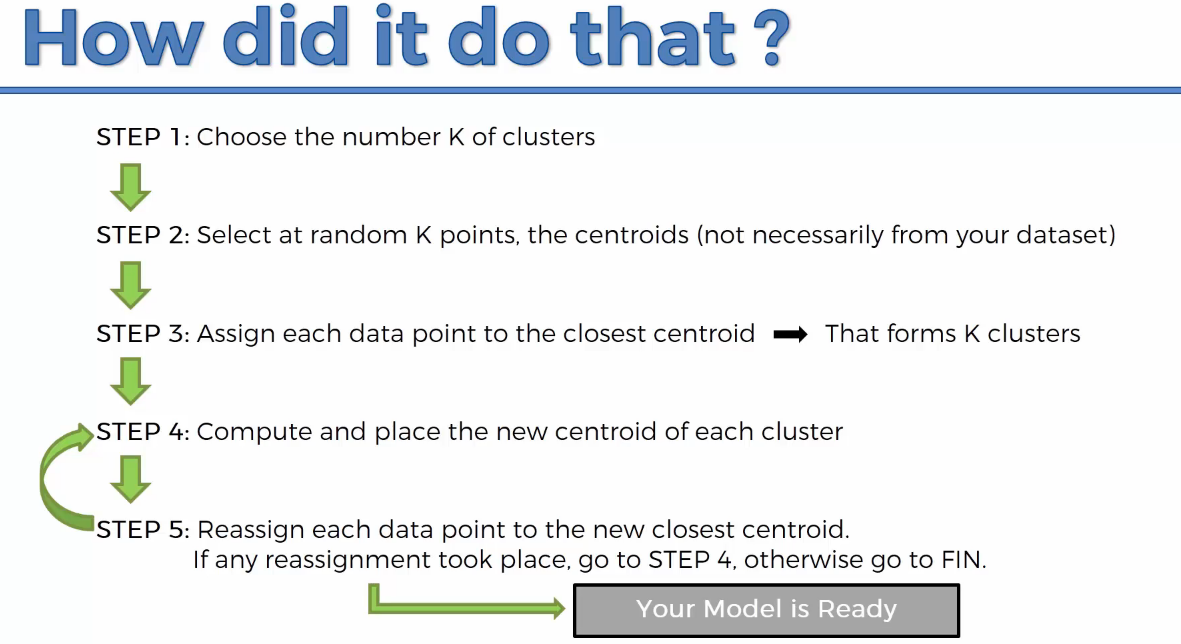
**K-Means:**

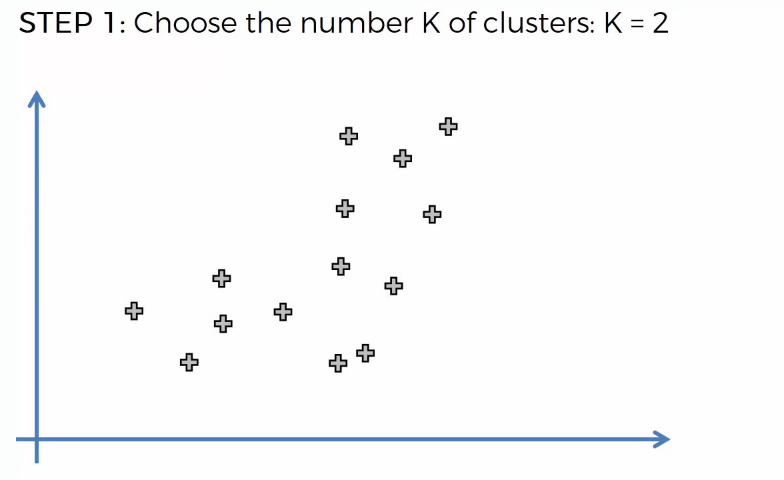
How to identify the groups

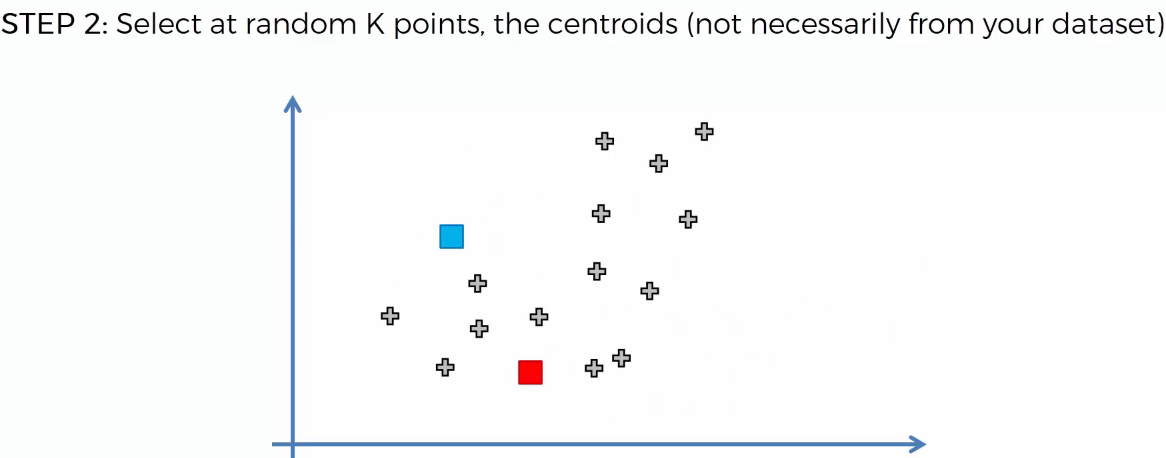


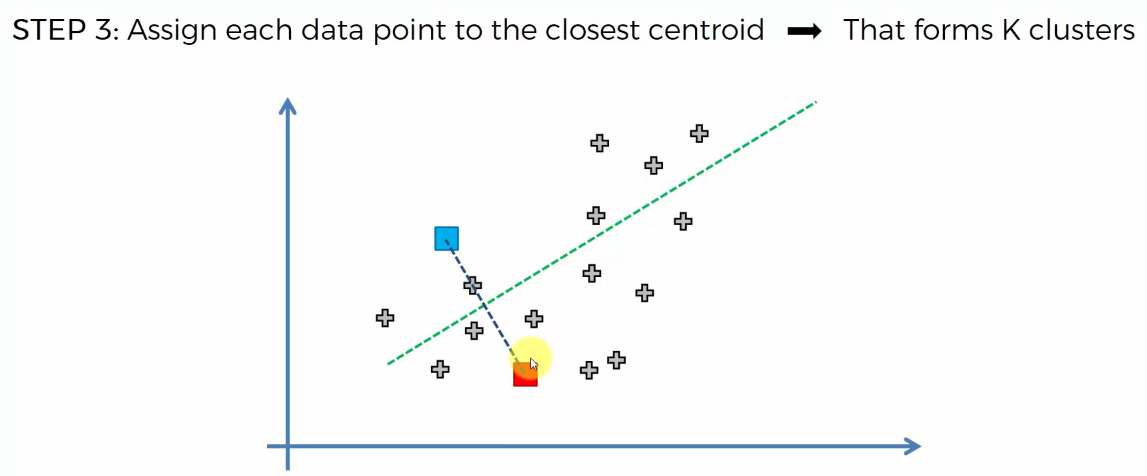


It can work with multi-dimensional data

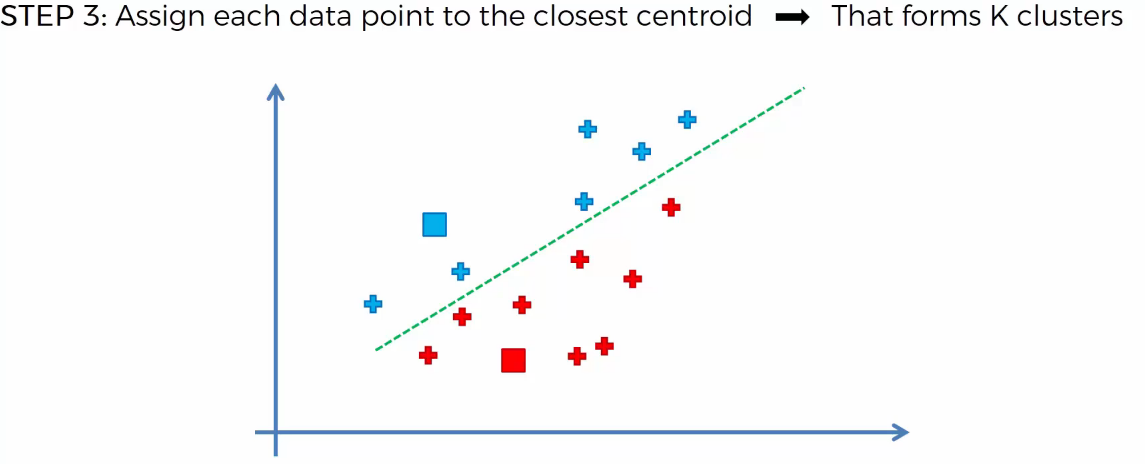




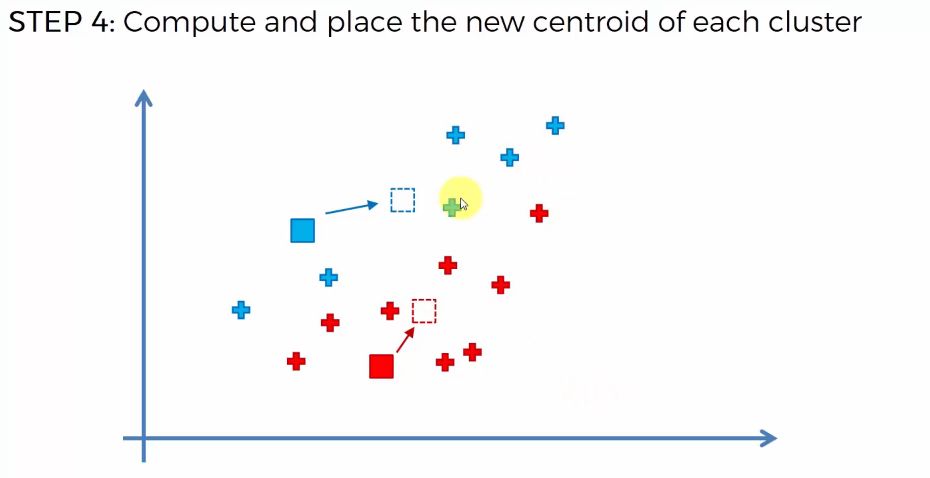


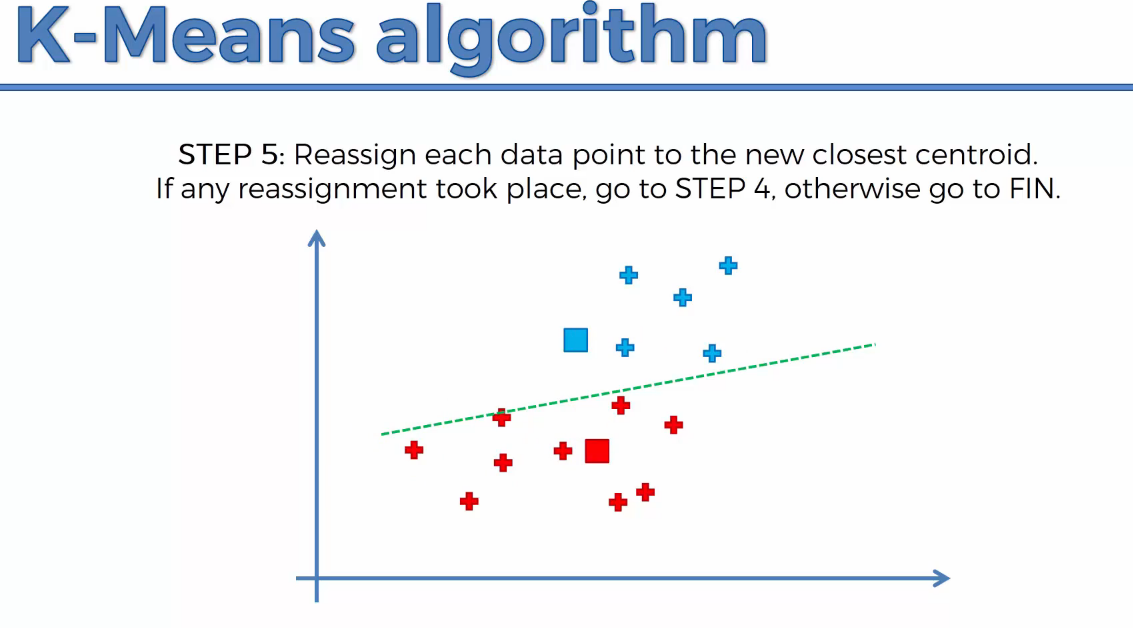


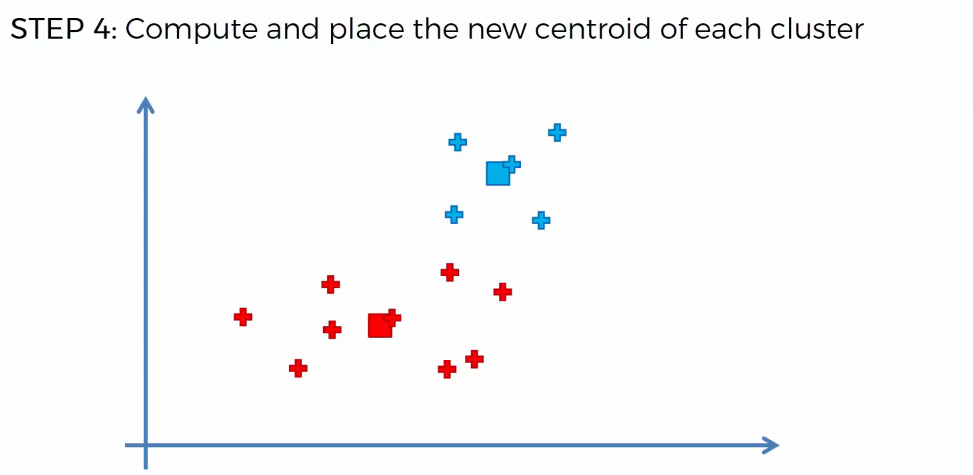
Use perpendicular to make it simple.

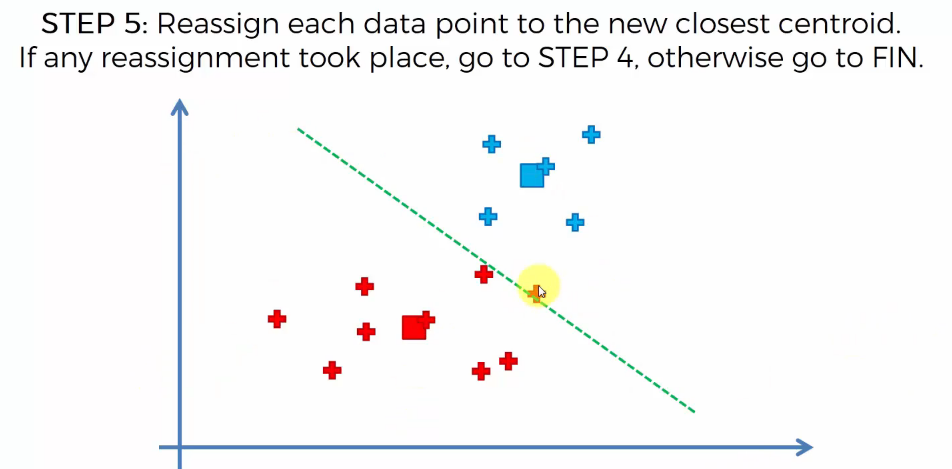


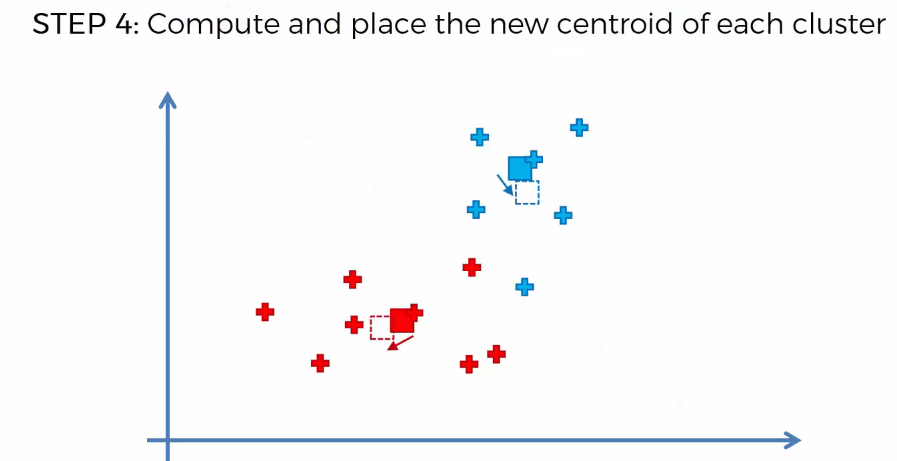
Find the center of gravity based upon the points.

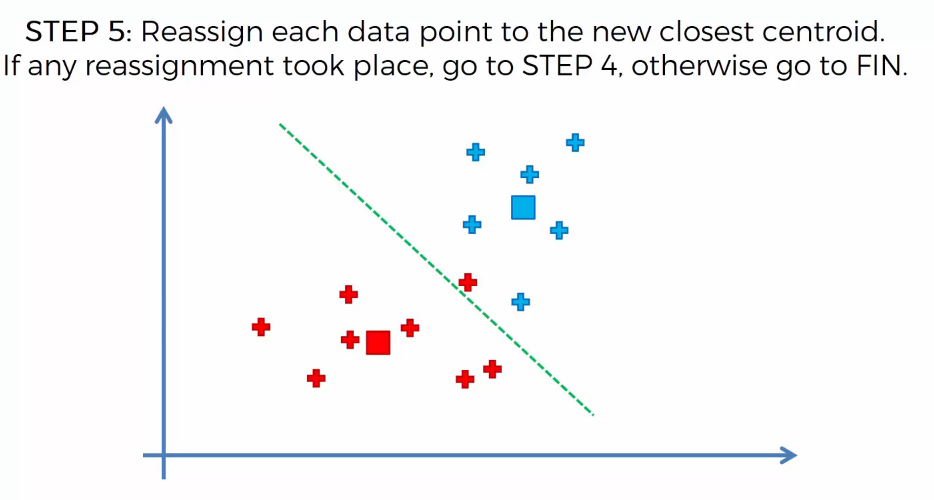


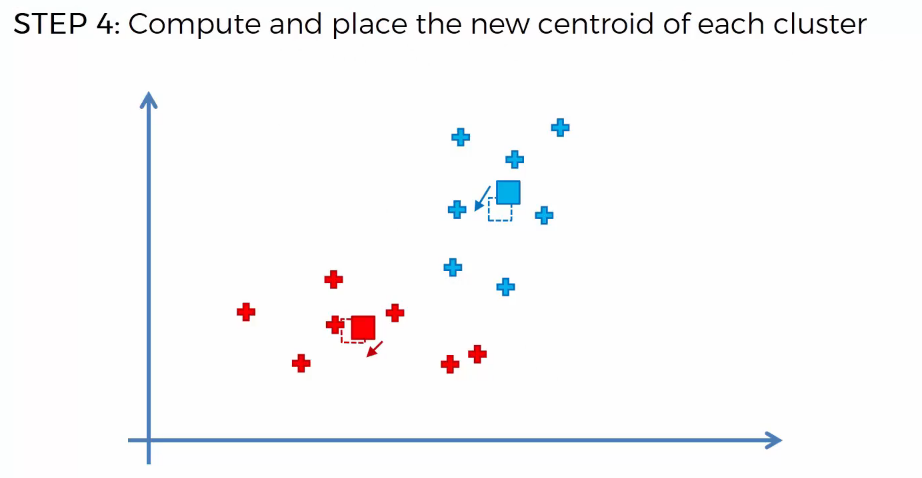


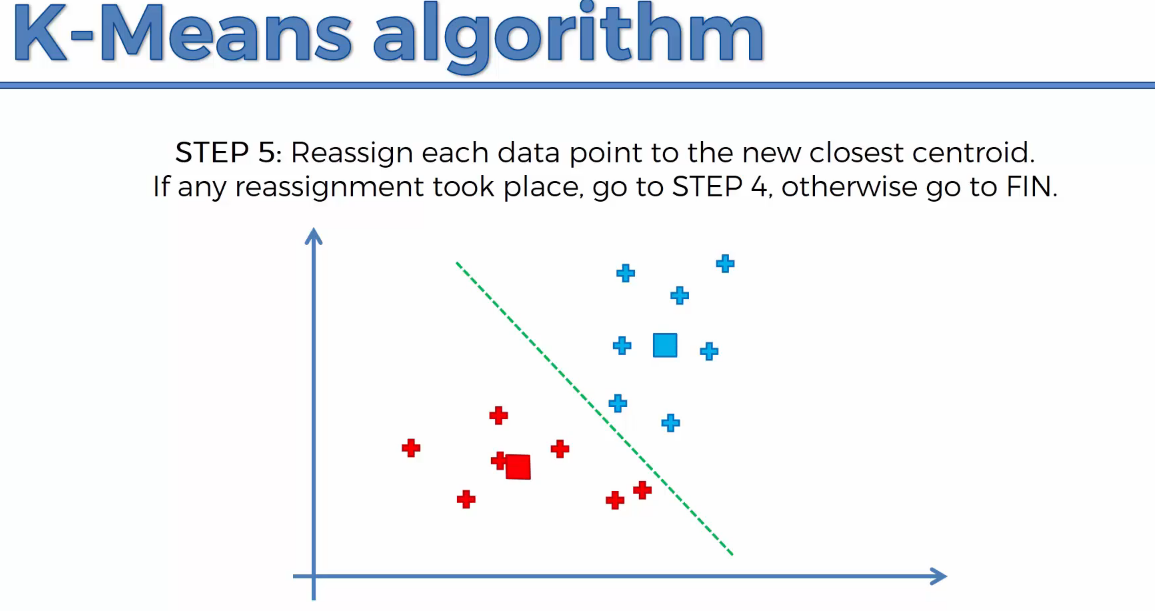




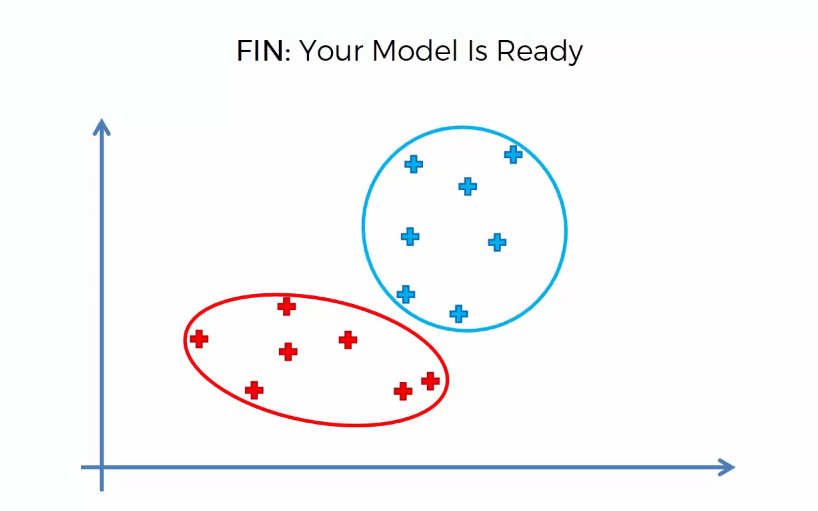


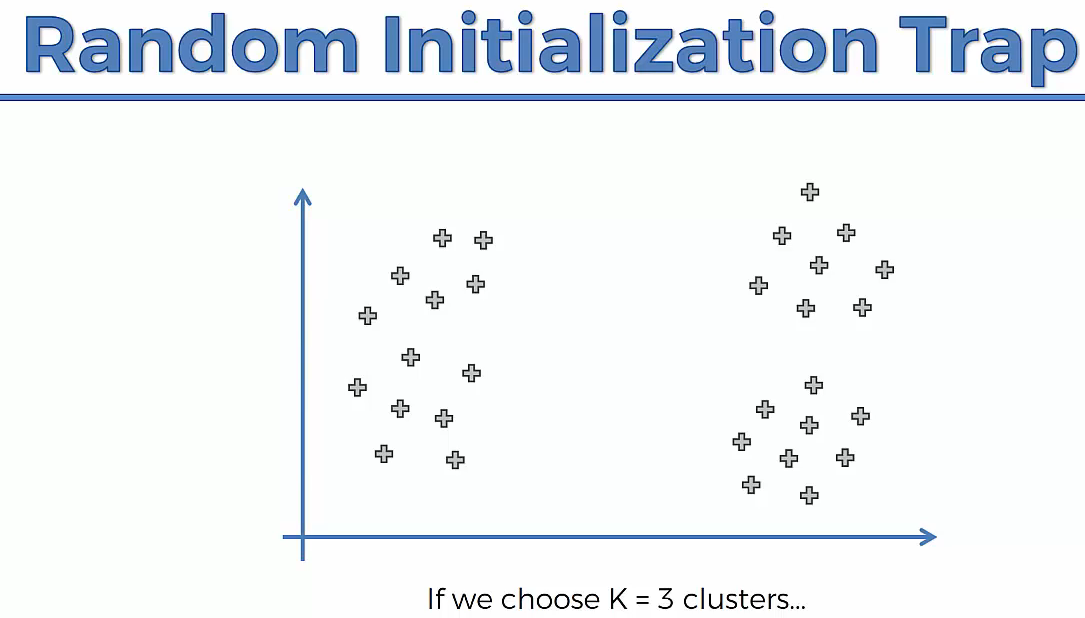


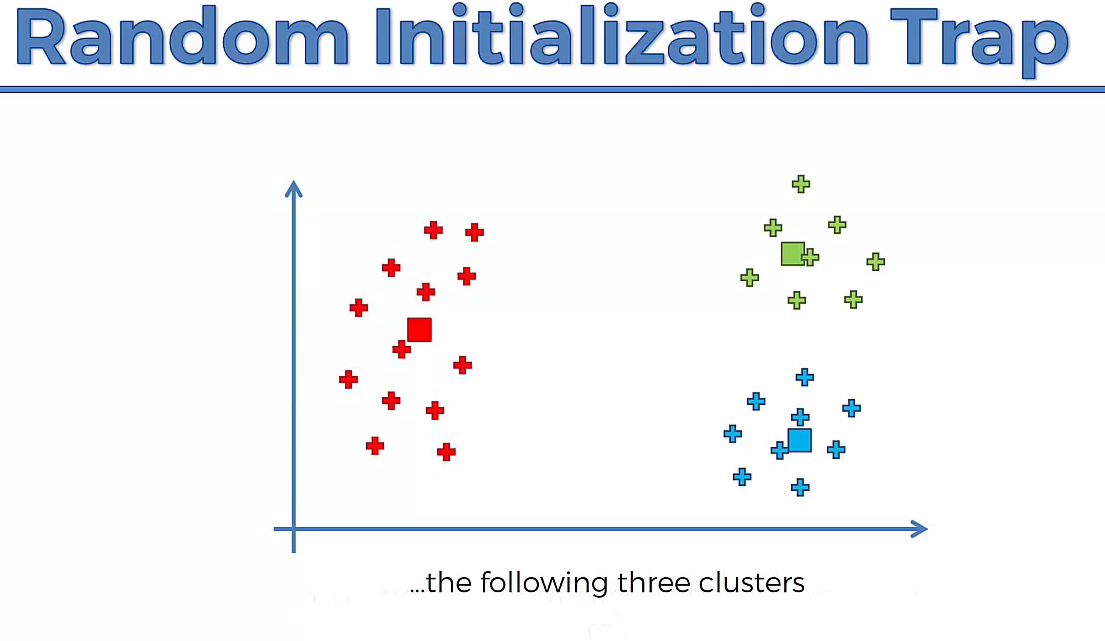


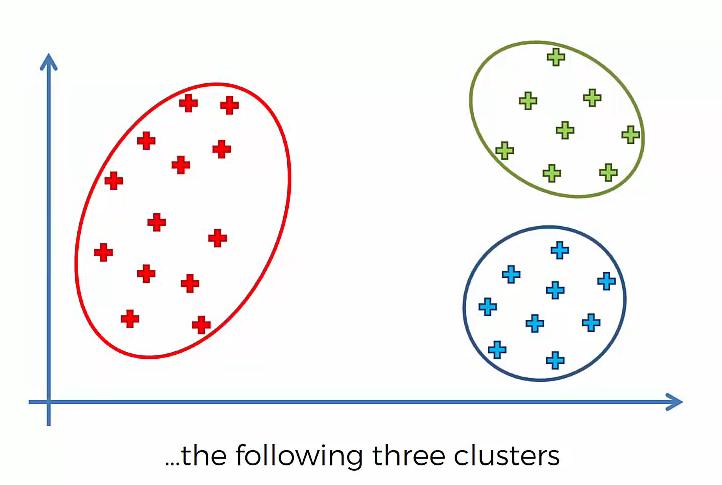


Algorithm is done!!!

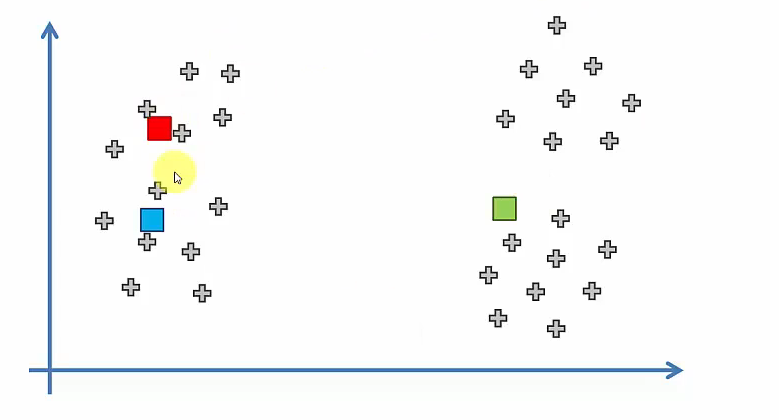


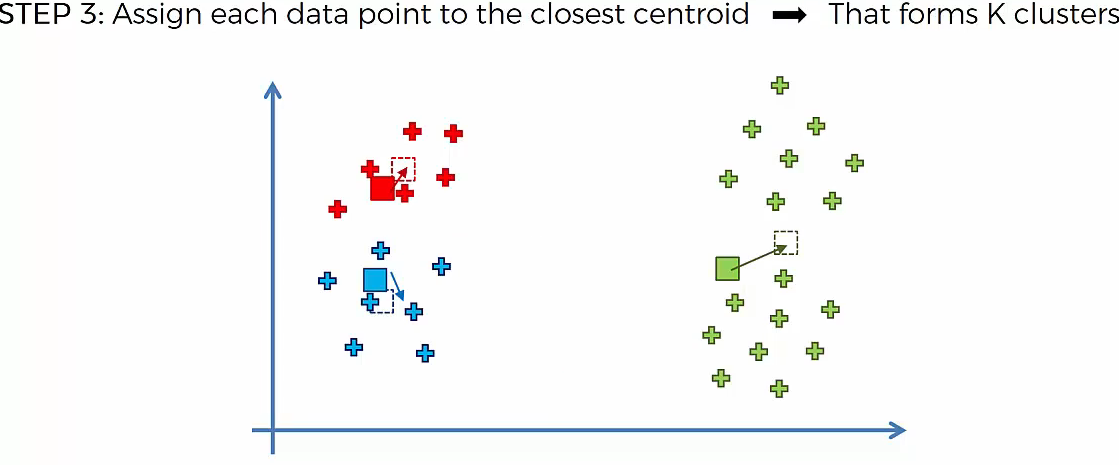


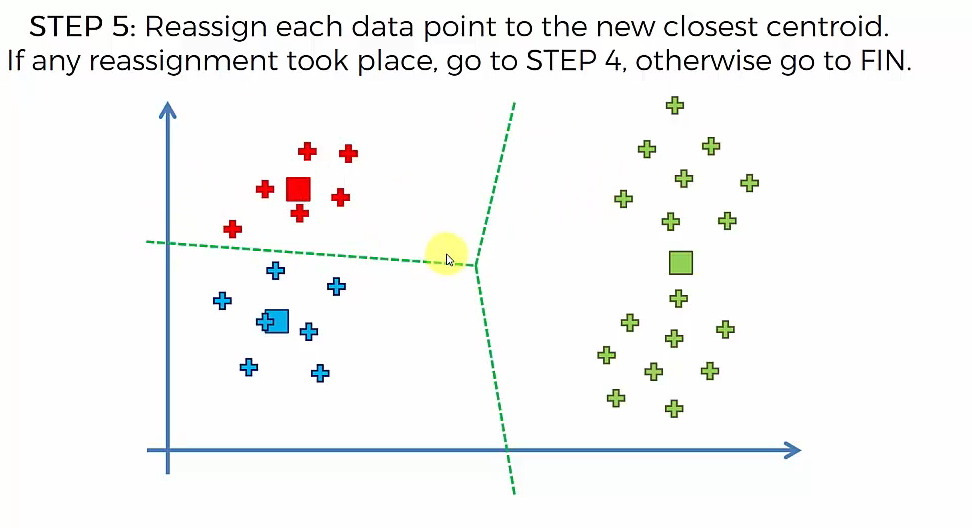


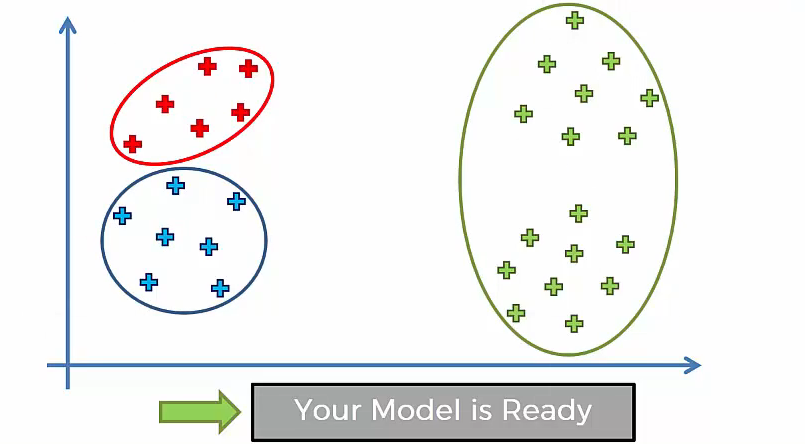




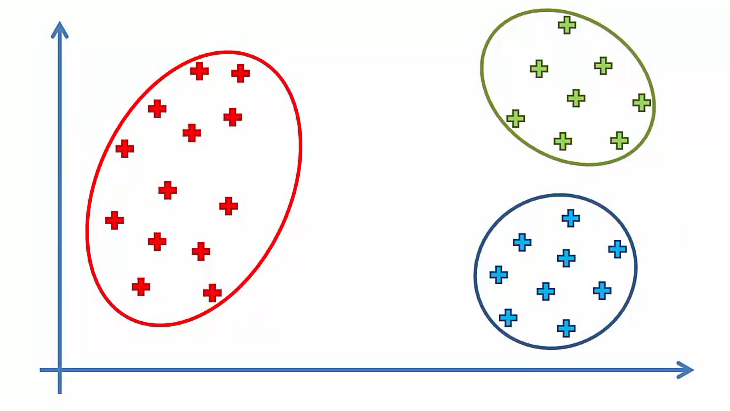






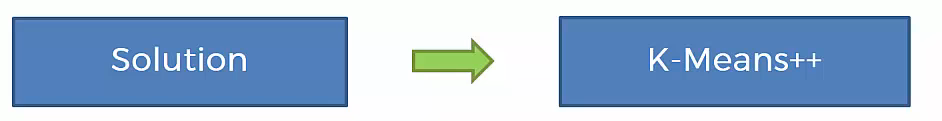


But before we got,

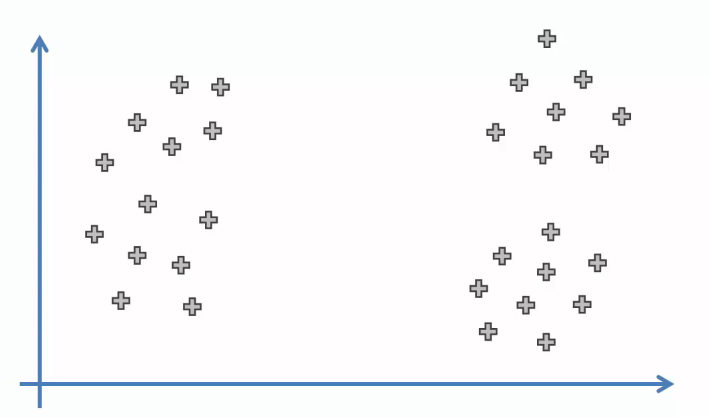
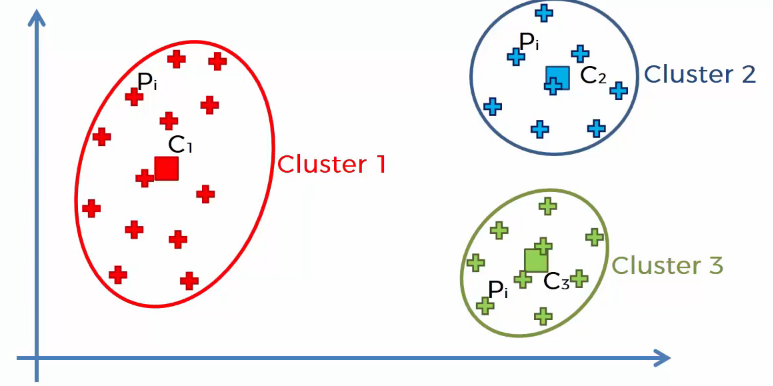


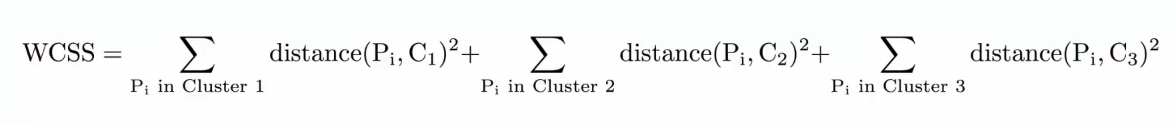
Selection of random points will impact the clustering

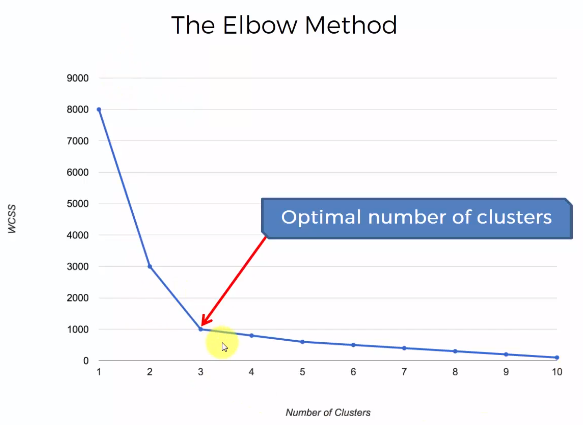
Solution,



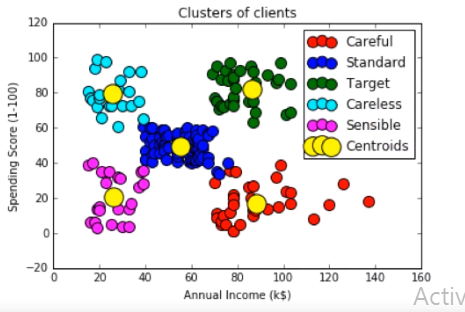


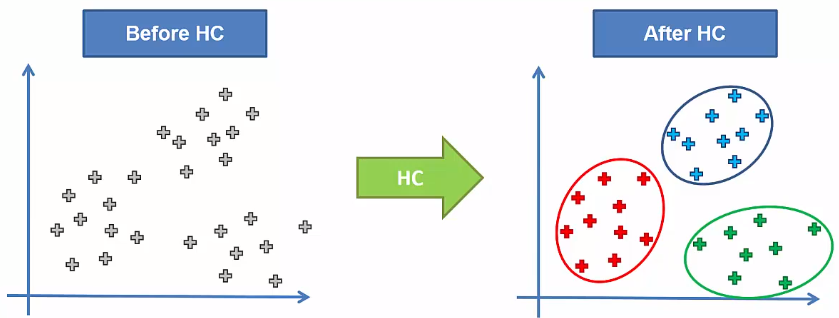




Look for the drop and choose the point as the number of clusters.



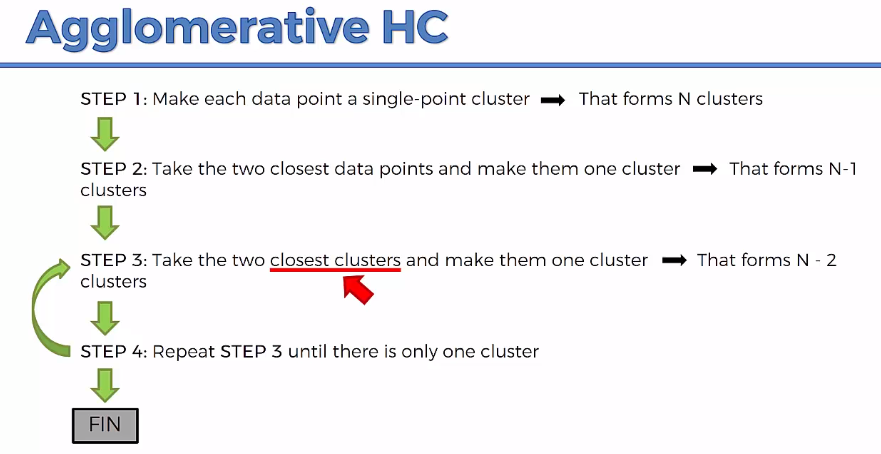
Hierarchical Clustering



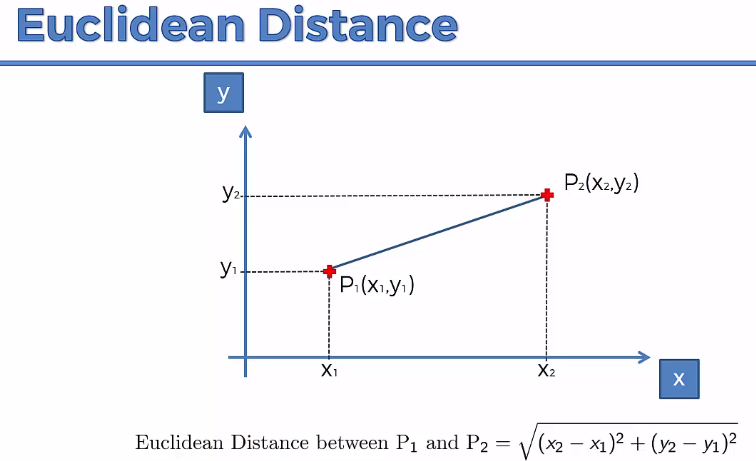
**Types:**

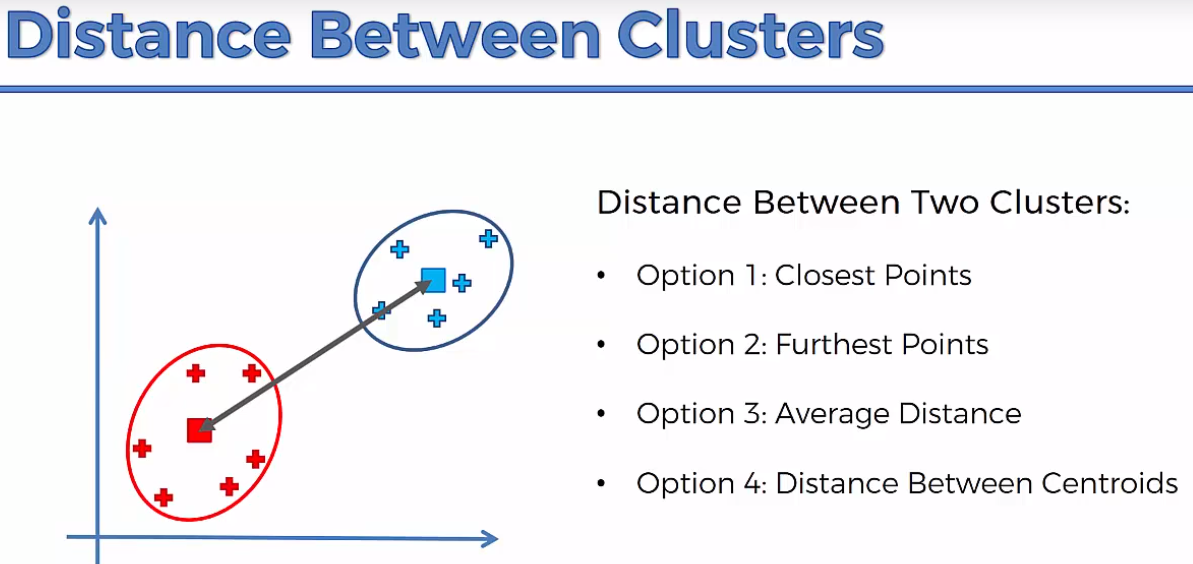
1. Agglomerative (Bottom to up)

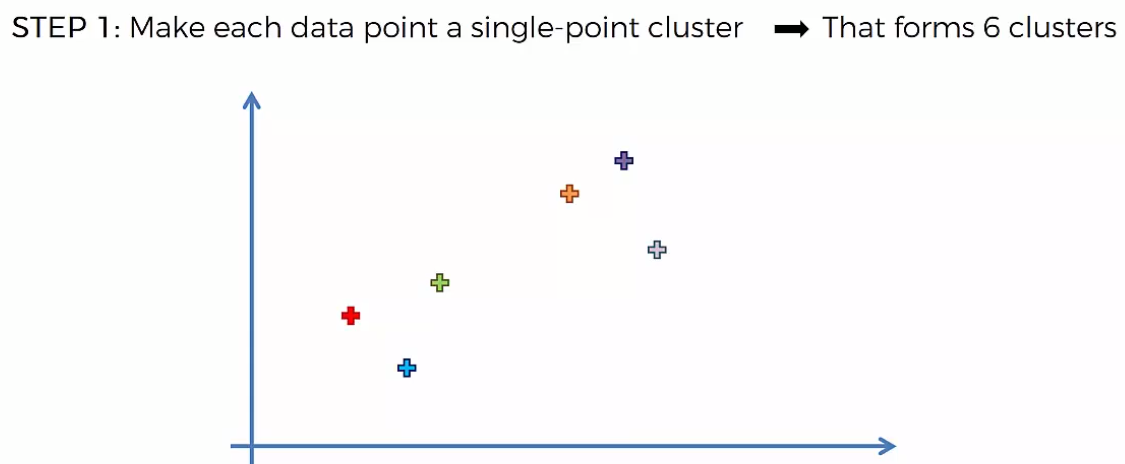
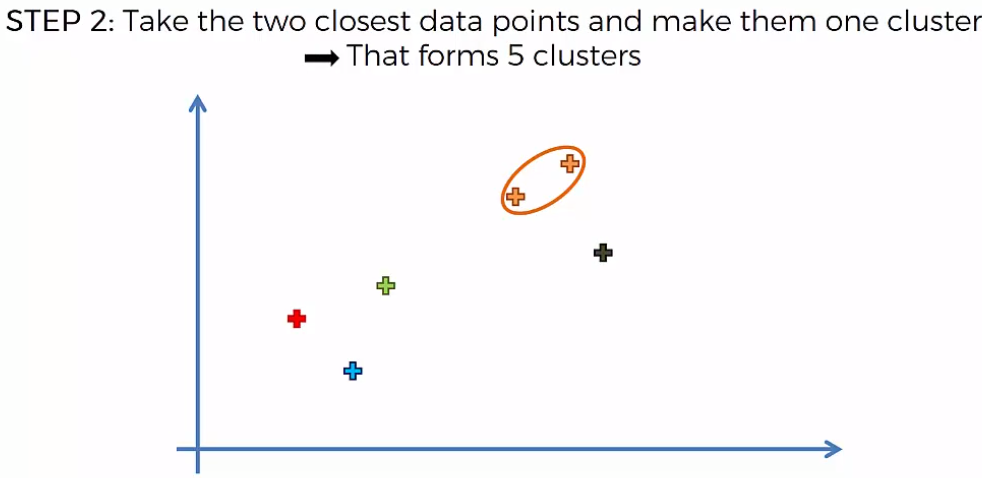
2. Divisive (Top to bottom)

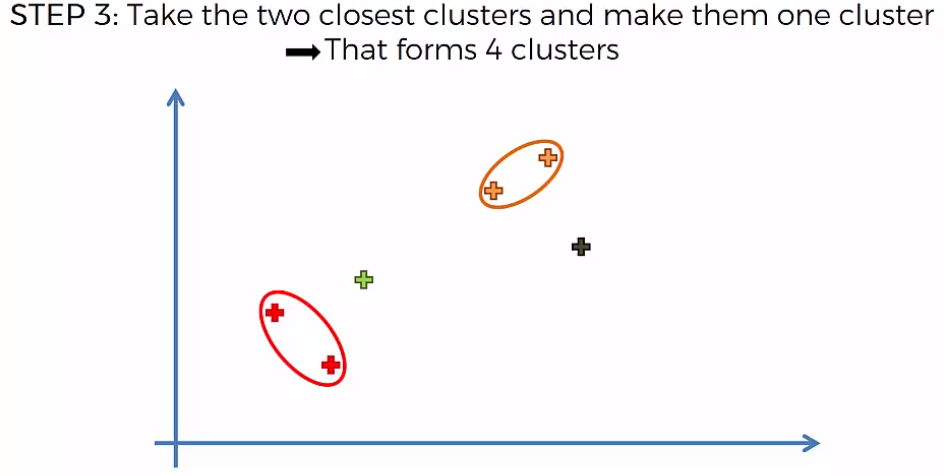


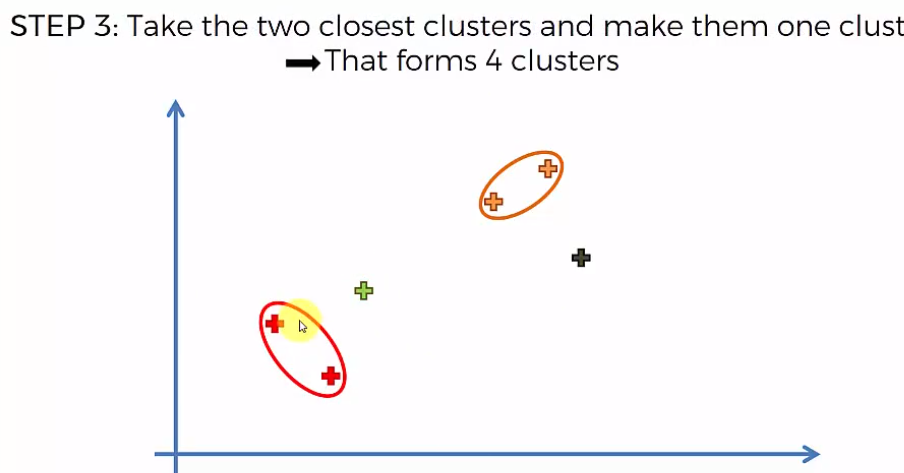
**Closest of clusters:**

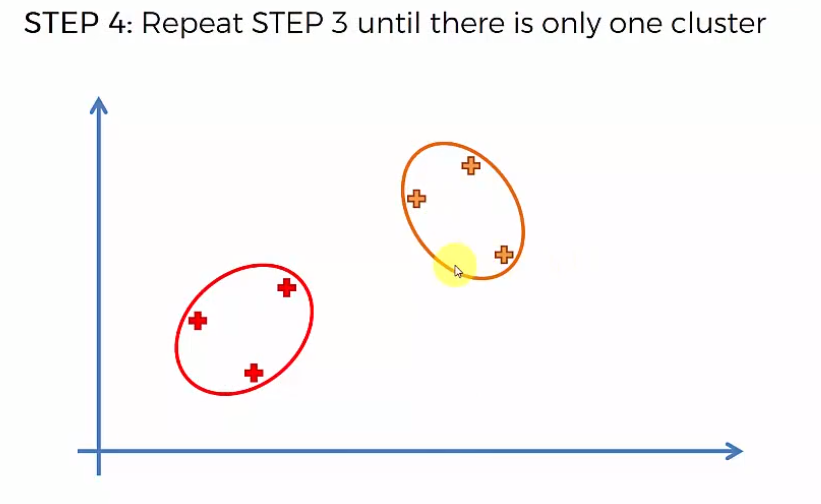
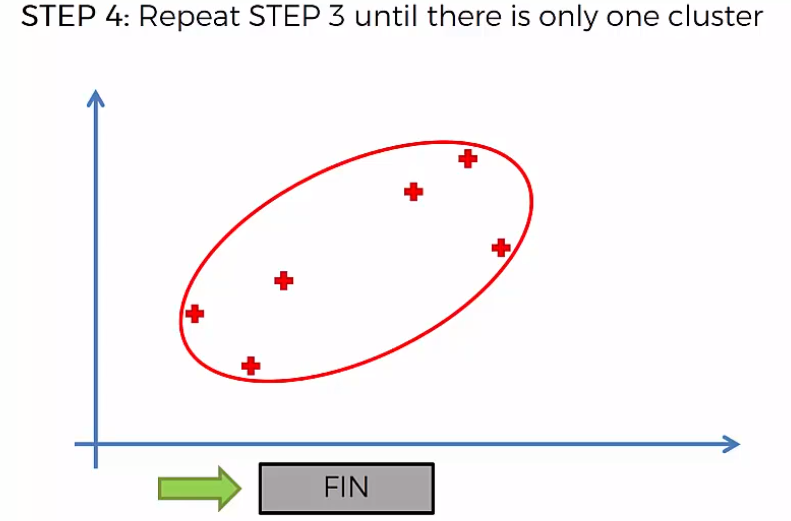






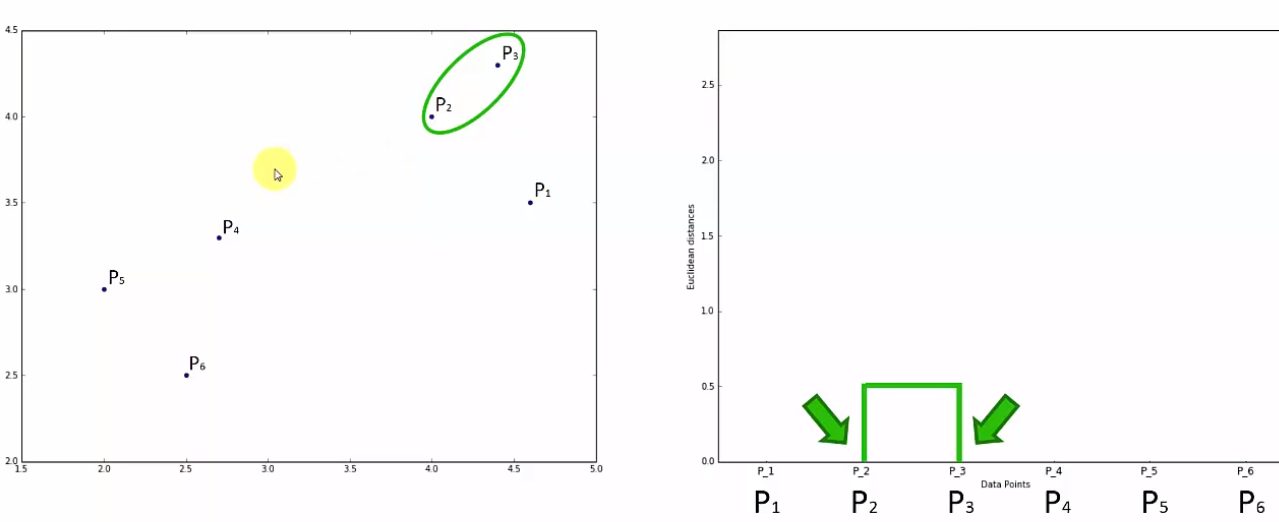


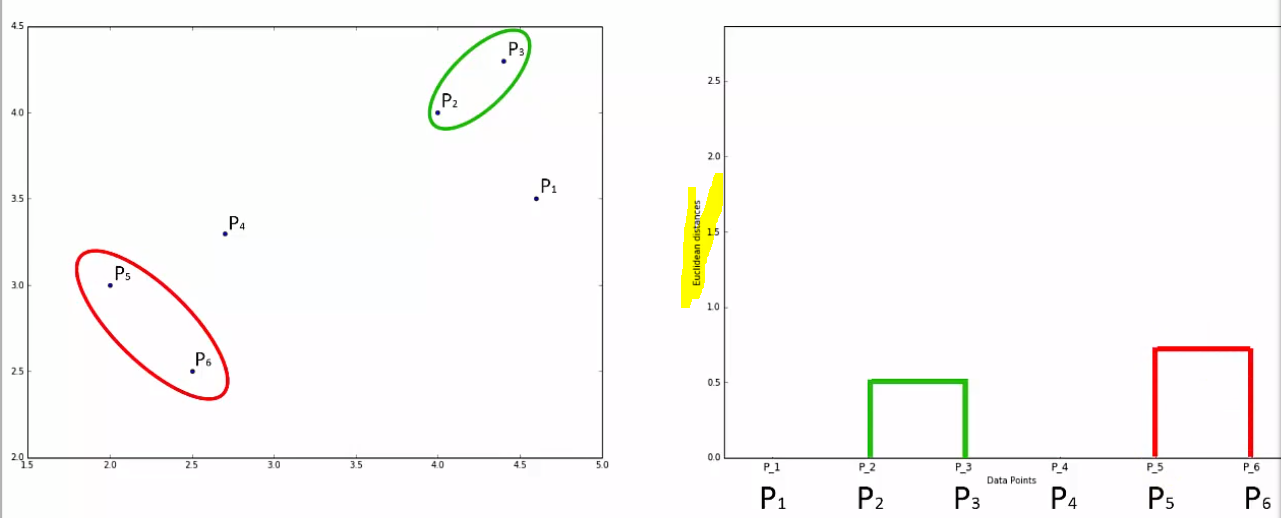
It maintains the memoryas dendrograms

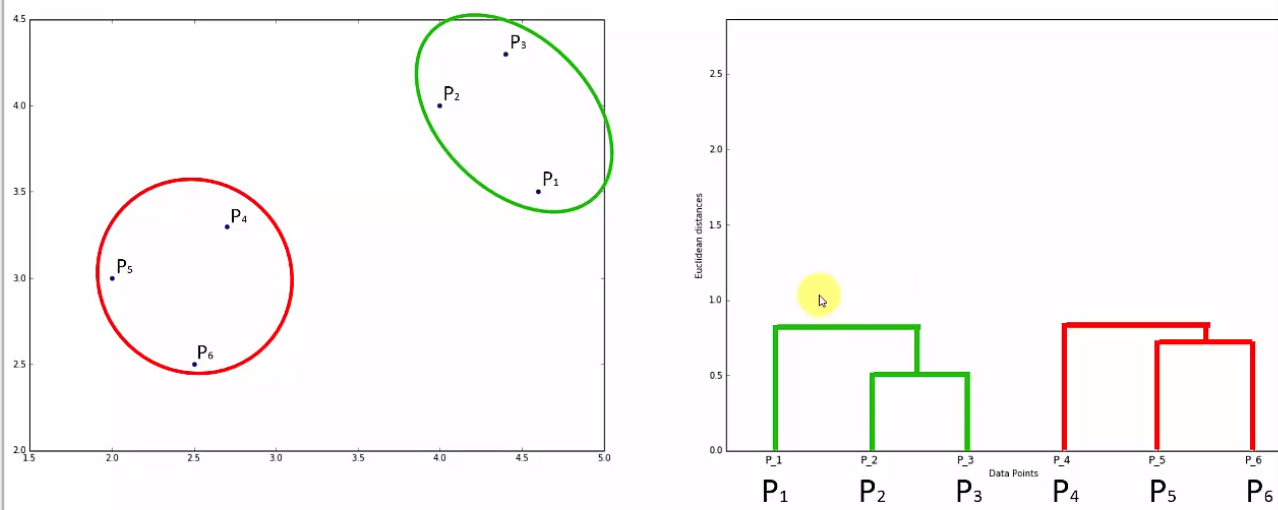


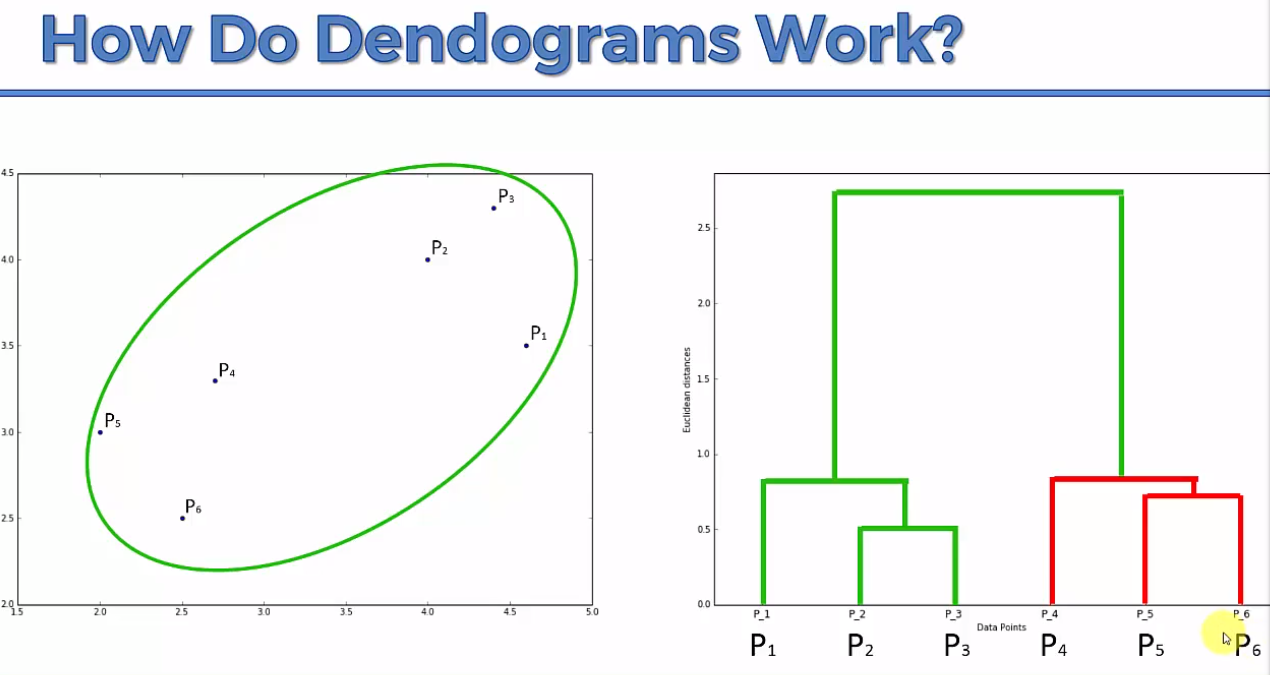
It is the memory of HC algorithm,



The bar height is the distance between the two points.

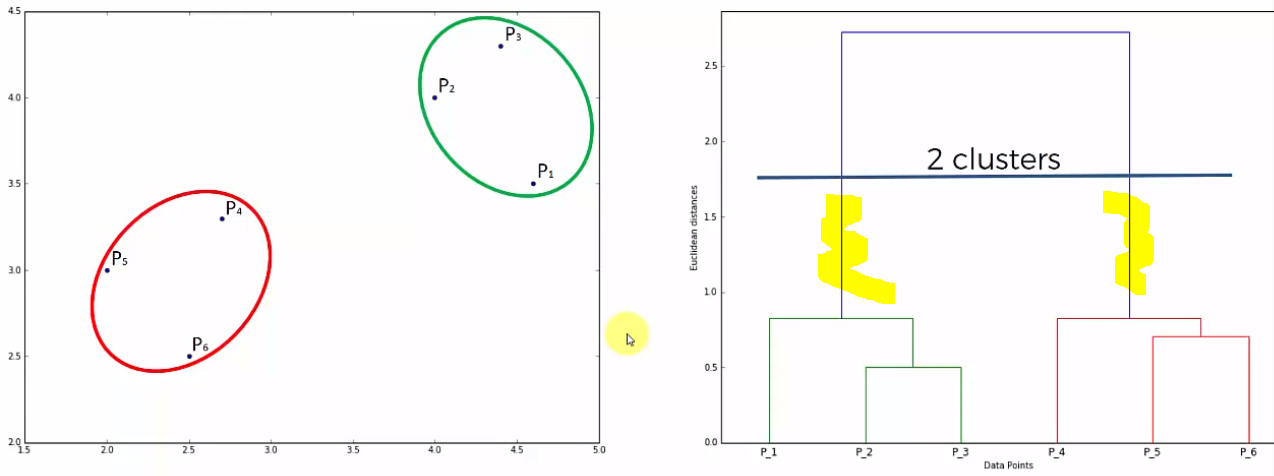




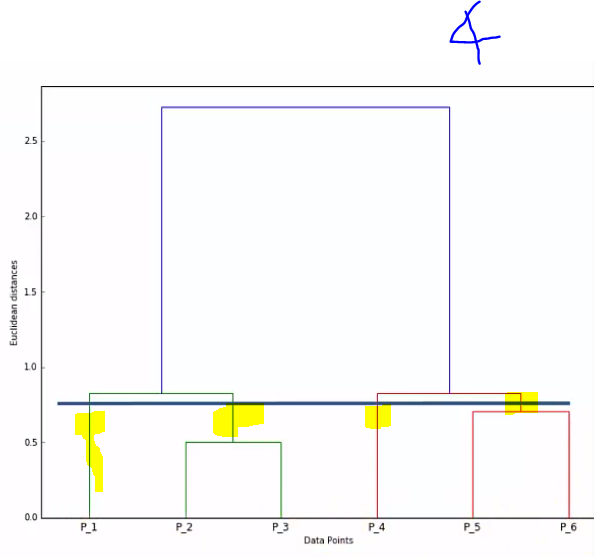


(Bottom – up Approach)

We can set the threshold for the dissimilarity (Any cluster having some distance greater than our threshold). Check the below example.



By seeing the dendrogram, crossing the threshold, we can say the number of clusters. Two clusters for the above example.



How to find the optimal # of cluster?

