**HIERARCHICAL CLUSTERING**

When we apply hierarchical clustering, results can be same as k-means but the process is different from k – means clustering.

There are two types of hierarchical clustering:

1. Agglomerative Clustering (bottom up approach)

2. Divisive (opposite of agglomerative clustering – starting at the top and dividing at the bottom)

Agglomerative Clustering

STEP 1 Make each data point a single-point cluster -> that forms n clusters

STEP 2 Take the two closest data points and make them one cluster -> that forms N - 1 clusters

STEP 3 Take the two closest clusters and make them one cluster -> that form N – 2 clusters

STEP 4 Repeat step 3 until there is only one clustering

FIN

Euclidaen distance in a 2-D space between 2 points is given by:

( (x2 - x1)2 – (y2 – y1)2 )1/2

Distance between two clusters can be calculated between any of the following:

Option 1: closest points

Option 2: furthest points

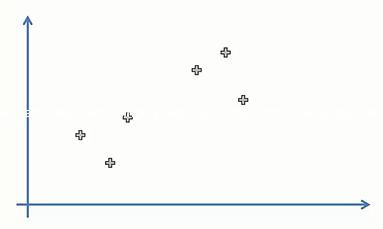
Option 3: average distance

Option 4: distance between centroids

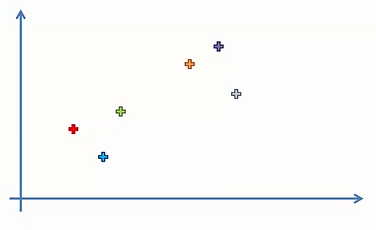
This significantly effects the outcome of the algorithm.

Step by step approach for agglomerative clustering

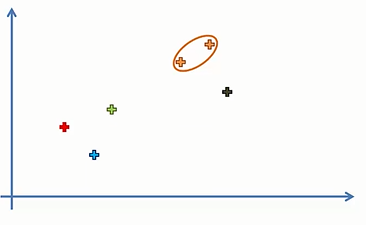
Consider the following dataset of N = 6 data points

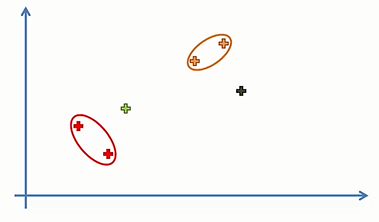


step 1 says make each data point a single-point cluster and that forms n clusters

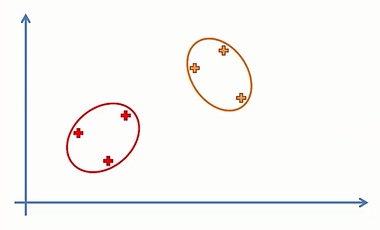
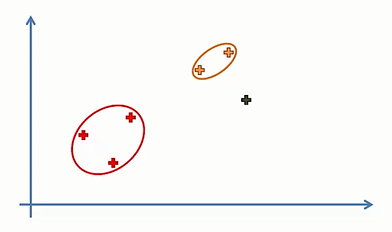


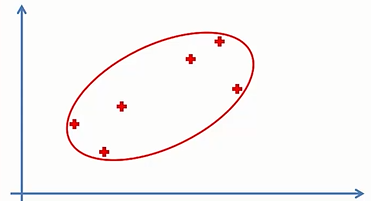
step 2 says take the two closest data points and make them one cluster and that forms N – 1, i.e 5 clusters

step 3 says take the two closest clusters and make them one cluster and that form N – 2, i.e 4 clusters



step 4 says repeat step 3 until there is only one clustering





This is the final graph, thus FIN

Purpose of this:

The way hierarchical clustering works is that it maintains a memory of how we went through this process and that memory is stored in a dendrogram.