There are two types of hierarchical clustering.

1. Agglomerative
2. Divisive (Reverse of the agglomerative)

Steps of Agglomerative clustering:

1. Make each data point a single cluster ------ It forms N number of clusters.
2. Take the two closest data points and make them one cluster --- It will then form N-1 clusters.
3. Take the two closest clusters that you now have and make them one cluster ---- it will make N-2 clusters
4. Repeat step 3 until there is one huge cluster left.

The hierarchical clustering algorithm maintains a memory for the whole process. That memory stored is called a dendograms.

How do the dendograms work?

While at the last of the hierarchical clustering, we get one huge cluster, then how we can divide the whole dataset from one big cluster to the optimal number of clusters? The answer for this is the dendograms memory.

The dendogram is the kind of memory of HC algorithm, it shows the connected points and connected clusters with the distances through the height on y-axis of the graph.

The horizontal line shows the connectivity, and the height of that line shows the Euclidean distances

How to choose optimal number of clustering?

When all the clusters are combined in the dendogram, and the whole tree is made then look at the horizontal level and set the threshold which are called (distance / dissimilarity threshold).

This threshold would describe that non of the cluster would have dissimilarity greater than the threshold you set.

The benefit of setting this threshold is that only by looking to the dendogram, we can say how many clusters would be made. The number of clusters is equal to the lines crossed by the dissimilarity line.

How to set the optimal threshold Or how to find the optimal number of clusters?

One of the approach is to look for the highest vertical distance that you find on the dendogram that does not cross other line(s).

Set the threshold on the highest line in the dendogram that does not cross other line….