**Hierarchical Clustering Intuition**

* If you have scatterplots or data points that we looked at previously, if you apply hierarchical clustering, you will get clusters again, very similar to k-means clustering but the process is a bit different.
* Note: There are two types of hierarchical clustering – **Agglomerative and Divisive.**
* Agglomerative is the bottom-up approach, whereas Divisive is a top-down approach.
* In this section, we will be focusing on agglomerative approach.

**Steps for Agglomerative Hierarchical Clustering –**

**Step 1:** Make each data point a single-point cluster. That forms N clusters. So, if we have N data points, we form N clusters for individual data points.

**Step 2:** Take the two closest datapoints and make them one cluster. That forms N-1 clusters.

**Step 3:** Take the two closest clusters and make them one cluster. That forms N-2 clusters.

**Step 4:** Repeat Step 3 until there is only one cluster. You keep repeating step 3 and combining points in to bigger and bigger clusters, until there is only one huge cluster left.

**FINISH**

* The one thing that stands out here are the words “Closest Clusters”. To find the closest cluster, we need to find the distance between them. How do we find/measure the distance between two clusters?
* Measuring distance between two clusters is not as obvious as it sounds. There are couple of options to measure distance between clusters:

**Option 1 –** Take closest points, measure them, and call them two closest clusters.

**Option 2 –** Take two furthest points, measure them, and call them two closest clusters.

**Option 3 –** Take the average of the distances of different points in the cluster and take the average of that distance.

**Option 4 –** Measure the distance between the centroids, and that would be the distance between the clusters.

* It is a very important part of the hierarchical clustering what you define as the distance between two clusters because that can significantly impact the output of your algorithm.

Let’s look a step-by-step approach in an example of **Agglomerative Clustering** –

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

A picture containing scatter chart

Description automatically generated

Shape

Description automatically generated with medium confidence

Diagram

Description automatically generated with low confidence

* The way the hierarchical clustering algorithm works is that it maintains the memory of how the one big cluster was formed.
* And that memory is stored in a dendrogram.

**How Do Dendrograms Work?**

Chart

Description automatically generated

* Here, on the left we have got a chart with 6 points, and on the left we have got another chart, which we are going to use to create a dendrogram.
* On the vertical axis we have got Euclidean distance.
* We will go through the Hierarchical Clustering Algorithm and create the dendrogram as we go.
* To start off with, every single point is a single cluster.
* Next, we will find the closest points and put them into one cluster. Those points would be P2 and P3.
* Now we would like to represent on the dendrogram that points P2 and P3 were the closest points on the scatterplot because dendrogram is the memory of the HC (Hierarchical Clustering) algorithm – it will store every single step we perform.
* We will plot a line between the two closest points on the dendrogram to show that they are the two closest points.

Chart, histogram

Description automatically generated

* The height of the line (the y-axis) represents the distance between the two points on the scatterplot.
* To keep it simple, the further two points are, the more dissimilar they are, and that is being captured by the dendrogram by the height of the y-axis. So basically, the dendrogram captures the dissimilarities between two points on the scatterplot by setting the height of the bar. And the bar itself shows that we connected P2 and P3.
* Then, we will move on to the next step in the HC algorithm, step 3 – find the next two closest points on the scatterplot and form their cluster. The next two closest points on the scatterplot are P5 and P6.

Chart, histogram

Description automatically generated

* The bar between P5 and P6 is placed higher because the distance between the two points on the scatterplot is more than the distance between P2 and P3.
* The next step is to repeat step 3 and look among all the other clusters and determine which are the closest. And we determined that cluster of P1 is closest to the P2-P3 cluster. Thus, we combine those two clusters into one cluster.
* We plot the same on the dendrogram as follows:

Chart, histogram

Description automatically generated

* And the height of y-axis connecting P1 with the cluster of P2, P3 represents the dissimilarities (distance) between those two clusters.
* In the next step we again repeat step 3 and find the closest among the other clusters to form one single cluster. And we can see that cluster P4 is closest to the cluster of P5, P6.
* So, we combine them into one single cluster and plot the same on the dendrogram as shown below:

Chart, histogram

Description automatically generated

* In the final step we combine the two remaining clusters, since there are no other clusters left, and represent them on the dendrogram as shown below:

Chart, histogram

Description automatically generated

* That is how we construct our dendrogram using the bottom-up approach, and this is what it means when it says that dendrogram contains the memory of the hierarchical clustering algorithm. So, just by looking at the dendrogram we can determine in which order the hierarchical clustering was formed.

**Using Dendrograms for Hierarchical Clustering:**

* We have already seen the working of Hierarchical Clustering Algorithm, and we have also understood the working of Dendrograms and how they are constructed, now we will learn how to put them together and get the maximum output from our hierarchical clustering algorithms.
* We will take the same example that we used earlier to understand the concept better.
* So, what we can do with dendrograms is look at horizontal levels, and set thresholds – distance thresholds, or dissimilarities thresholds because the y-axis measure the Euclidean distance between the two clusters – for our dissimilarities and define that we don’t want the dissimilarity to exceed a certain value/level.
* To keep it simple, we don’t want to have within a cluster dissimilarity above a certain threshold.

Chart

Description automatically generated

* In our example we can see that after setting a threshold, of let’s say 1.75, on the y-axis we will only end up with two clusters that have dissimilarity level less than 1.75.

**How do we find the optimal number of clusters using the Dendrogram?**

* One of the standard approaches is to look for the highest vertical distance that you can find on the dendrogram – basically, any vertical line that will not cross any of the horizontal lines on the dendrogram.
* So, as mentioned before, we take the longest vertical line that will not cross any of the horizontal line and set a threshold for the vertical line, and use that threshold to calculate the optimal number of clusters.