**1. Agglomerative Hierarchical Clustering**

Hierarchical Clustering is same as K-Means clustering but the process is different. There are two types of Hierarchical Clustering – Agglomerative and Divisive Hierarchical 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 forms N-2 clusters.

Step 4: Repeat step 3 until there is only one cluster

To calculate the closeness between the data points as mentioned in the above algorithm, we use Euclidean distance.

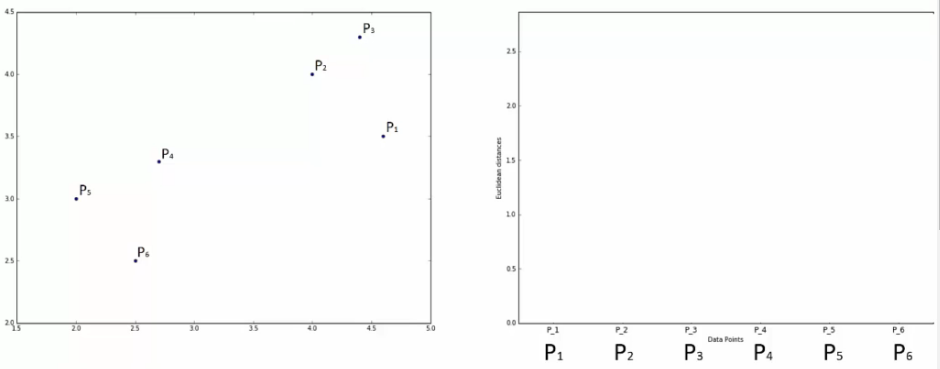
Euclidean distance between points P1 and P2 =

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

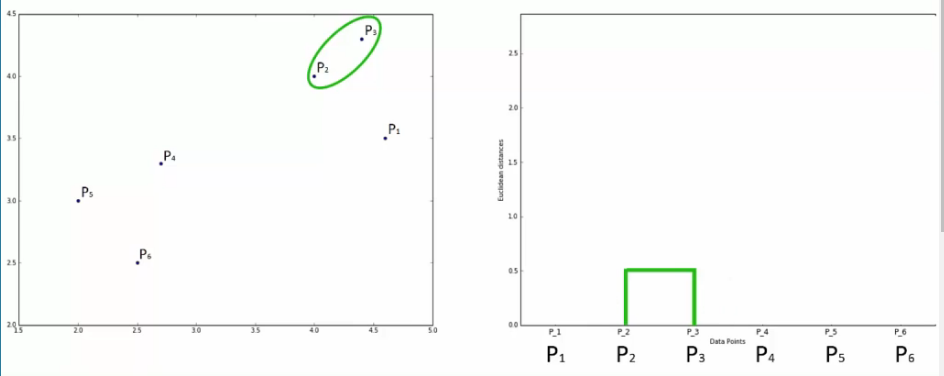
**1.1 Dendrogram**

A dendrogram is a diagram representing a tree. In hierarchical clustering, it illustrates the arrangement of the clusters produced by the corresponding analyses.

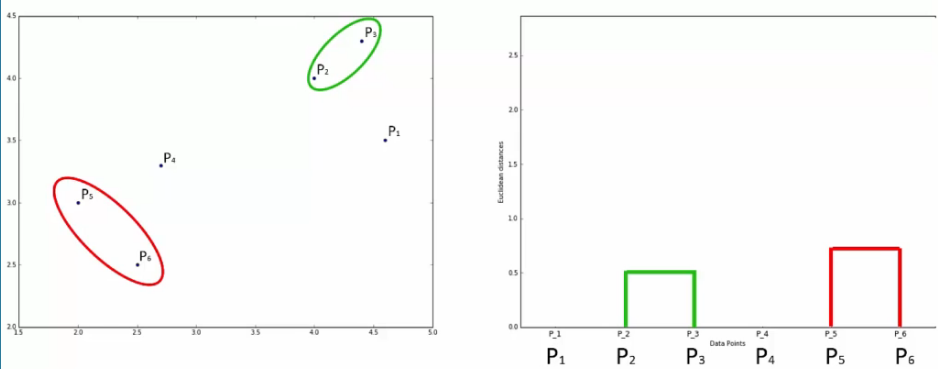
Consider the following example where on the left we have got 6 points and on the right we have a dendrogram. We will follow the steps in AHC and will plot the dendrogram simultaneously.



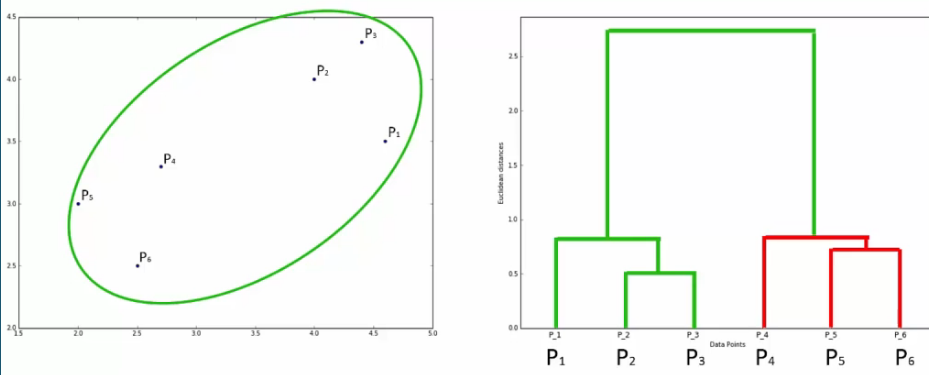
First, we assign all points as individual clusters. Next, we select the two closest points and cluster them together. The height of green box indicates the Euclidean distance between the two points.



We now find the next two closest points and form a cluster. Simultaneously, we mention it in the dendrogram.

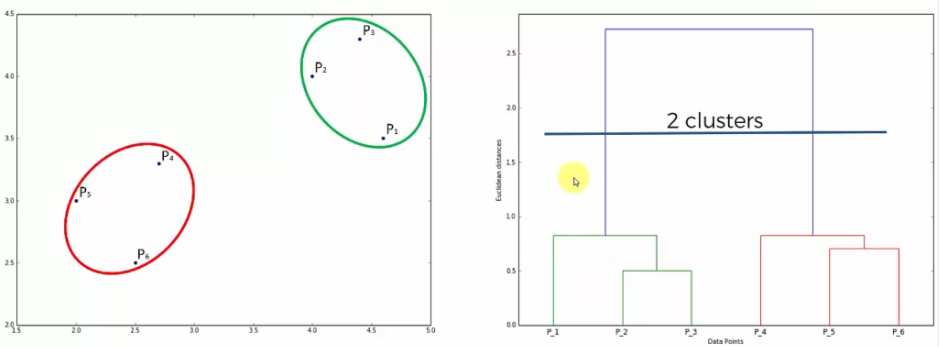


We will follow the steps in a similar way and finally we get the following result:



**1.2 Hierarchical Clustering using Dendrograms**

Dendrograms can be used to set a threshold for forming clusters. The threshold value will be the Euclidean distance. A cluster won’t be allowed to form if the distance between the points forming that cluster is greater that the threshold.



In the above example, we have set the threshold to approximately 1.7. Therefore, we will get only 2 clusters instead of three.

An interesting observation to note is that we can tell how many clusters will the given data distribution have based on the number of lines the threshold line passes through. In the above example, the threshold line passes through 2 vertical lines. Hence, we have 2 clusters.

But will the dendrogram be able to give us any hint about the optimal number of clusters? Yes, and the hints are the vertical lines. As said before, the vertical lines give us an idea about the dissimilarity between any two points. The higher the dissimilarity, further apart are the two data points and hence lower the chances of them being in the same cluster.

To find the optimal number of clusters, find the longest vertical line that doesn’t get crossed by any of the horizontal lines. And take any threshold that will pass through that line (in above example – 1.7).