Hierarchical Clustering

- Broadly categorizing, there are two ways of performing Hierarchical Clustering.

  1. **Agglomerative Clustering:**
     - ➔ The word agglomerative suggests combining things
     - ➔ It is a bottom-up approach
     - ➔ Agglomerative clustering starts with the assumption that every data point is a cluster
     - ➔ Then, it groups the clusters which are close to each other until there is only a single cluster left

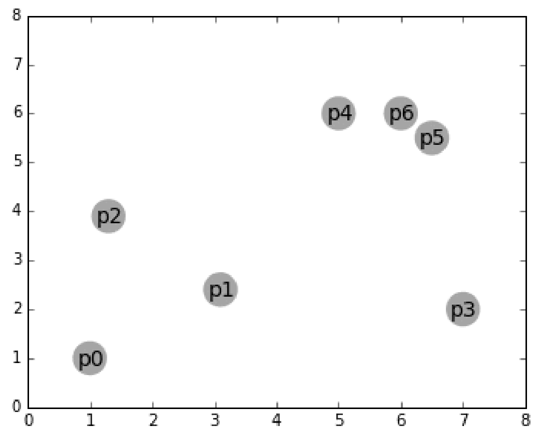
  2. **Divisive Clustering:**
     - ➔ It is the complete opposite of the agglomerative approach
     - ➔ It is a top-down approach
     - ➔ It starts with one big cluster that contains all the data points.
     - ➔ It then divides the points into different clusters till each data point is a cluster itself

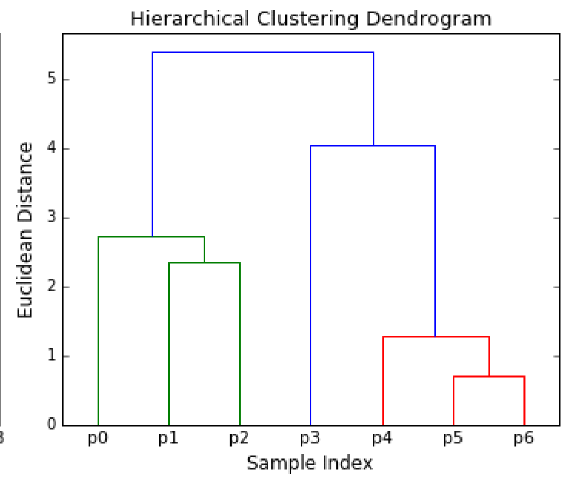
Agglomerative Clustering

- The steps involved in Agglomerative Clustering are:

  1. Assume each point is a cluster (n datapoints -> n clusters)
  2. Compute Proximity Matrix ($Pn*n$)
  3. Repeat until a single cluster is left:
     a. Merge the closest clusters
     b. Update the proximity matrix

- If we visualize this, this looks like a Tree, but there is another name that is often used in Data Mining terminology which is called Dendrogram.





Proximity Matrix

- Proximity matrix is a matrix of distances or similarity.
- The word proximity suggests how close things are
- Say, at any point we're having C_m clusters. For each of the pairs of clusters, the proximity matrix P will indicate the similarity between clusters C_i and C_j.

- Initially the proximity matrix P will be $N*N$ matrix.
- Suppose cluster Ci and Cj, where $i{\neq}j$, are similar and they have the smallest value in the proximity matrix, then those clusters will be combined and proximity matrix will get updated
- The new matrix will be a $N{-}1*N{-}1$ matrix, as two clusters have combined.

- One can use the following distances for computing the values of proximity matrices.

   1. Using Euclidean Distance between the centroids of two clusters Ci and Cj.
   2. Maximum distance between two points xi and xj, such that $xi \in Ci$ and $xj \in Cj$.
   3. Minimum distance between two points xi and xj, such that $xi \in Ci$ and $xj \in Cj$.

$$\sum_{x_i \in C_i} \sum_{x_j \in C_j} \frac{dist(x_i x_j)}{|C_i||C_j|}$$

4. Average Distance:

5. Ward's Distance: $\sum_{x_i \in C_i} \sum_{x_j \in C_j} \frac{dist(x_i x_j)^2}{|C_i||C_j|}$

Limitations of Hierarchical Clustering

1. With large datasets, Agglomerative Clustering does not work well
   a. Space Complexity = $O(n)$: Proximity Matrix
   b. Time Complexity = $O(n^2)$
2. Unlike K-means where we try to minimize **within-cluster distance**, there is **no mathematical objective** that is being minimized in Agglomerative clustering.