***HierarchicalClustering Theorem***

Hierarchical clustering, as the name suggests is an algorithm that builds hierarchy of clusters. This algorithm starts with all the data points assigned to a cluster of their own. Then two nearest clusters are merged into the same cluster. In the end, this algorithm terminates when there is only a single cluster left. Strategies for hierarchical clustering generally fall into two types:

**Agglomerative**: This is a "**bottom up**" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

**Divisive**: This is a "**top down**" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

In general, the merges and splits are determined in a greedy manner. The results of hierarchical clustering are usually presented in a **dendrogram**.

The standard algorithm for hierarchical agglomerative clustering (HAC) has a time complexity O(n power3) and requires O( N power 2)

**Diff between HC and K-mean?**

Hierarchical clustering can’t handle big data well but K Means clustering can. This is because the time complexity of K Means is linear i.e. O(n) while that of hierarchical clustering is quadratic i.e. O(n2).

In K Means clustering, since we start with random choice of clusters, the results produced by running the algorithm multiple times might differ. While results are reproducible in Hierarchical clustering.

K Means is found to work well when the shape of the clusters is hyper spherical (like circle in 2D, sphere in 3D).

K Means clustering requires prior knowledge of K i.e. no. of clusters you want to divide your data into. But, you can stop at whatever number of clusters you find appropriate in hierarchical clustering by interpreting the dendrogram

The k-means algorithm is parameterized by the value k, which is the number of clusters create. The algorithm begins by creating k centroids. It then iterates between an assign step (where each sample is assigned to its closest centroid) and an update step (where each centroid is updated to become the mean of all the samples that are assigned to it. This iteration continues until some stopping criteria is met; for example, if no sample is re-assigned to a different centroid.

The k-means algorithm makes a number of assumptions about the data, which are demonstrated in this scikit-learn example: demonstration of k-means assumptions. The most notable assumption is that the data is 'spherical,' see how to understand the drawbacks of K-means for a detailed discussion.

HCA- instead, builds clusters incrementally, producing a dendogram. the algorithm begins by assigning each sample to its own cluster (top level). At each step, the two clusters that are the most similar are merged; the algorithm continues until all of the clusters have been merged. Unlike k-means, we don't need to specify a k parameter: once the dendogram has been produced, you can navigate the layers of the tree to see which number of clusters makes the most sense to your particular application.

**What is Dendrogram?**

A dendrogram is a type of tree diagram showing hierarchical clustering — relationships between similar sets of data. They are frequently used in biology to show clustering between genes or samples, but they can represent any type of grouped data.

