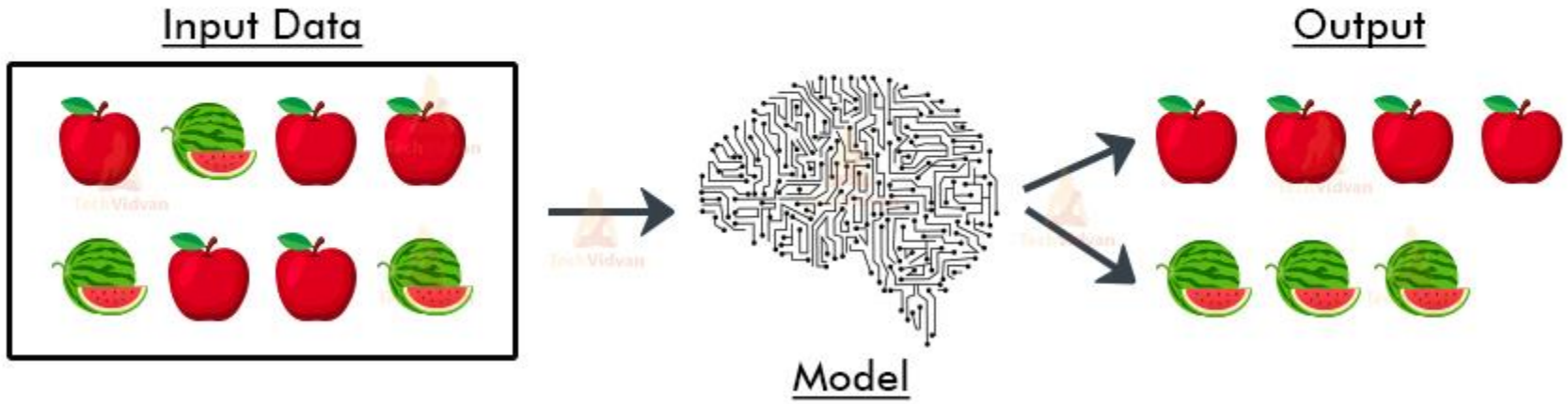
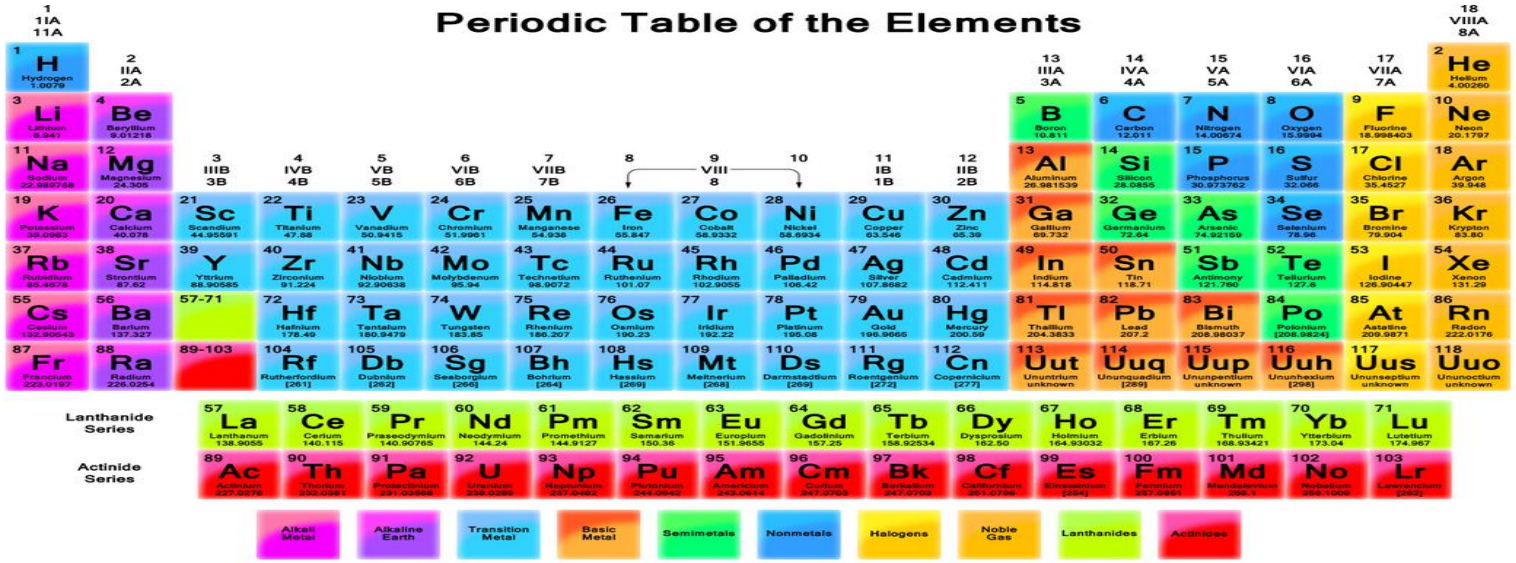
Unsupervised Learning in ML



Clustering

Clustering is the process of dividing the dataset into groups, consisting of similar data-points

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters.



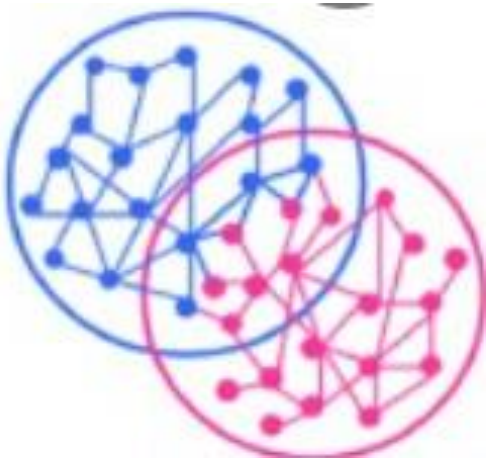
Types of Clustering

1. Exclusive Clustering
2. Overlapping Clustering
3. Hierarchical Clustering

Exclusive Clustering: Exclusive Clustering is the hard clustering in which data point exclusively belongs to one cluster.
For example, K-Means Clustering.

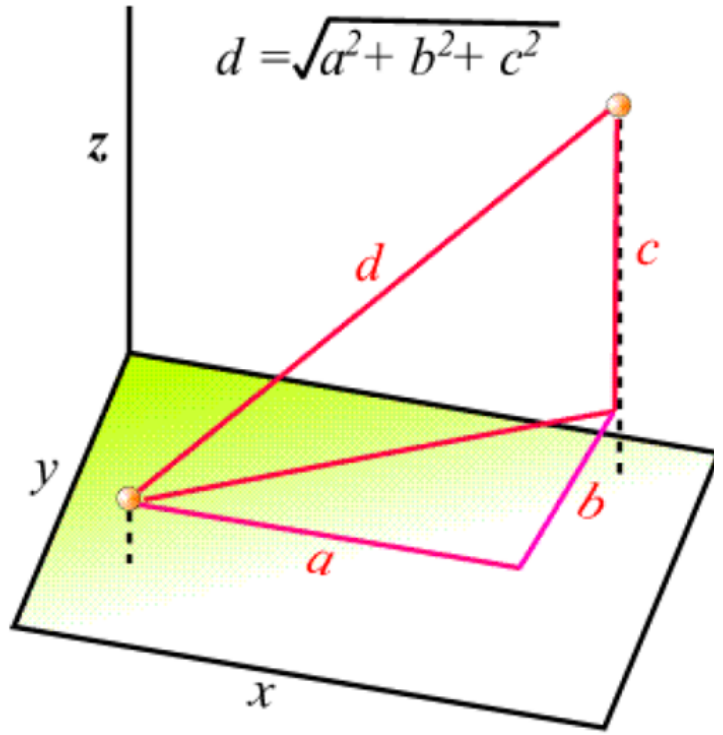
Overlapping Clustering:

Overlapping clustering is the soft cluster in which data point belongs to multiple clusters.
For example, C-Means Clustering.



we can see that some of the blue data points and some of the pink data points are overlapped.

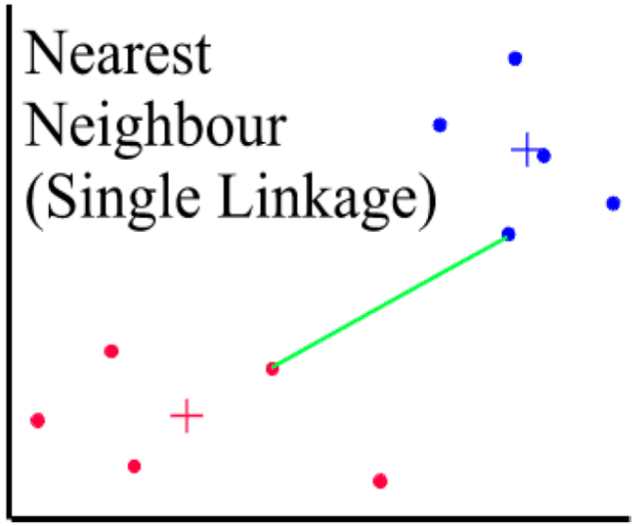
Euclidean Distance



$$d_{ij} = \sqrt{\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2 + \cdots + \left(x_{ip} - x_{jp}\right)^2}$$

Distances Between Clusters: ‘single linkage’ (‘nearest neighbor’)

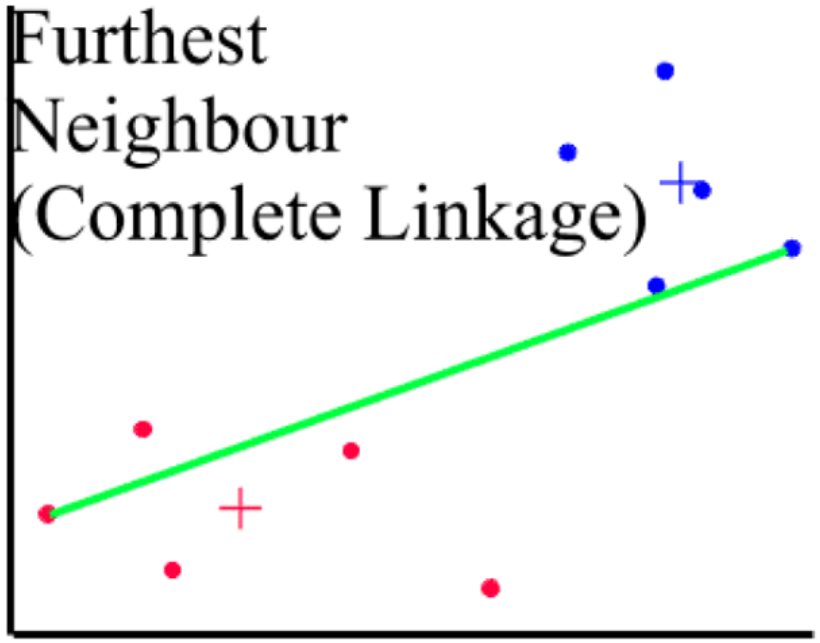
Distance between 2 clusters = **minimum distance** between members of the two clusters



$$\Delta\left(\mathbf{C}_{\alpha},\mathbf{C}_{\beta}\right)=\min_{\mathbf{x}\in\mathbf{C}_{\alpha},\mathbf{y}\in\mathbf{C}_{\beta}}\left\{\Delta\left(\mathbf{x},\mathbf{y}\right)\right\}$$

Distances Between Clusters: 'complete linkage' ('farthest neighbor')

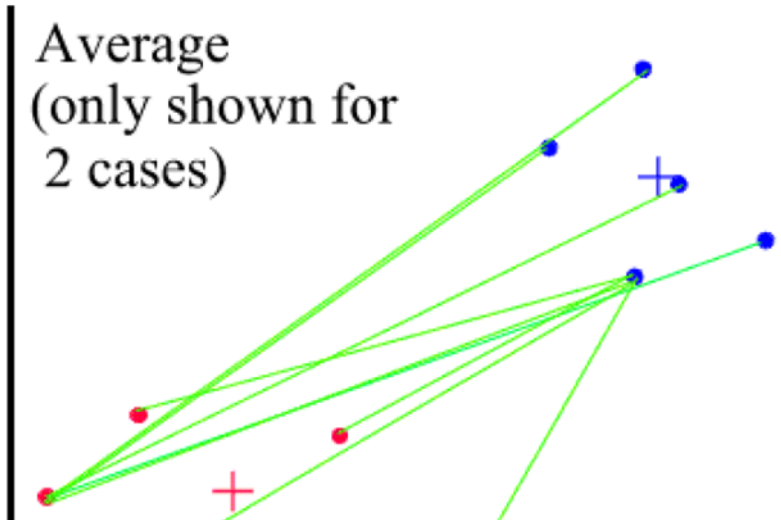
Distance between 2 clusters = **greatest** **distance** between members of the two clusters



$$\Delta\left(\mathbf{C}_{\alpha},\mathbf{C}_{\beta}\right)=\max_{\mathbf{x}\in\mathbf{C}_{\alpha},\mathbf{y}\in\mathbf{C}_{\beta}}\left\{\Delta\left(\mathbf{x},\mathbf{y}\right)\right\}$$

Distances Between Clusters: ‘average linkage’

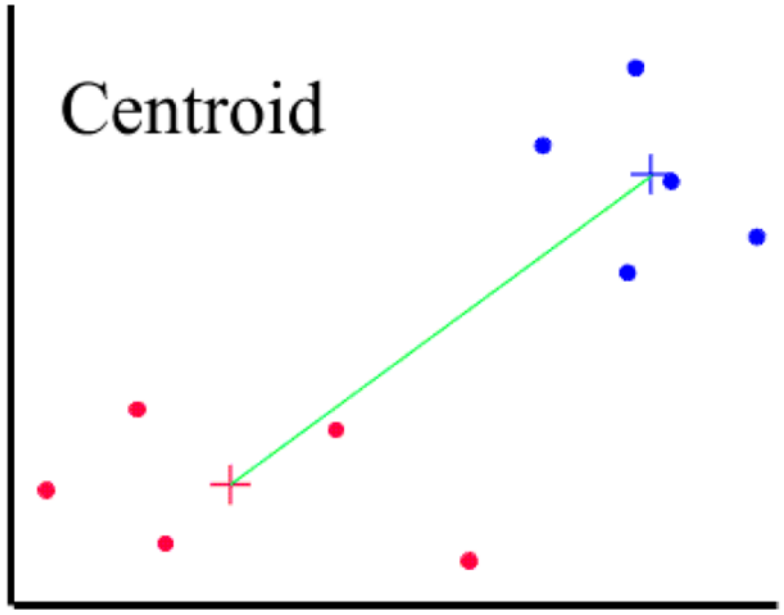
Distance between 2 clusters = **average** of all distances between members of the two clusters



$$\Delta\left(\mathbf{C}_\alpha,\mathbf{C}_\beta\right)=\frac{1}{\left|\mathbf{C}_\alpha\right|\left|\mathbf{C}_\beta\right|}\sum_{\mathbf{x}\in\mathbf{C}_\alpha}\sum_{\mathbf{y}\in\mathbf{C}_\beta}\Delta\left(\mathbf{x},\mathbf{y}\right)$$

Distances Between Clusters: ‘centroid linkage’

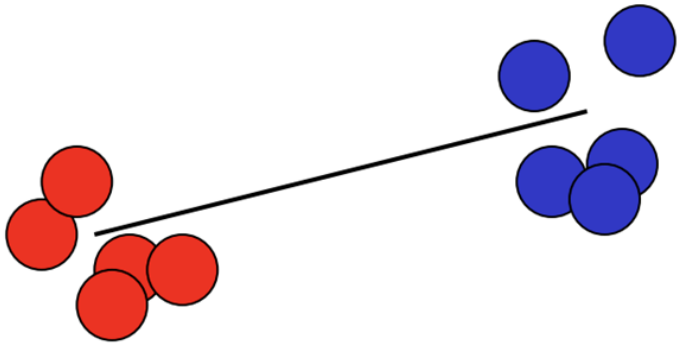
Distance between 2 clusters =
distance between their **centroids** (centers)



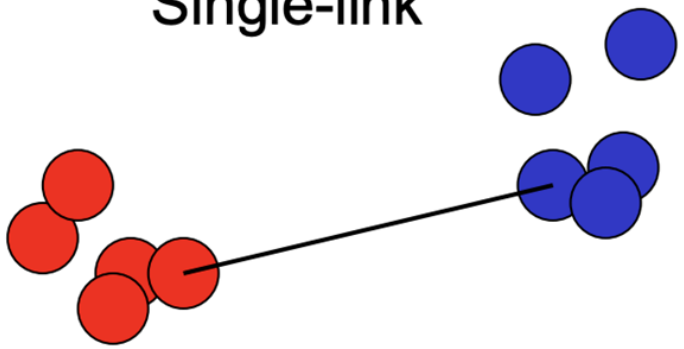
$$\Delta\left(\mathbf{C}_\alpha, \mathbf{C}_\beta\right) = \left\|\mathbf{m}_\alpha - \mathbf{m}_\beta\right\|$$

All linkage methods

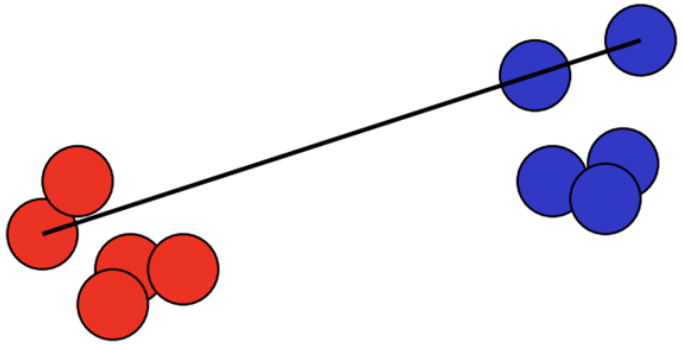
Distance between centroids



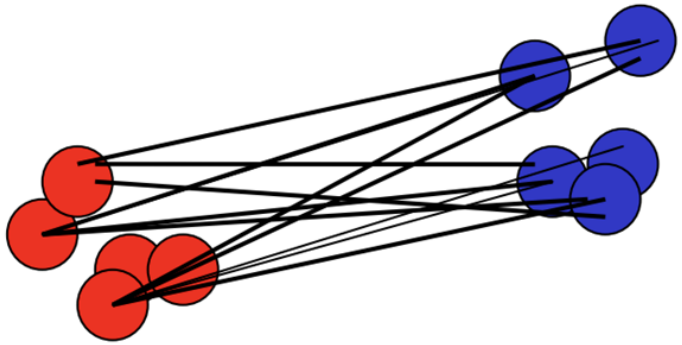
Single-link



Complete-link



Mean-link



Hierarchical clustering Technique:

Hierarchical clustering is one of the popular and easy to understand clustering technique. This clustering technique is divided into two types:

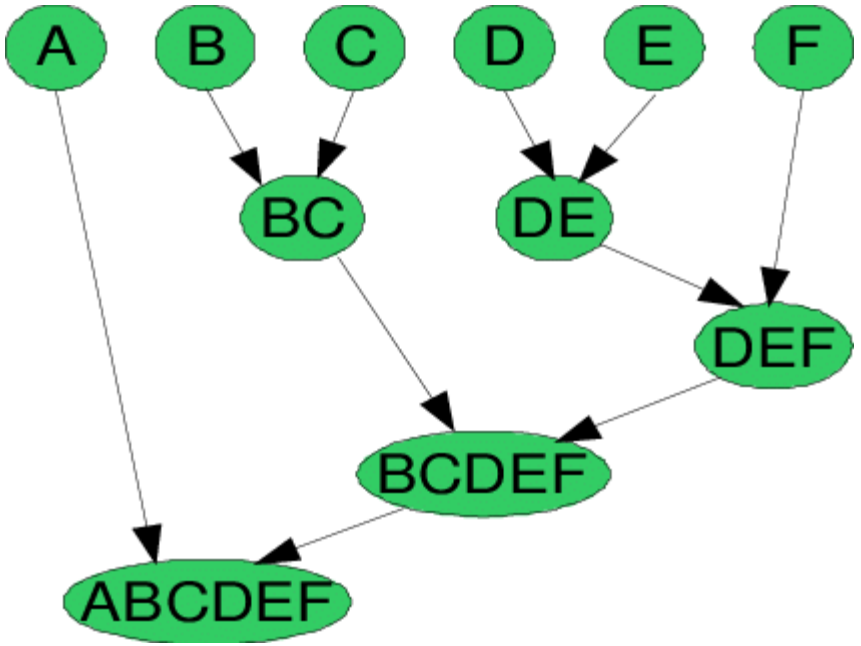
1.Agglomerative

2.Divisive

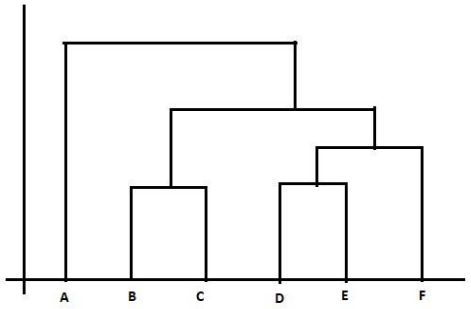
Agglomerative Hierarchical clustering Technique:

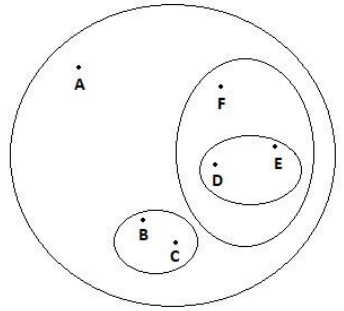
initially each data point is considered as an individual cluster. At each iteration, the similar clusters merge with other clusters until one cluster or K clusters are formed.

Key operation is the computation of the proximity of two clusters
To understand better let's see a pictorial representation of the Agglomerative Hierarchical clustering Technique. Lets say we have six data points {A,B,C,D,E,F}.



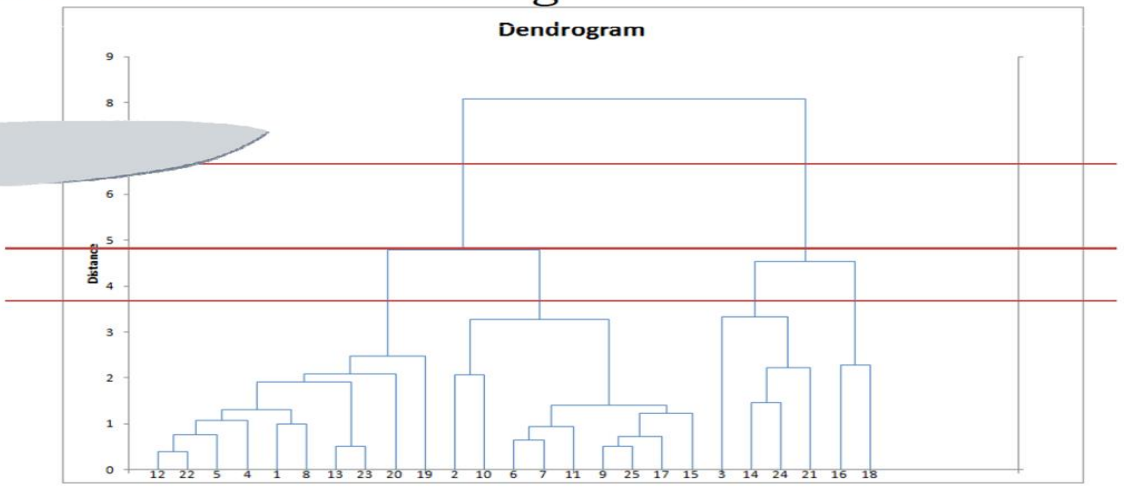
The Hierarchical clustering Technique can be visualized using a **Dendrogram.**
A **Dendrogram** is a tree-like diagram that records the sequences of merges or splits.





From Dendrograms to Clusters

- After dendrogram is obtained, **cut** it to create clusters. **How?**
- Examine *distance levels*
  - Cutpoint determines # clusters
  - Obtain statistics on resulting clusters



2. Divisive Hierarchical clustering Technique:

Since the Divisive Hierarchical clustering Technique is not much used in the real world, I'll give a brief of the Divisive Hierarchical clustering Technique.

In simple words, we can say that the Divisive Hierarchical clustering is exactly the opposite of the **Agglomerative Hierarchical clustering.** In Divisive Hierarchical clustering, we consider all the data points as a single cluster and in each iteration, we separate the data points from the cluster which are not similar. Each data point which is separated is considered as an individual cluster. In the end, we'll be left with n clusters.

As we're dividing the single clusters into n clusters, it is named as **Divisive Hierarchical clustering.**

So, we've discussed the two types of the Hierarchical clustering Technique.