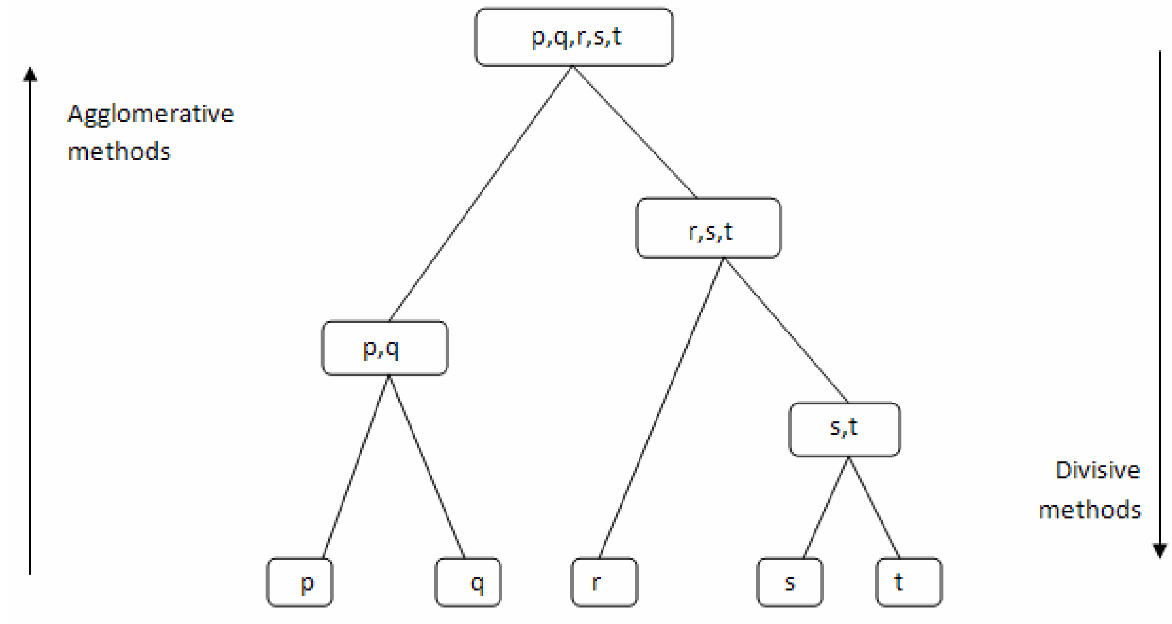
Hierarchical Clustering

Unlike Kmeans, in Hierarchical clustering we don't need to define the number of clusters at the beginning

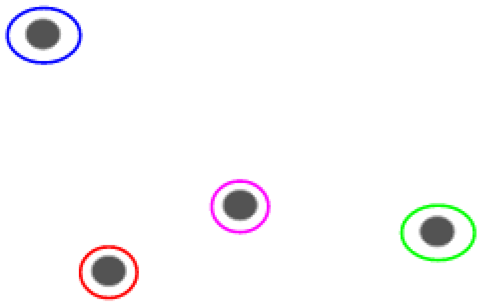
Types



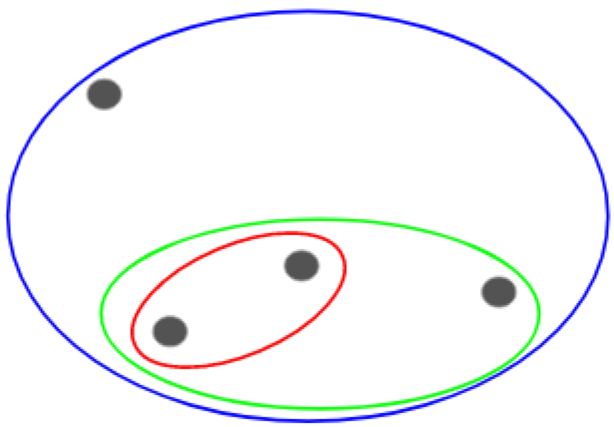
Agglomerative Hierarchical Clustering

The Agglomerative Hierarchical Clustering is the most common type of hierarchical clustering used to group objects in clusters based on their similarity. It's a bottom-up approach where each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

Suppose there are 4 data points. We will assign each of these points to a cluster and hence will have 4 clusters in the beginning.



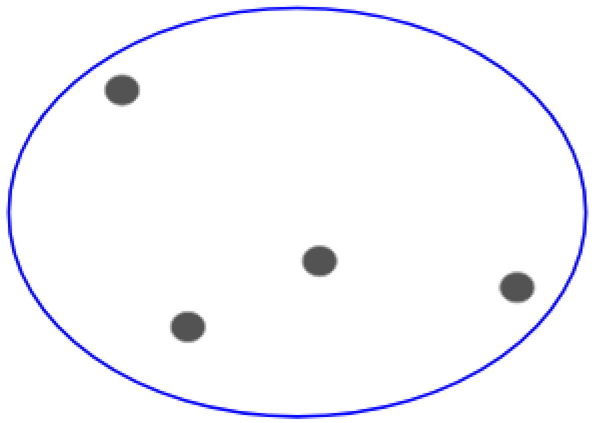
Then, at each iteration, we merge the closest pair of clusters and repeat this step until only a single cluster is left.



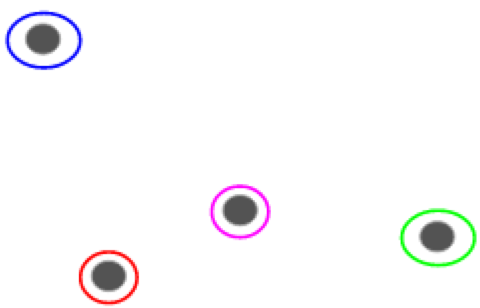
Divisive Hierarchical Clustering

Divisive hierarchical clustering is not used much in solving real-world problems. It works in the opposite way of agglomerative clustering. In this, we start with all the data points as a single cluster. At each iteration, we separate the farthest points or clusters which are not similar until each data point is considered as an individual cluster. Here we are dividing the single clusters into n clusters, therefore the name divisive clustering.

It doesn't matter if we have 10 or 1000 data points. All these points will belong to the same cluster at the beginning:



Now, at each iteration, we split the farthest point in the cluster and repeat this process until each cluster only contains a single point:



Hierachical Clustering use two important parameters

- Measure of distance (similarity)

Similarity can be calculated using the below metrics

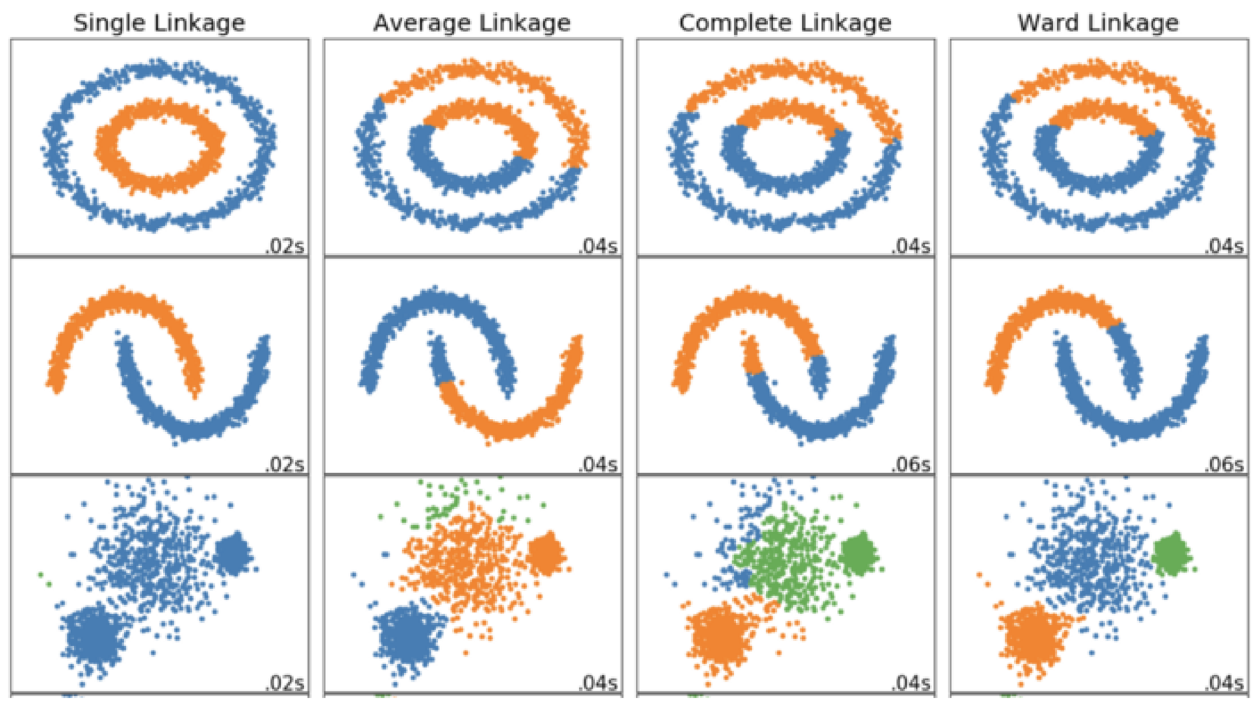
Hamming Distance

Manhattan Distance (Taxicab or City Block)

Minkowski Distance

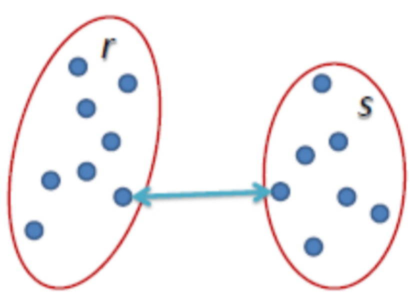
- Linkage Criteria

After selecting a distance metric, it is necessary to determine from where distance is computed. The linkage criteria refer to how the distance between clusters is calculated.



Single Linkage

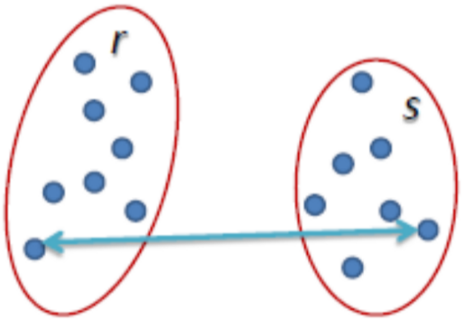
The distance between two clusters is the shortest distance between two points in each cluster



$$L(r,s) = \min(D(x_{ri}, x_{sj}))$$

Complete Linkage

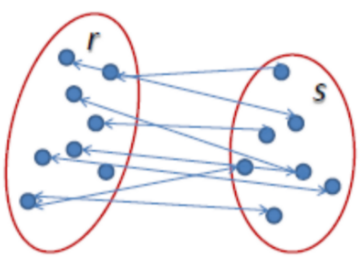
The distance between two clusters is the longest distance between two points in each cluster



$$L(r,s) = \max(D(x_{ri}, x_{sj}))$$

Average Linkage

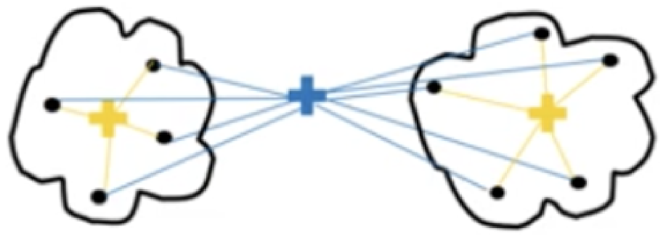
The distance between clusters is the average distance between each point in one cluster to every point in other cluster



$$L(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} D(x_{ri}, x_{sj})$$

Ward Linkage

The distance between clusters is the sum of squared differences within all clusters



How does it work?

Agglomerative hierarchical clustering is more used than divisive hierarchical clustering. Below is the working of Agglomerative hierarchical clustering

- First, we assign all the points to an individual cluster

- Next, we will look at the smallest distance in the proximity matrix and merge the points with the smallest distance. We then update the proximity matrix

ID	1	2	3	4	5
1	0	3	18	10	25
2	3	0	21	13	28
3	18	21	0	8	7
4	10	13	8	0	15
5	25	28	7	15	0

ID	(1,2)	3	4	5
(1,2)	0	18	10	25
3	18	0	8	7
4	10	8	0	15
5	25	7	15	0

- We will repeat step 2 until only a single cluster is left.

How to choose the number of clusters?

To choose the number of clusters in hierarchical clustering, we make use of concept called dendrogram.

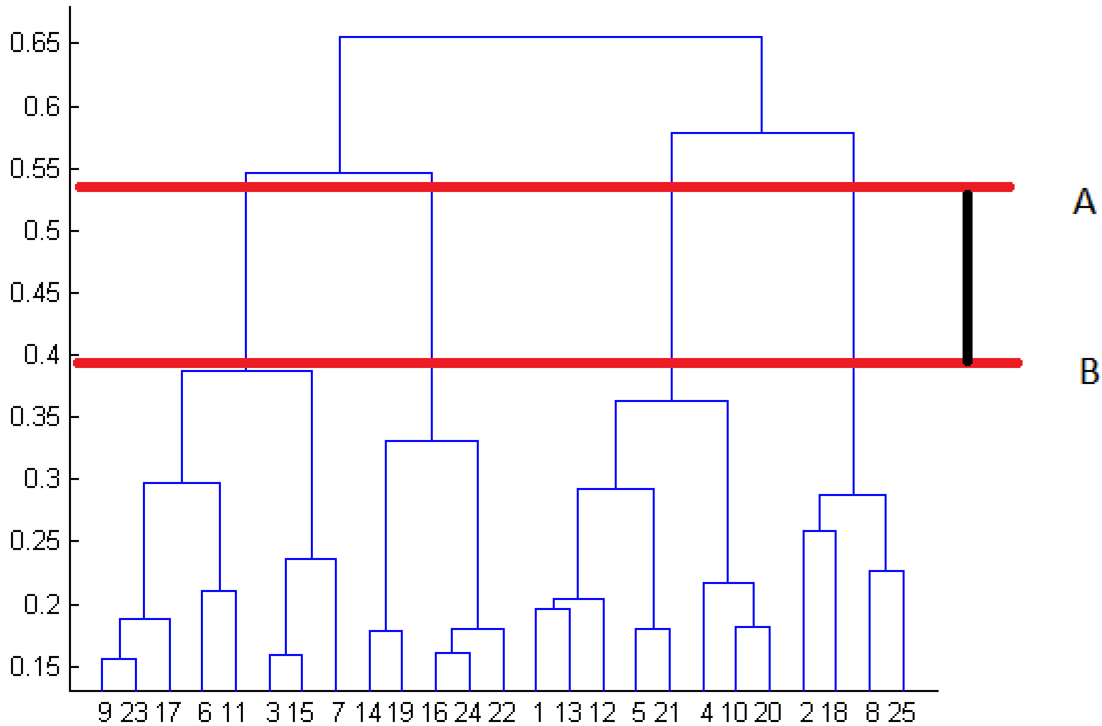
We can use a dendrogram to visualize the history of groupings and figure out the optimal number of clusters. Dendrogram is a tree like diagram that shows the hierarchical relationship between the observations.

- We look for the largest distance that we can vertically without crossing any horizontal line

- Draw a horizontal line at both extremities

- The optimal number of clusters is equal to the number of vertical lines going through the horizontal line

In the below case, best choice for no. of clusters will be 4.



The important point to note while reading dendrogram is that:

- Height of the blocks represents the distance between clusters, and

- Distance between observations represents dissimilarities.