HIERARCHICAL CLUSTERING

→ Hierarchical clustering is a hard clustering method unlike K-means (which is a Flat Clustering method), we don't pick a set number of clusters but rather we arrange the data in a hierarchy where on top of the hierarchy we have a single big cluster while at the bottom of the hierarchy as many clusters as many observations the dataset have.

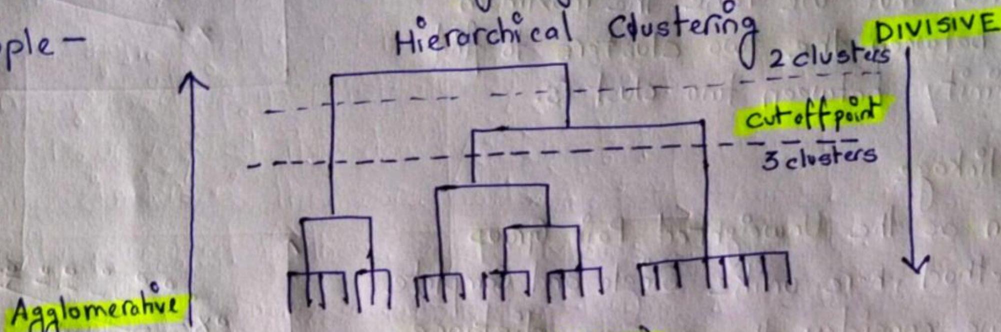
Agglomerative vs Divisive Clustering →

→ 2 most using methods of hierarchical clustering are:

I) Agglomerative Hierarchical Clustering Algorithm (AGNES)

II) Divisive Hierarchical Clustering Algorithm (DIANA)

Example -



Agglomerative Hierarchical Clustering (AGNES) -

It is a bottom up approach where we initially assign different clusters to each observations and then on the basis of similarity we consolidate/merge clusters until we are left with one single cluster.

Divisive Hierarchical Clustering Algorithm aka Divisive Analysis Clustering (DIANA)

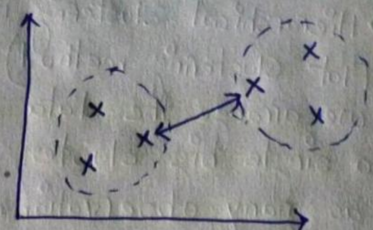
The opposite of Agglomerative method is the Divisive method which is a top-down method where initially we consider all the observations as a single cluster. We then divide this one big single cluster into smaller cluster. The cluster can be divided continuously until we have one cluster for each observation.

→ In both the methods we use a threshold which provide us with a number of clusters. If our ~~clust~~ threshold is high, we get closer, more generalised clusters. If our threshold is low, we can get a lot of clusters which are too fine grained to be of any meaning.

Type of Linkages → In both the method we either require to find the similarity (agglomerative method) or dissimilarity (divisive method) among the clusters. This is done by calculating the distance between clusters. There are multiple way of calculating the distance such as Single linkage, Complete linkage and Average linkage.

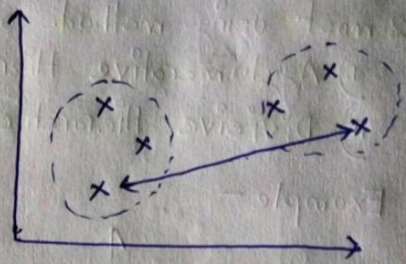
Single linkage (Nearest Neighbour)

When we perform clustering using single linkage we find proximity between two clusters by calculating the shortest distance between them. Here we consider the two closest data points of the two clusters to calculate the distance.



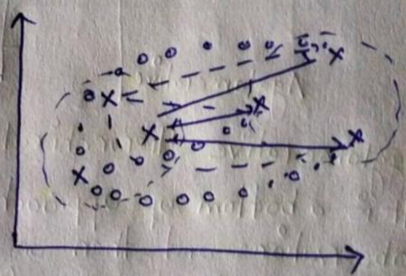
Complete linkage (Farthest Neighbour)

It is the opposite of Single linkage where we consider the two farthest point of both the two cluster. Thus we take the maximum distance between the two clusters to find the proximity between two clusters.



Average linkage

Also known as the Unweighted Pair Group Mean method, here unlike Single and Complete linkage, we consider the average distance. For this, we calculate the average distance from each data point of a cluster to all the datapoints of the other cluster.



Overview →

Linkage	Description
Single	Minimal intercluster dissimilarity. Compute all pairwise dissimilarity between observation in cluster A and cluster B and record the smallest of these dissimilarities.
Complete	Maximal intercluster dissimilarity. Compute all pairwise dissimilarities between the observation in cluster A and the observation in cluster B and record the largest of these dissimilarities
Average	Mean intercluster dissimilarity. compute all pairwise dissimilarities between the observations in cluster A and the observation in cluster B and record the average of these dissimilarities.

Agglomerative Hierarchical Clustering → Using single linkage we will calculate distance between clusters.

For example, Consider a dataset with two feature X and Y.

Data points	P1	D2	D3	D4	D5	D6	D7
X	0.7	2.45	3.47	5.23	5.98	7.778	8.97
Y	3.2	2.89	1.12	5.24	6.23	8.97	6.12

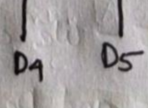
we will start with as many clusters as many observations, in our case 7 clusters. Now our goal is to perpetually consolidate the data points to form clusters until we are left with single cluster.

Step 1 – First create what is known as distance matrix. Here we compute the distance from each observations of the dataset. Distance metric using in this example is euclidean distance.

formula of euclidean distance = $\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$

	D_1	D2	D3	D4	D5	D6	D7
D_1	0						
D_2	1.78	0					
D_3	3.46	2.04	0				
D_4	4.97	3.64	4.48	0			
D_5	6.09	4.86	5.69	1.24	0		
D_6	8.92	7.83	8.66	4.24	3.00	0	
D_7	8.77	7.28	7.43	3.84	2.99	2.78	0

Step 2 – Agglomerative clustering is a bottom up approach therefore we merge observations together based on their similarity (minimum distance). So minimum value is 1.24 which is distance between D4 and D5. Thus we merge the two datapoints to form a cluster.



Step 3 – This step we introduce the linkage method. Here we need to update the distance matrix by recalculating the distance with respect to newly formed cluster.

for example, we have to calculate the distance D_1 to the cluster D4, D5. If we use single linkage method, we look at the distance between D1 & D4 and D1 & D5 and select minimum distance as the distance between D1 and D4, D5.

From above distance matrix, we can observe D1 have two options: 4.97 (Distance from D1 to D4) and 6.09 (Distance from D1 to D5).

As we are using single linkage, we choose minimum distance between, so we choose 4.97 and consider it as the distance between D1 and D4,D5. If we are using complete linkage than maximum value would be selected as distance between D1 and D4, D5. which would be 6.09. If we use average linkage then the average of these two distances would have been taken. Thus the distance between D1 and D4, D5 would have come out to be 5.53=(4.97+6.09)/2.

In this example, we are creating clusters using single linkage method.

Update distance matrix,

	D1	D2	D3	D4, D5	D6	D7
D1	0					
D2	1.78	0				
D3	3.46	2.04	0			
D4, D5	4.97	3.64	4.48	0		
D6	8.92	7.83	8.66	3.00	0	
D7	8.77	7.28	7.43	2.99	2.78	0

Step 4 – From now on we will simply repeat step 2 and step 3 until we are left with one cluster. We look again, minimum value is 1.78 indicating new cluster can be formed between D1 and D2.

So, final distance matrix will looks like

	D1, D2, D3.
D1, D2, D3	0
D4, D5, D6, D7	3.64.

So, this is the final cluster.

Step 5 – We can get the optimize clusters through Dendogram.

Merits and Demerits of hierarchical Clustering →

→ It implements without requiring a predetermined number of clusters.

→ It is sensitive to noise/outliers.

→ It require standardization of data as distance metrics such as euclidean distance is used which require data to be on same scale.

→ Sometimes it become difficult to identify right number of cluster in dendogram.

→ Methode like using cohension which indicates the goodness of the clusters.