

Steps to Perform Hierarchical Clustering

Step 1: First, we assign all the points to an individual cluster:



Different colors here represent different clusters. You can see that we have 5 different clusters for the 5 points in our data.

Step 2: 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

Here, the smallest distance is 3 and hence we will merge point 1 and 2:



Let's look at the updated clusters and accordingly update the proximity matrix:

Student_ID	Marks
(1,2)	10
3	28
4	20
5	35

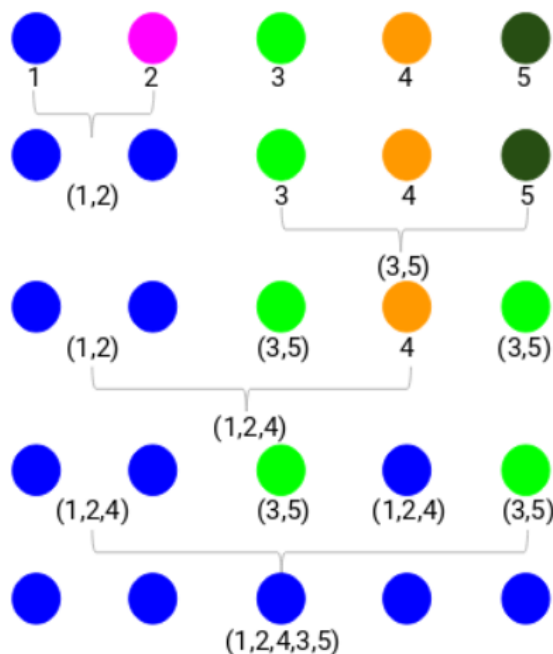
Here, we have taken the maximum of the two marks (7, 10) to replace the marks for this cluster. Instead of the maximum, we can also take the minimum value or the average values as well. Now, we will again calculate the proximity matrix for these clusters:

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

Step 3: We will repeat step 2 until only a single cluster is left.

So, we will first look at the minimum distance in the proximity matrix and then merge the closest pair of clusters. We will get the merged clusters as shown below after repeating these steps:

So, we will first look at the minimum distance in the proximity matrix and then merge the closest pair of clusters. We will get the merged clusters as shown below after repeating these steps:



We started with 5 clusters and finally have a single cluster. **This is how agglomerative hierarchical clustering works.** But the burning question still remains – how do we decide the number of clusters? Let's understand that in the next section.

In []: