19/5/25.

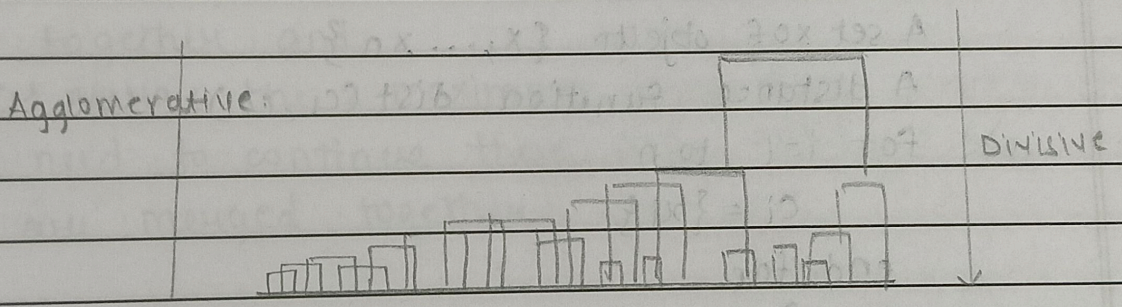
Practical No:-2.

Title / Aim:- Implementing Hierarchical clustering.

Objective:- To study and implement Hierarchical clustering.

Theory:-

Hierarchical clustering involves creating clusters that have a predetermined ordering from top to bottom. for eg:-, all files & folders on the hard disk are organized in a hierarchy. There are 2 types of hierarchical clustering. Divisive and Agglomerative.




Divisive Method.

- In divisive or top-down clustering method we assign all the observations to a single cluster & then partition the cluster to 2 least similar clusters using a flat clustering method (eg:- k-Means). Finally, we proceed recursively on each cluster until there is one cluster for each observation. There is evidence that divisive algorithms produce more accurate

hierarchies than agglomerative algorithms in circumstances but is conceptually more complex

* Agglomerative method.

- In agglomerative or bottom-up clustering method we assign each observation to its own cluster. Then, compute the similarity (eg:- distance) between each of the clusters and join the two most similar clusters.

Finally, repeat steps 2 and 3 until there is only a single cluster left. The related algorithm is shown below.

Given:-

A set x of objects $\{x_1, \ldots x_n\}$

A distance function dist(c1, c2)

for i=1 to n

$c_i = \{x_i\}$

end for

$c = \{c_1, \ldots, c_n\}$

while c.size > 1 do

- $(c_{min1}, c_{min2})$ = minimum dist($c_i, c_j$) for all $c_i, c_j$ in c
- remove $c_{min1}$ and $c_{min2}$ from c
- add $\{c_{min1}, c_{min2}\}$ to c
- l = l + 1

end while.

- As we have discussed above, firstly, the data P2 and P3 combine together and form a cluster correspondingly a dendogram is created which connects P2 & P3 with a rectangular shape. The height is decided according to the Euclidean distance between the data points.
- In the next step P5 & P6 form a cluster, and the corresponding dendogram is created. It is higher than of previous as the Euclidean distance between P5 & P6 is a little bit greater than P2 & P3.
- Again, two new dendograms are created that combines P1, P2 & P3 in one dendrogram & P4, P5 & P6 in another dendrogram.
- At last, the final dendrogram is created that combines all the datapoints together.
- We can cut the dendrogram tree structure at any level as per our requirement.

Conclusion:-

Hierarchical clustering effectively groups similar data points without requiring a predefined number of clusters. It provides a clear tree like structure (dendrogram) for better data interpretation and analysis.