# Hierarchical Clustering

Hierarchical clustering is an unsupervised machine learning algorithm used to group data points into clusters based on similarity. It seeks to build a hierarchy of clusters. Hierarchical clustering does not require us to pre-specify the number of clusters. There are two main types of hierarchical clustering—

(1) **Agglomerative Hierarchical Clustering (Bottom-up approach)**

Agglomerative clustering is a bottom-up approach where each data point starts as its own cluster, and pairs of clusters are merged step by step based on distance metric until all points are merged into one cluster.

Steps Involved:

(a) Assign each point to its own cluster.
(b) Calculate the distance between all pairs of clusters.
(c) Merge the two closest clusters.
(d) Recompute the distances between the remaining clusters.
(e) Repeat steps (c) & (d) until all points are merged into a single cluster.

(2) **Divisive Hierarchical Clustering (Top-Down approach)**

Divisive clustering is a top-down approach where all data points are initially part of a single cluster. Then, the cluster is recursively split into smaller clusters based on some distance metric, until each data point forms its own cluster.

Steps Involved:

(a) Assign all data points to a single cluster.
(b) Split the cluster into two based on distance metric.
(c) Repeat the process for each cluster until all clusters contain only one data point. ~~or~~

* **Linkage Criteria:** To decide how to compute distances between two clusters, different linkage criteria can be used. These define how the distance between clusters is calculated, influencing the final shape of

hierarchy. The main types are –

(A) Single Linkage : The minimum distance between points in two ~~each~~ cluster.

(B) Complete Linkage : The maximum distance between points in two clusters.

(C) Average Linkage : The average distance between all points in two clusters.

(D) Centroid Linkage : The distance between the centroids (mean points) of two clusters.

(E) Ward's Method : A method that minimizes the variance of the clusters being merged, leading to a compact and spherical cluster. This is the default method in scikit-learn.

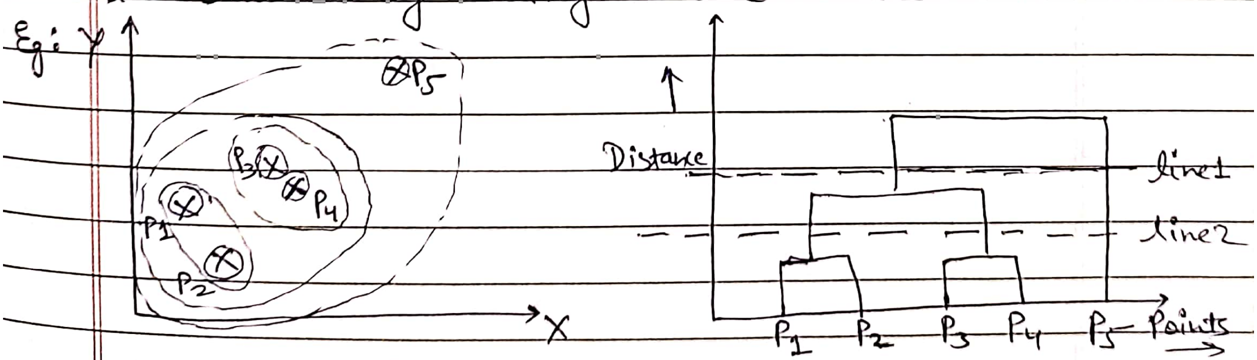
** **Dendogram :** The result of hierarchical clustering can be ~~defi~~ visualized using a dendogram, a tree-like diagram that records the sequence of cluster merges or splits. The vertical axis represents the distance or dissimilarity between clusters, while the horizontal axis shows the data points.

In the dendogram :

→ The height of the merge represents the distance between the clusters.

→ A cut-off threshold can be applied to divide the dendogram into clusters by cutting the tree at a certain level.

Eg:




In this example, we are taking 5 random points from $P_1$ to $P_5$ and using agglomerative hierarchical clustering and then drawing its dendogram.

If our threshold is such that it cuts line 1, then we will have 2 clusters ($P_1, P_2, P_3, P_4$) and $P_5$, if we cut at line 2, we will have 3 clusters [($P_1, P_2$), ($P_3, P_4$) & $P_5$]

classmate
Date
Page 03

## Advantages of Hierarchical Clustering:

(a) Unlike K-Means, we don't have to pre-define the number of clusters to be assigned.

(b) Dendogram provides a visual representation of the data's hierarchical structure.

(c) It works well with clusters of different shapes and sizes and densities without assuming spherical cluster.

## Disadvantages:

(a) It is computationally expensive (usually $O(n^3)$) and doesn't scale well for large datasets.

(b) Once a merge or split is made, it can't be undone.

(c) Noise and outliers can affect the structure of hierarchy significantly.