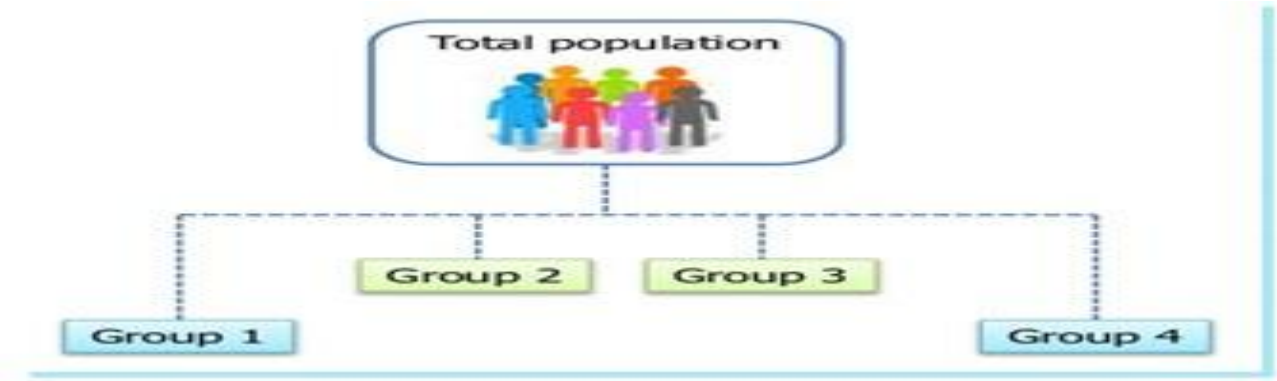
Hierarchical Clustering Algorithm

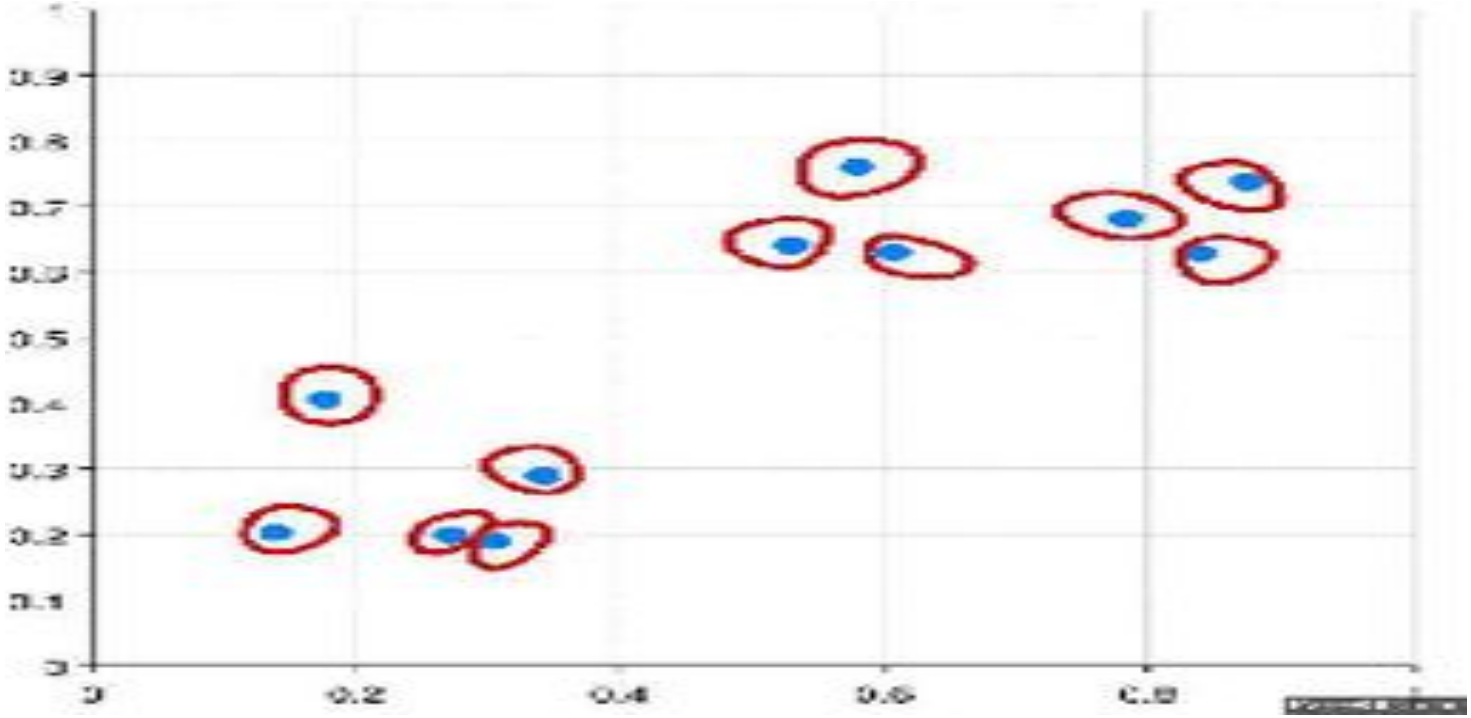


- HCA is an unsupervised clustering algorithm which involves creating clusters that have predominant ordering
- The algorithm groups similar objects into groups called clusters. The endpoint is a set of clusters or groups, where each cluster is distinct from each other cluster, and the objects within each cluster are broadly similar to each other.
- For example, items shown in the image above should be as similar as possible in terms of attributes of the items in each group, and objects in group 1 and group 2 should be as dissimilar as possible.
- Another Example- All files and folders on our hard disk are organized in a hierarchy.

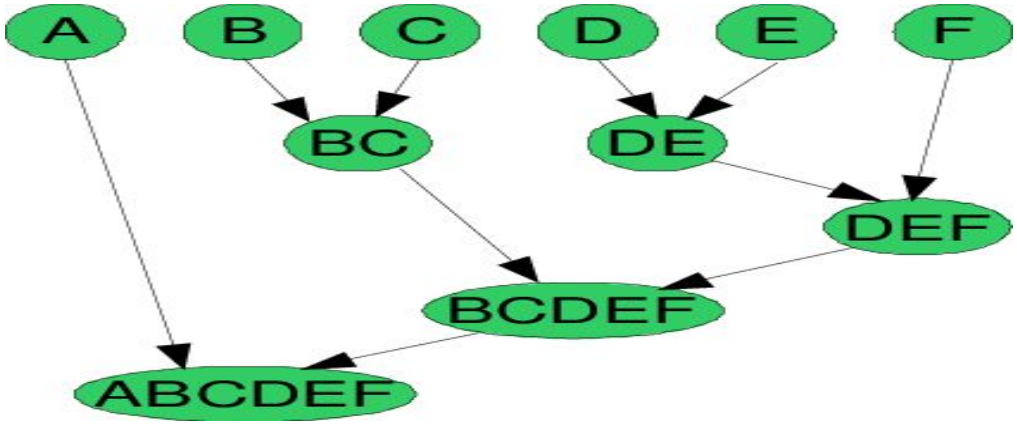
Hierarchical Clustering

Make each data point a single-point cluster → forms N clusters

1. Take the two closest data points and make them one cluster → forms N-1 clusters
2. Take the two closest clusters and make them one cluster → Forms N-2 clusters.
3. Repeat step-3 until you are left with only one cluster.



Hierarchical Clustering Technique



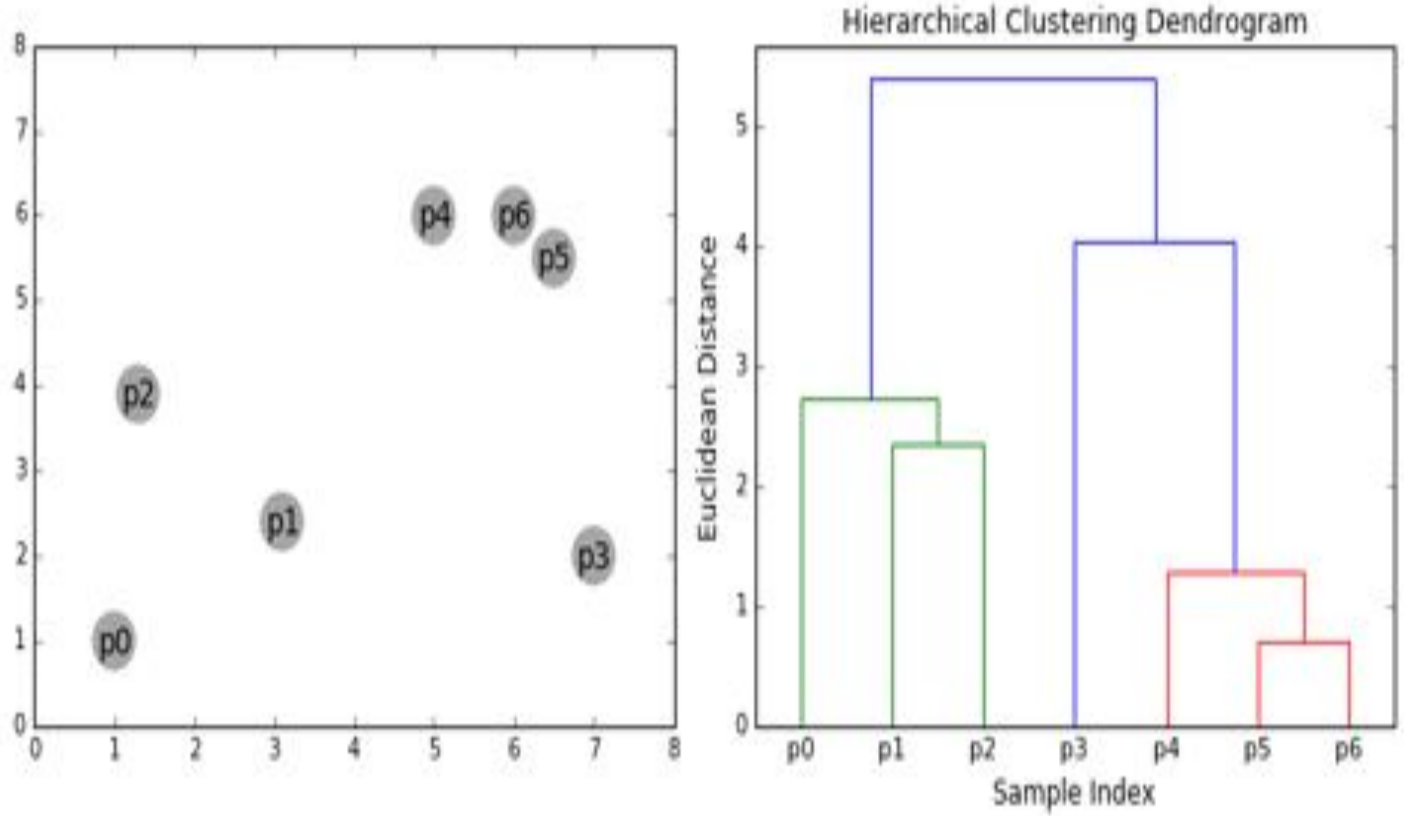
Lets say we have six data points {A,B,C,D,E,F}.

- Step- 1: Calculate the proximity of individual points and consider all the six data points
- Step- 2: similar clusters are merged together and formed as a single cluster. Let's consider B,C, and D,E are similar clusters that are merged in step two. Now, we're left with four clusters which are A, BC, DE, F.
- Step- 3: We again calculate the proximity of new clusters and merge the similar clusters to form new clusters A, BC, DEF.
- Step- 4: Calculate the proximity of the new clusters. The clusters DEF and BC are similar and merged together to form a new cluster. We're now left with two clusters A, BCDEF.
- Step- 5: Finally, all the clusters are merged together

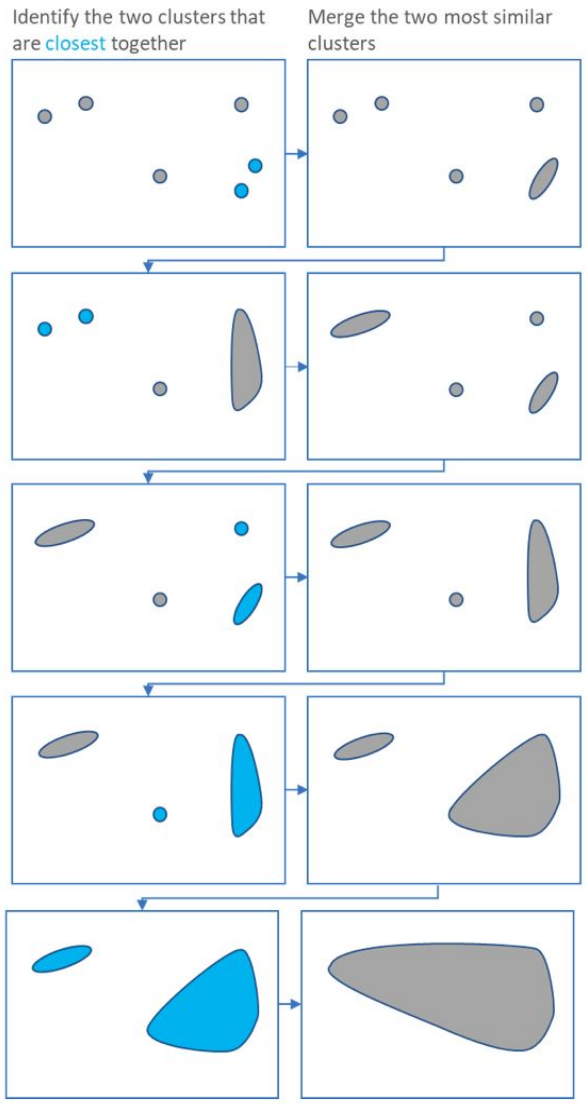
What is a Dendrogram?

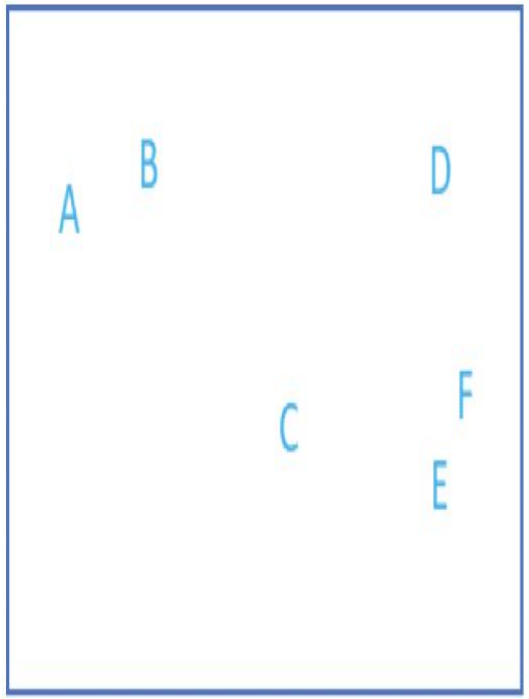
•A Dendrogram is a type of tree diagram showing hierarchical relationships between different sets of data.

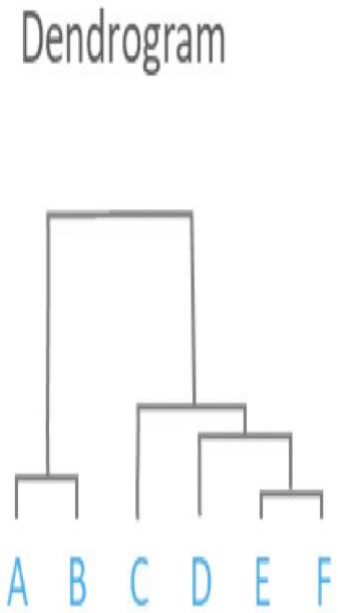
•A Dendrogram contains the memory of hierarchical clustering algorithm, so just by looking at the Dendrogram you can tell how the cluster is formed.



More Example







Industry use Hierarchical Clustering

Business Problem 1: A bank wants to group loan applicants into high/medium/low risk based on attributes such as loan amount, monthly instalments, employment tenure, the number of times the applicant has been delinquent in other payments, annual income, debt to income ratio

- **Business Benefit:** Once the segments are identified, the bank will have a loan applicants' dataset with each applicant labelled as high/medium/low risk.
- Based on these labels, the bank can easily make a decision on
  - Whether to give loan to an applicant
  - how much credit to extend,
  - as well as the interest rate

Business Problem 2: The enterprise wishes to organize customers into groups/segments based on similar traits, product preferences and expectations. Segments are constructed based on customer demographic characteristics, psychographics, past behaviour and product use behaviour.

- **Business Benefit:** Once the segments are identified, marketing messages and products can be customized for each segment.
  - The better the segment(s) chosen for targeting by a particular organization, the more successful the business will be in the market.
  - Hierarchical Clustering can help an enterprise organize data into groups to identify similarities and, equally important, dissimilar groups and characteristics,
  - So that the business can target pricing, products, services, marketing messages and more.

INTERNSHIPSTUDIO

Q.1 Explain Hierarchical Clustering?

Q.2 How the clusters are formed in Hierarchical Clustering?

Q.3 Explain the steps of Hierarchical clustering ?

Q.4 What is a Dendrogram?