Hierarchical learning

By :- Aditya Tyagi

Bhanu Gupta

What is Clustering?

- Clustering is an unsupervised machine learning technique that divides the population into several clusters such that data points in the same cluster are more similar and data points in different clusters are dissimilar.
- Points in the same cluster are closer to each other.
- Points in the different clusters are far apart.

About Hierarchical Clustering

- Hierarchical clustering, also known as hierarchical cluster analysis or HCA, is another unsupervised machine learning approach for grouping unlabeled datasets into clusters.

- The hierarchy of clusters is developed in the form of a tree in this technique, and this tree-shaped structure is known as the dendrogram.

- Simply speaking, Separating data into groups based on some measure of similarity, finding a technique to quantify how they're alike and different, and limiting down the data is what hierarchical clustering is all about.

- Hierarchical clustering method functions in two approaches-
  - Agglomerative
  - Divisive

Types of Hierarchical learning

- Divisive Hierarchical learning
- Agglomerative Hierarchical learning

Divisive Hierarchical learning

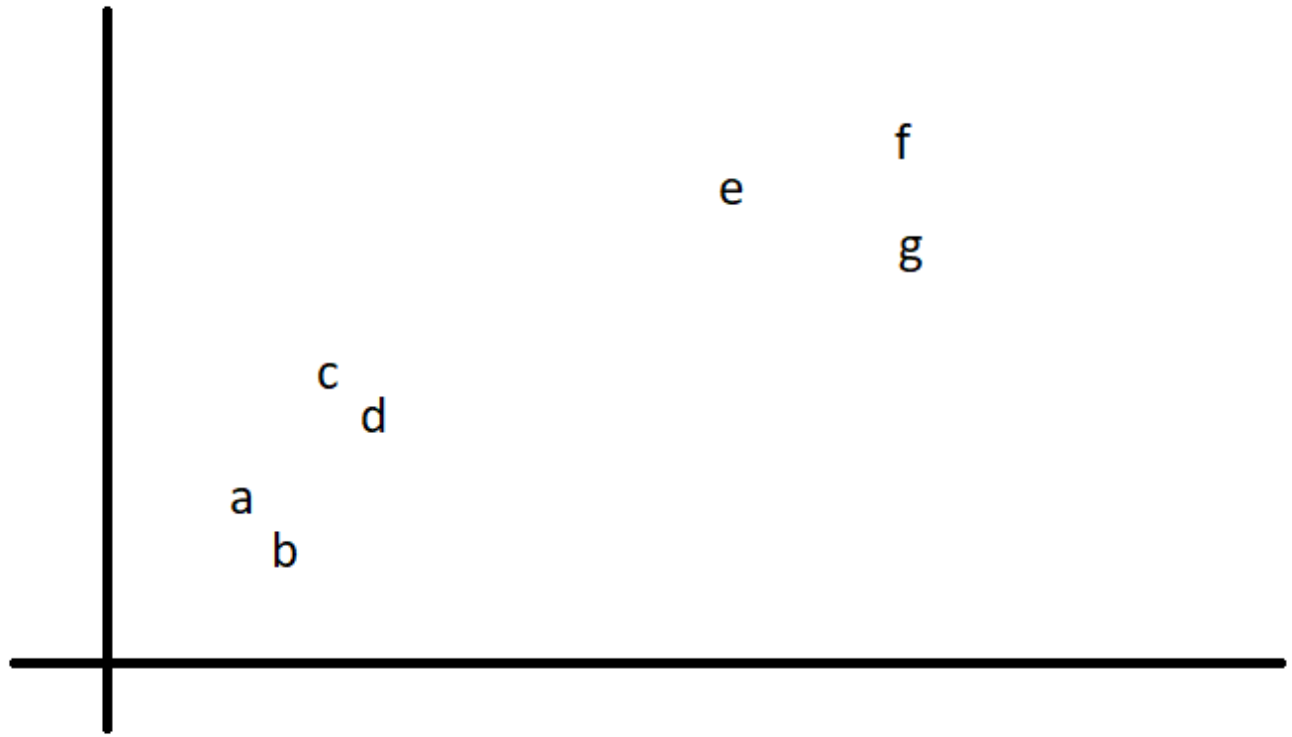
- The divisive clustering algorithm is a top-down clustering approach, initially, all the points in the dataset belong to one cluster and split is performed recursively as one moves down the hierarchy.
- Clustering continues unitll small groups of similar clusters are obtained.

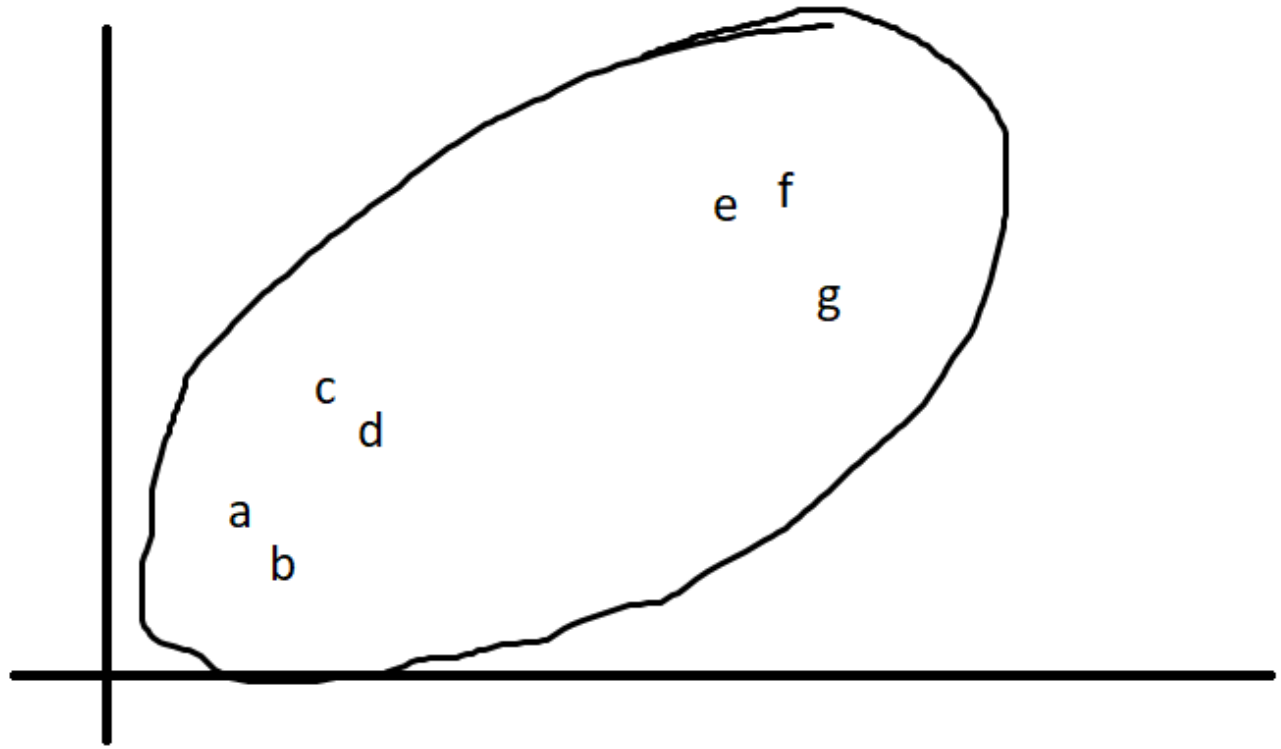
Steps of Divisive Clusterin g:

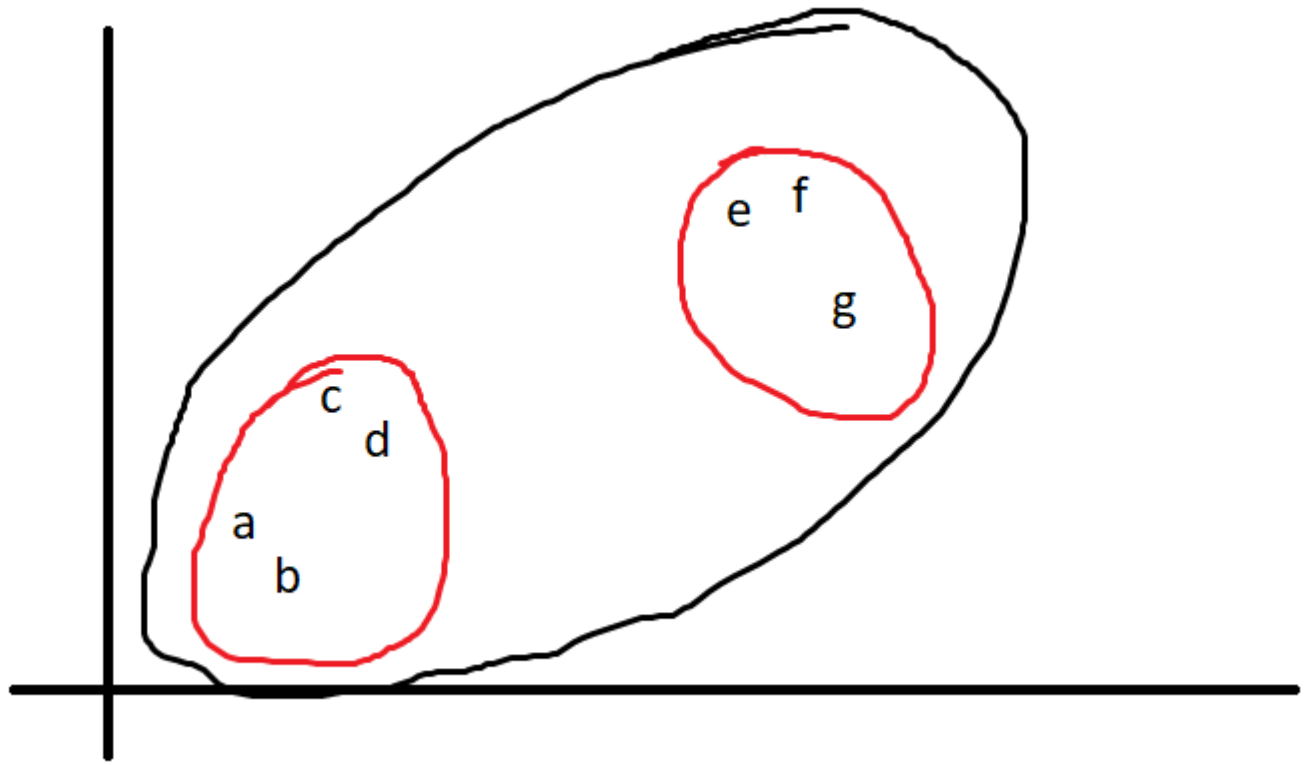
Initially, all points in the dataset belong to one single cluster.

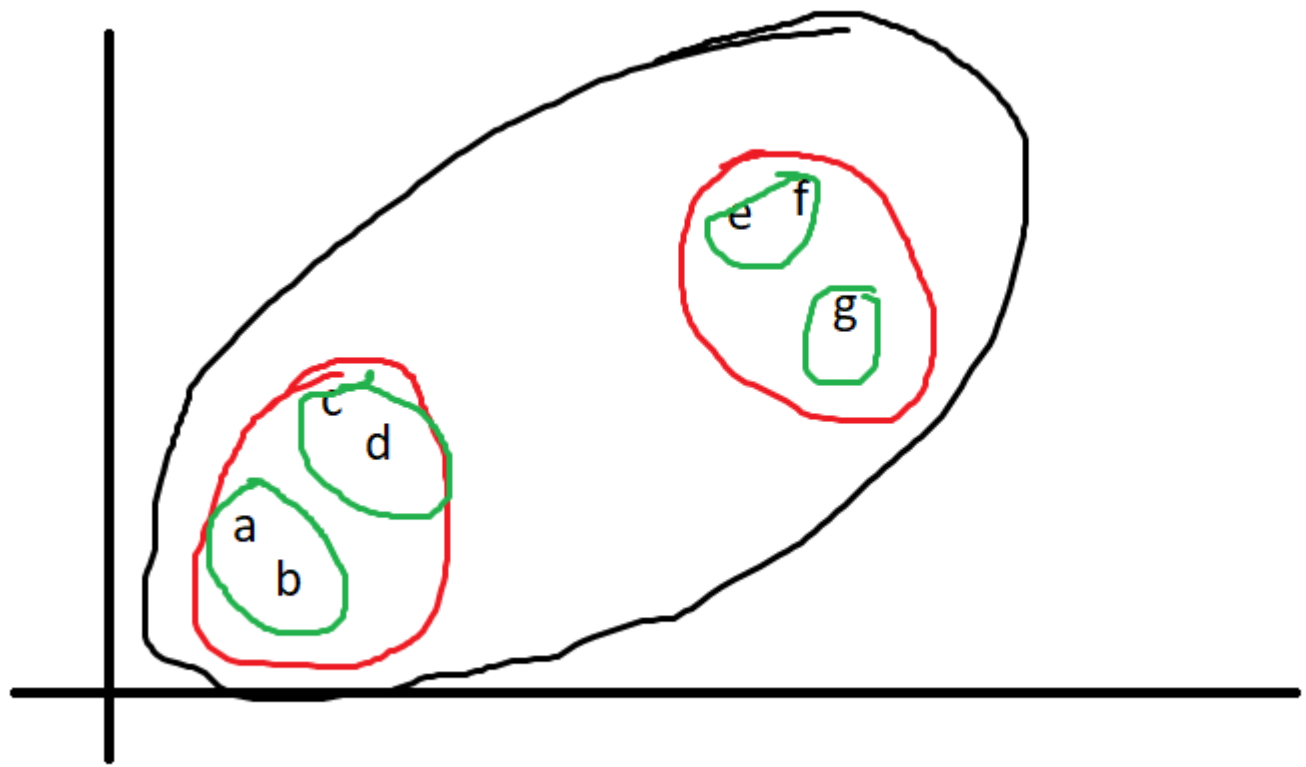
Partition the cluster into two least similar cluster

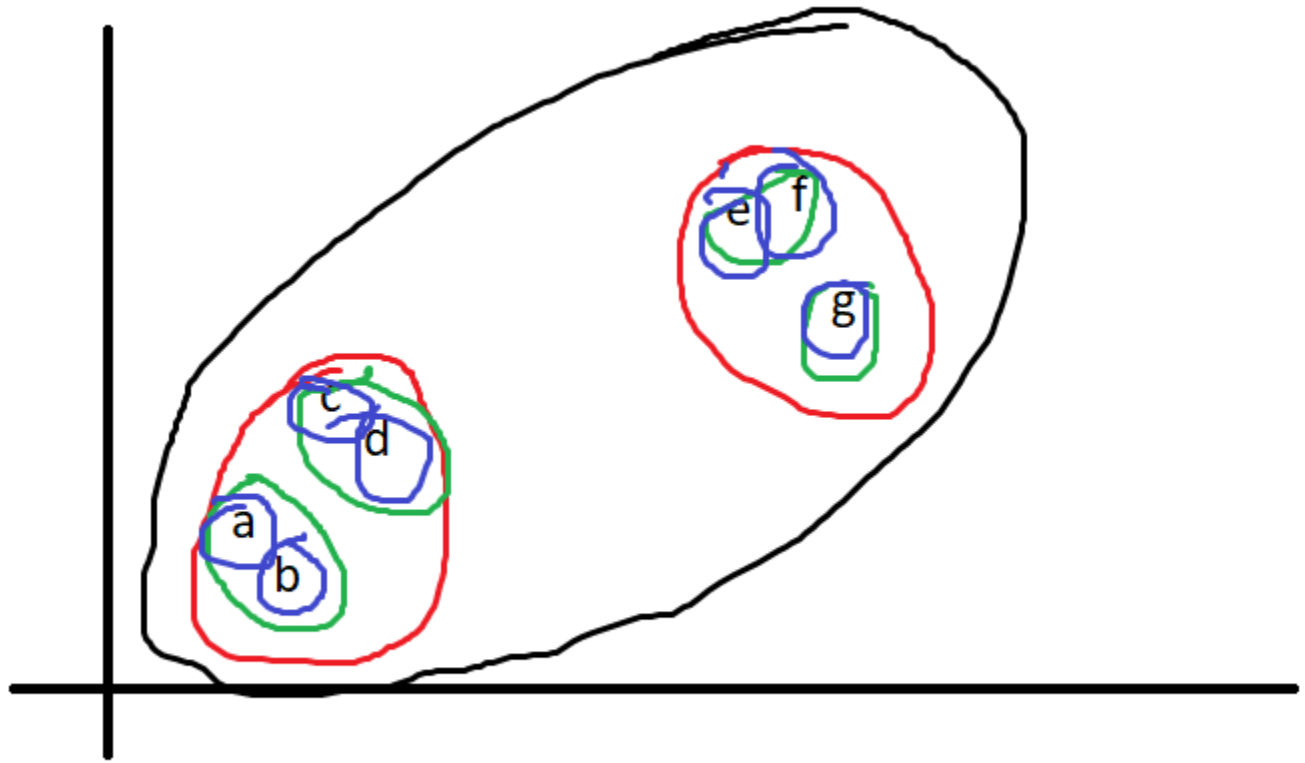
Proceed recursively to form new clusters until the desired number of clusters is obtained.

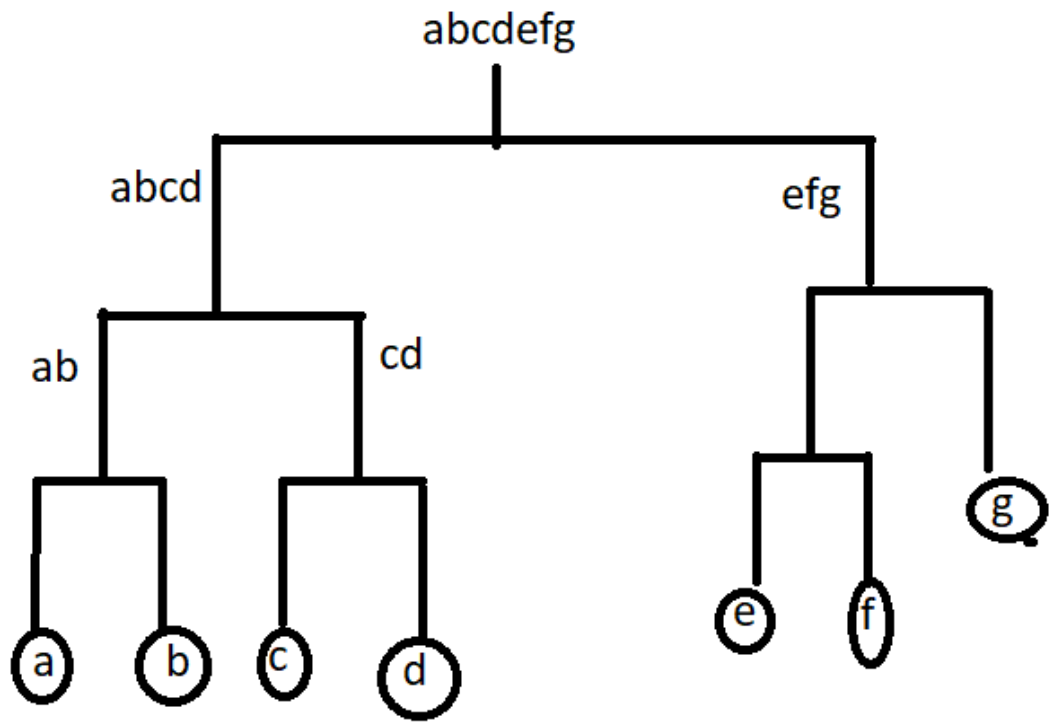

f
e
g
c
d
a
b


e
f
g
c
d
a
b


e
f
g
c
d
a
b


e
f
g
c
d
a
b


a
b
c
d
e
f
g



This representation is called Dendogram

Agglomerative clustering:

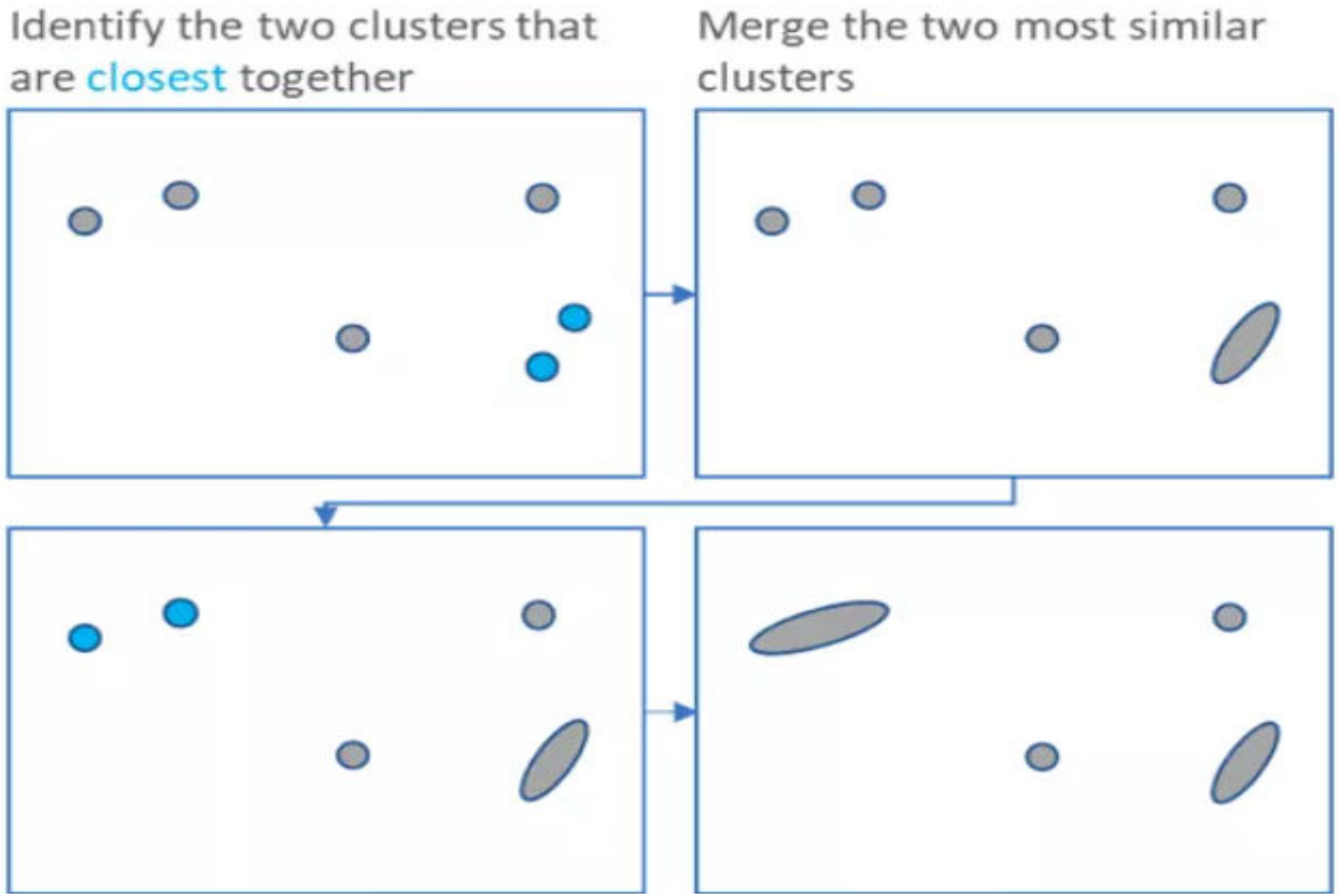
- Agglomerative Clustering is a bottom-up strategy in which each data point is originally a cluster of its own, and as one travels up the hierarchy, more pairs of clusters are combined. In it, two nearest clusters are taken and joined to form one single cluster.

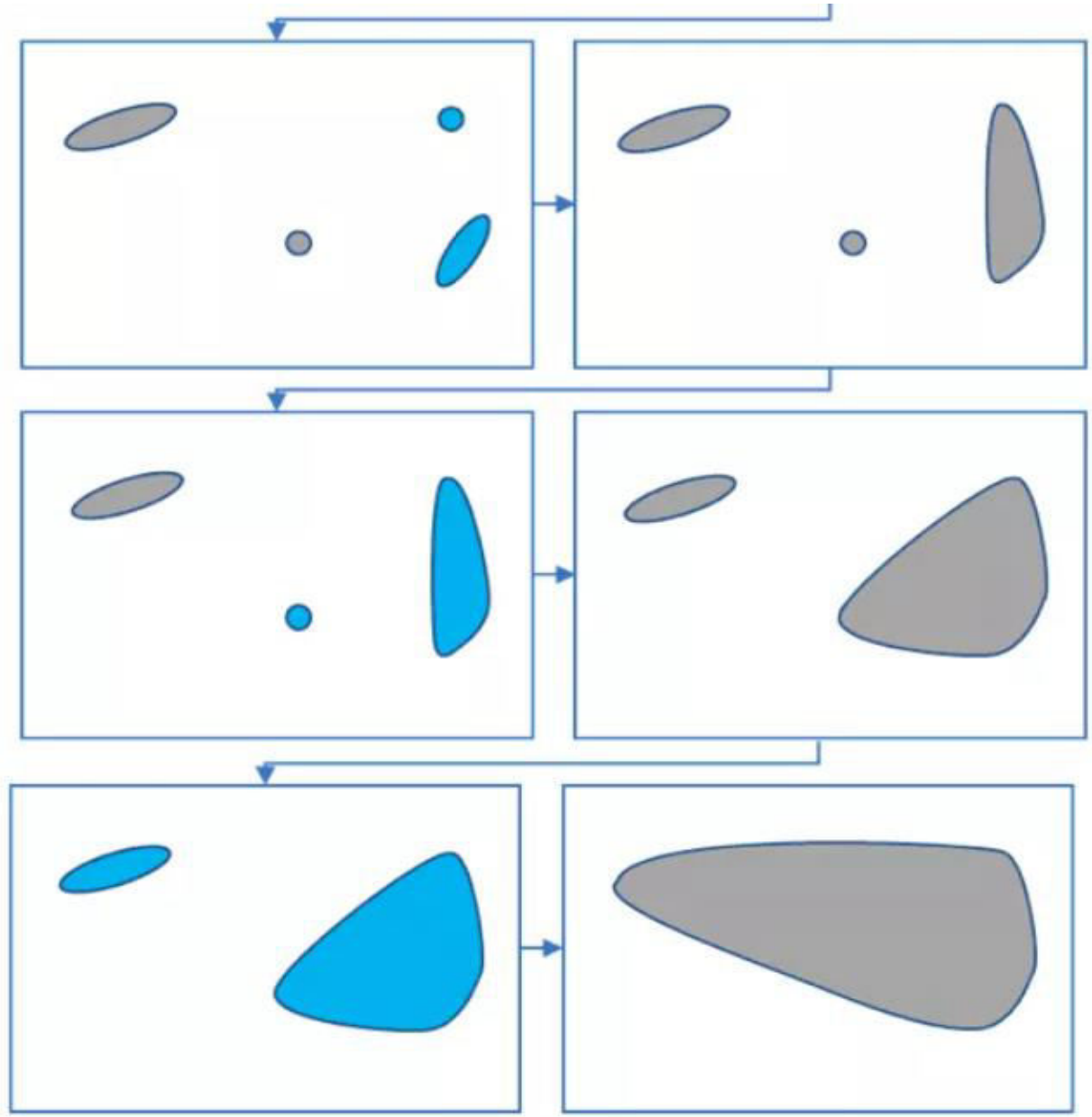
How hierarchical clustering works

Hierarchical clustering starts by treating each observation as a separate cluster. Then, it repeatedly executes the following two steps:

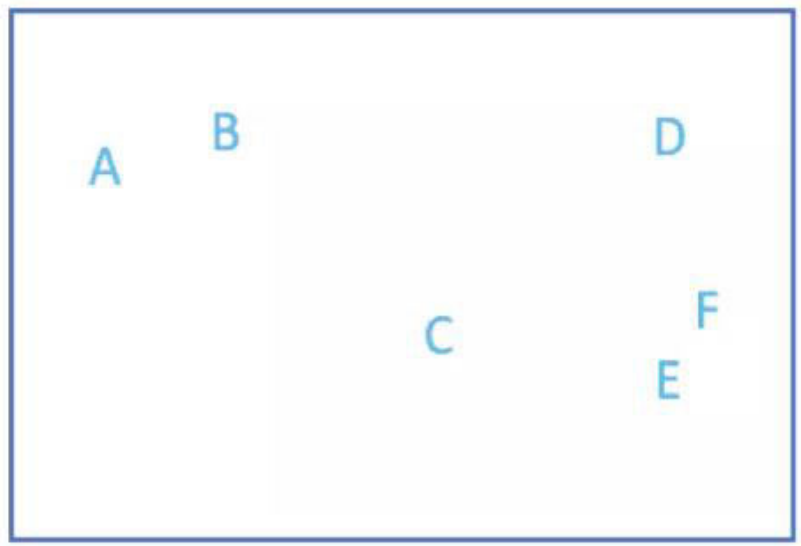
(1) identify the two clusters that are closest together, and

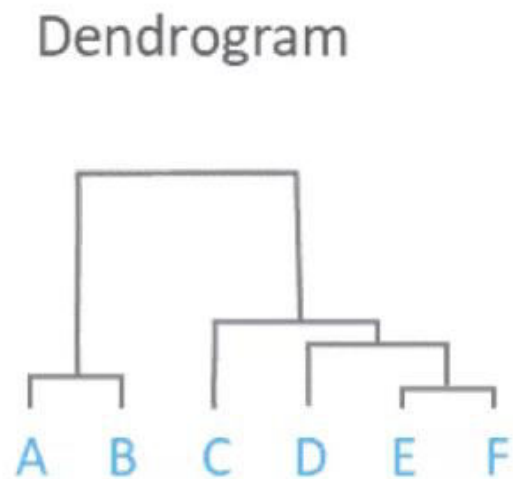
(2) merge the two most similar clusters. This iterative process continues until all the clusters are merged together. This is illustrated in the diagrams below.


Identify the two clusters that are closest together
Merge the two most similar clusters



The main output of Hierarchical Clustering is a *dendrogram*, which shows the hierarchical relationship between the clusters:





Measures of distance (similarity)

- In the example above, the *distance* between two clusters has been computed based on the length of the straight line drawn from one cluster to another. This is commonly referred to as the *Euclidean distance.* Many other *distance metrics* have been developed.
- The choice of distance metric should be made based on theoretical concerns from the domain of study. That is, a distance metric needs to define similarity in a way that is sensible for the field of study. For example, if clustering crime sites in a city, city block distance may be appropriate. Or, better yet, the time taken to travel between each location. Where there is no theoretical justification for an alternative, the Euclidean should generally be preferred, as it is usually the appropriate measure of distance in the physical world.

Linkage Criteria

- After selecting a distance metric, it is necessary to determine from where distance is computed. For example, it can be computed between the two most similar parts of a cluster (*single-linkage*), the two least similar bits of a cluster (*complete-linkage*), the center of the clusters (*mean* or *average-linkage*), or some other criterion. Many linkage criteria have been developed.
- Where there are no clear theoretical justifications for the choice of linkage criteria, *Ward's method* is the sensible default. This method works out which observations to group based on reducing the sum of squared distances of each observation from the average observation in a cluster. This is often appropriate as this concept of distance matches the standard assumptions of how to compute differences between groups in statistics (e.g., *ANOVA*, *MANOVA*).

Agglomerative versus divisive algorithms

- Hierarchical clustering typically works by sequentially merging similar clusters, as shown above. This is known as *agglomerative hierarchical clustering.* In theory, it can also be done by initially grouping all the observations into one cluster, and then successively splitting these clusters. This is known as *divisive hierarchical clustering.* Divisive clustering is rarely done in practice.

Applications of Hierarchical Clustering

- **Fake news is not a new phenomenon, but it is growing more prevalent. Thanks to technological advancements like social media, fake news is being manufactured and circulated at an alarming rate.**

- **Here to tackle this problem, technology or specifically speaking hierarchical clustering is used.**

Identifying criminal activity:

- **Criminal activities when dealt with with technologies provide an effective solution for them. Sometimes, a certain area of province or district seems to be more affected by criminal activities than the other ones.**

- **Here, we can use hierarchical clustering to identify those criminal activities. The system can group similar activities by analyzing GPS data. You may then categorize the groups based on their qualities into those that are genuine and those that are fake.**