

Hierarchical Methods ***(Agglomerative and Divisive)***

- Hierarchical clustering is another unsupervised learning algorithm, which is used to group the unlabeled datasets into a cluster.
- The hierarchical clustering method works by *combining data objects* into a *tree of clusters*.
- The result of hierarchical clustering is a *tree-like structure*, and this tree-like structure is known as the *dendrogram*, which illustrates the hierarchical relationships among the clusters.

Why hierarchical clustering?

- We have seen in the partitioning clustering (eg: K-means) that there are some challenges with these algorithm, which are a predetermined number of clusters, and it always tries to create the clusters of the same size.
- To solve these two challenges, we can opt for the hierarchical clustering algorithm because, in this algorithm, we ***do not need*** to have knowledge about the ***predefined number*** of clusters.

Types of Hierarchical Clustering:

Basically, there are two types of hierarchical Clustering:

- **Agglomerative** Hierarchical Clustering (AHC)
 - It follows **bottom-up** approach
- **Divisive** Hierarchical clustering (DHC)
 - It follows **top-down** approach

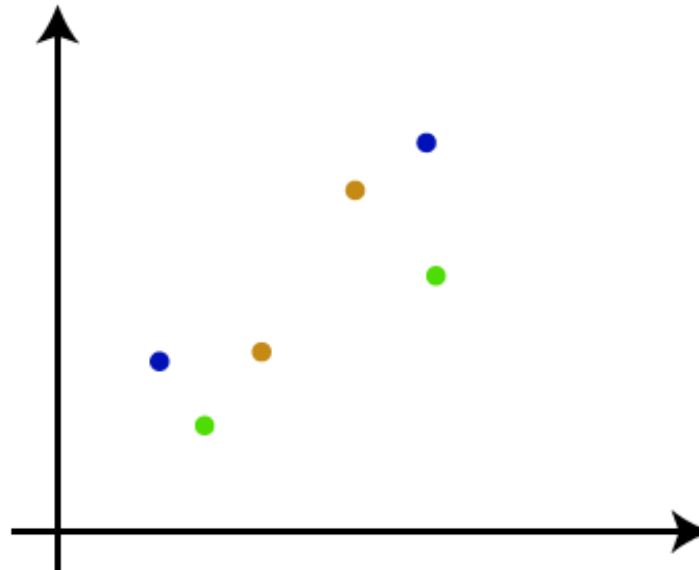
Agglomerative Hierarchical Clustering

- In this technique, initially *each data point* is considered as an *individual cluster*. At each iteration, the similar datapoints / clusters *merge* with other clusters until one cluster.
- In other words, the Agglomerative clustering follows a *bottom-up* approach, it means, this algorithm considers each datapoint as a single cluster at the beginning, and then start combining the closest pair of clusters together. It does this until all the clusters are merged into a single cluster that contains all the datapoints.

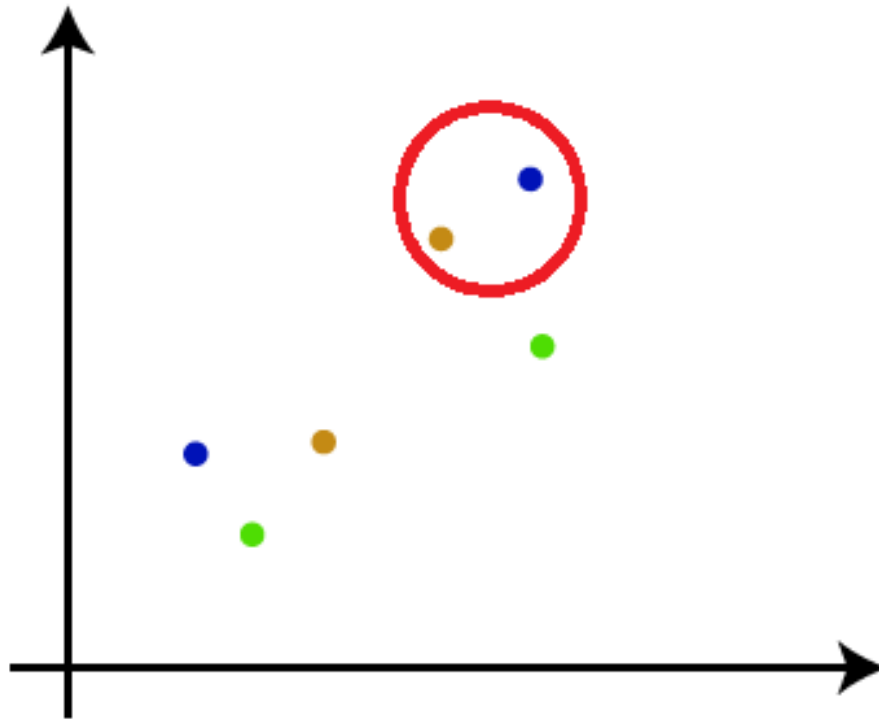
How the Agglomerative Hierarchical clustering Work?

The working of the AHC algorithm can be explained using the below steps:

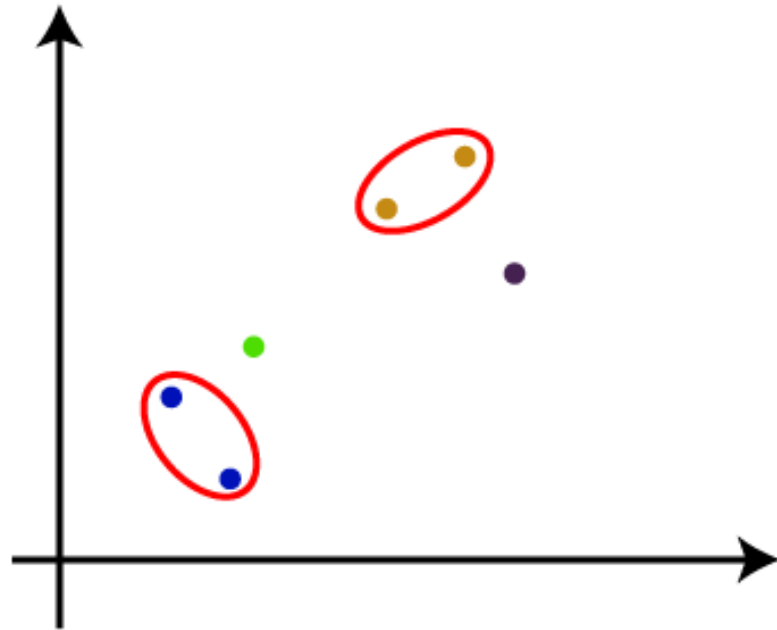
Step-1: Create *each data point* as a *single cluster*. Let us say there are **N** data points, so the number of clusters will also be **N**.



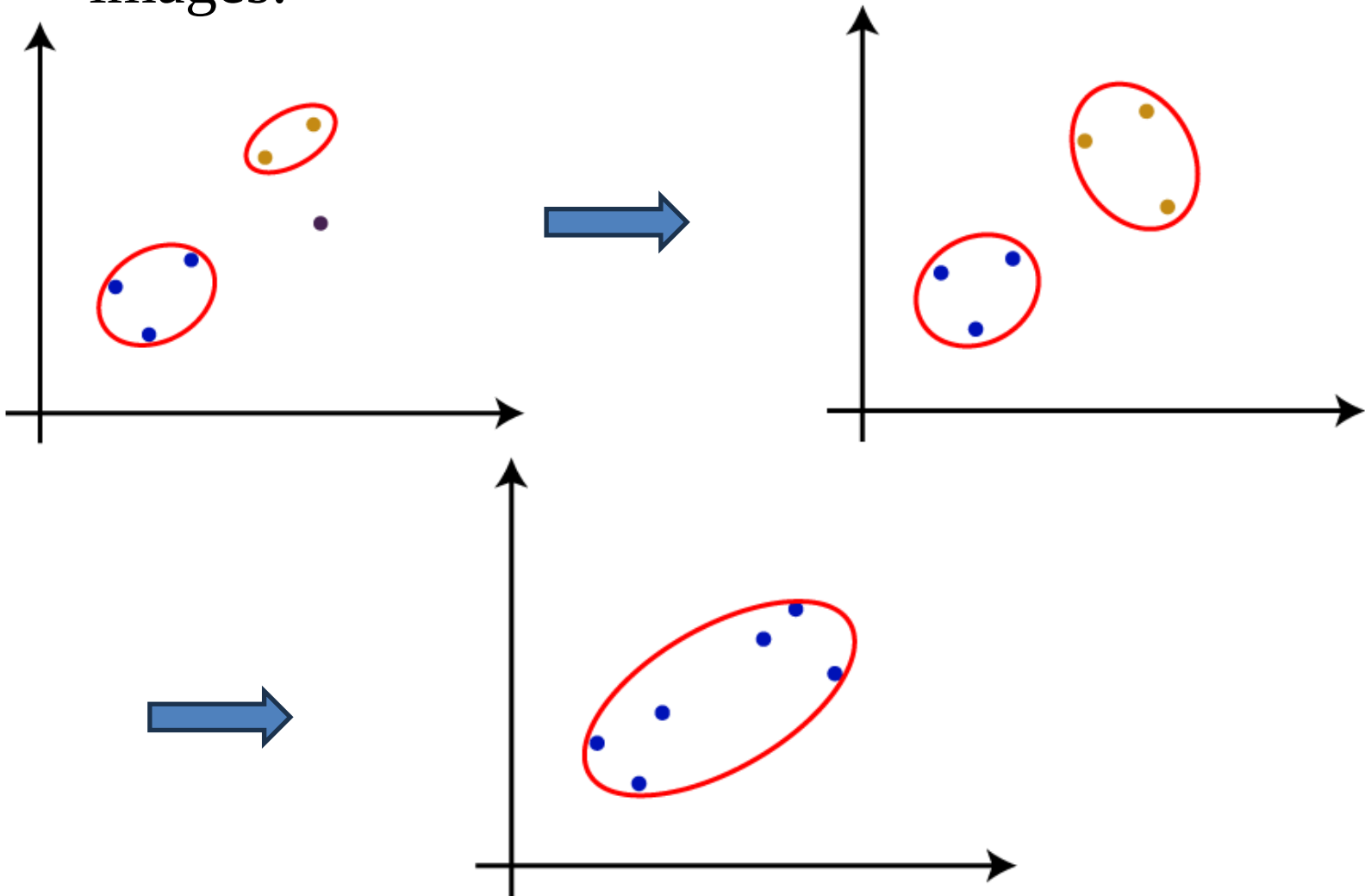
Step-2: Take *two closest data points* or clusters and *merge them* to form one cluster. So, there will now be $N-1$ clusters.



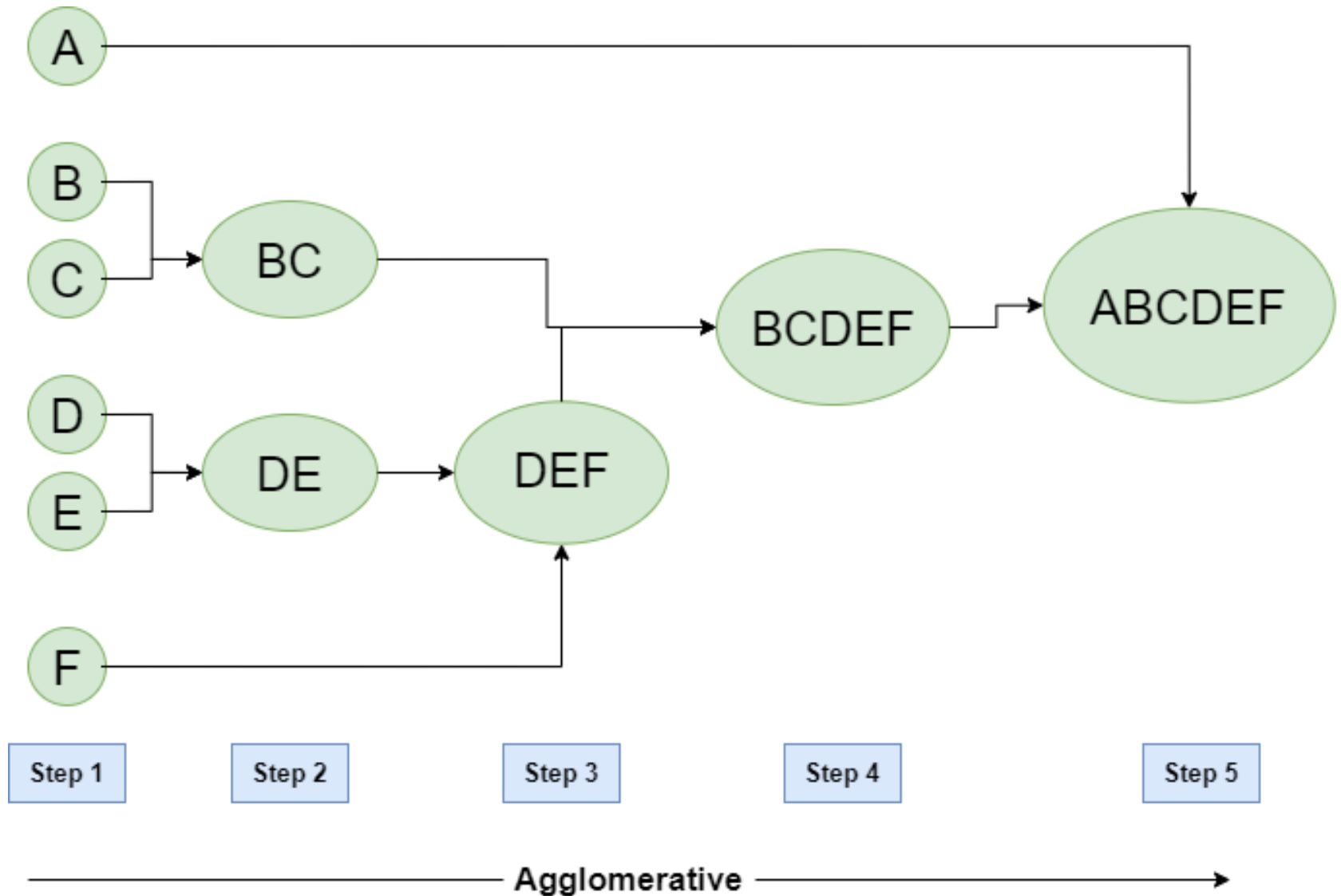
Step-3: *Again*, take the two closest clusters and merge them together to form one cluster. There will be $N-2$ clusters.



- **Step-4:** Repeat Step 3 until only *one cluster* left. So, we will get the following clusters. Consider the below images:



Example: Let us assume, we have six data points *A*, *B*, *C*, *D*, *E*, and *F*.



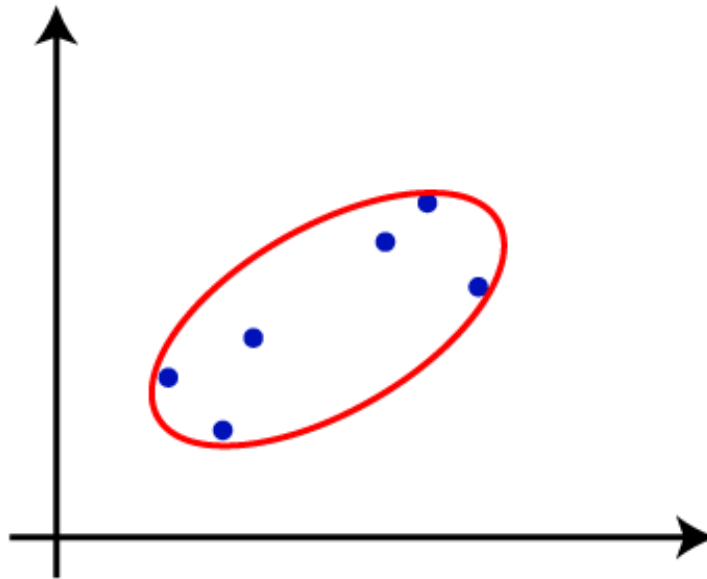
Divisive Hierarchical Clustering

- In simple words, we can say that the Divisive Hierarchical clustering is exactly the ***opposite*** of the Agglomerative Hierarchical clustering.
- In Divisive Hierarchical clustering, we consider all the data points as a ***single cluster*** and in each iteration, we ***separate*** the data points from the cluster which are ***not similar***. Each data point which is separated is considered as an individual cluster. In the end, we'll be left with **N** clusters.
- The Divisive Hierarchical clustering follows a ***top - down approach***.

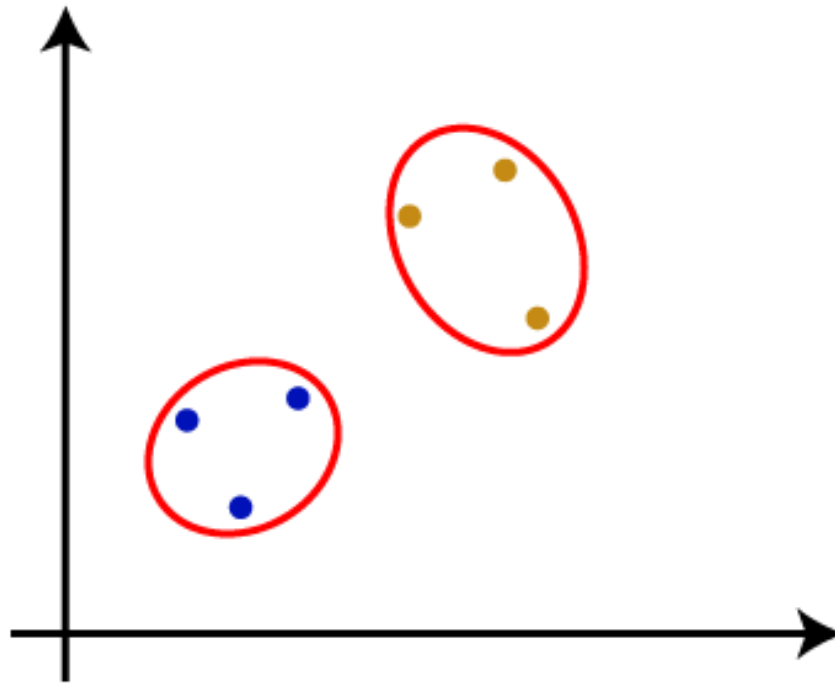
How the Divisive Hierarchical clustering Work?

The working of the DHC algorithm can be explained using the below steps:

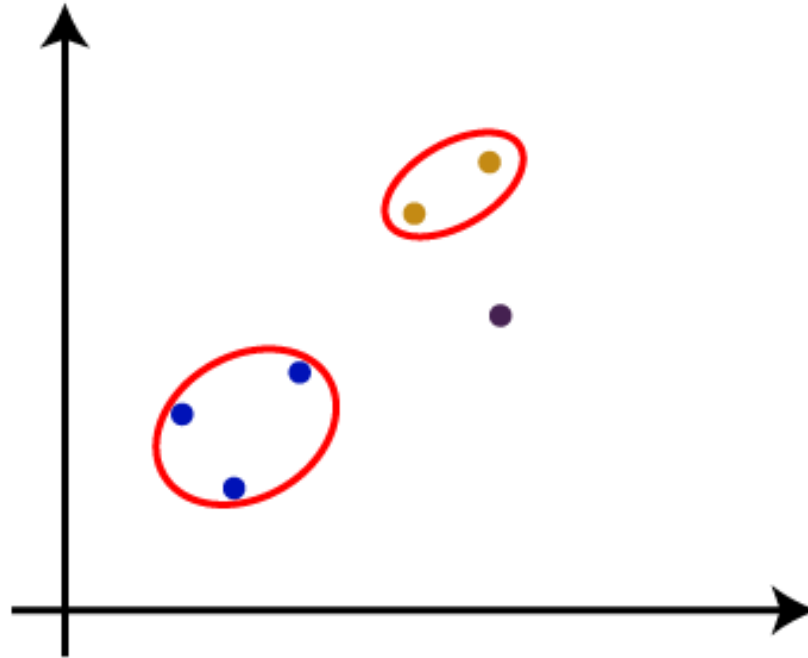
Step-1: Consider all the datapoints as a *single cluster*. Let us say there are **N** data points, so the number of clusters is 1.



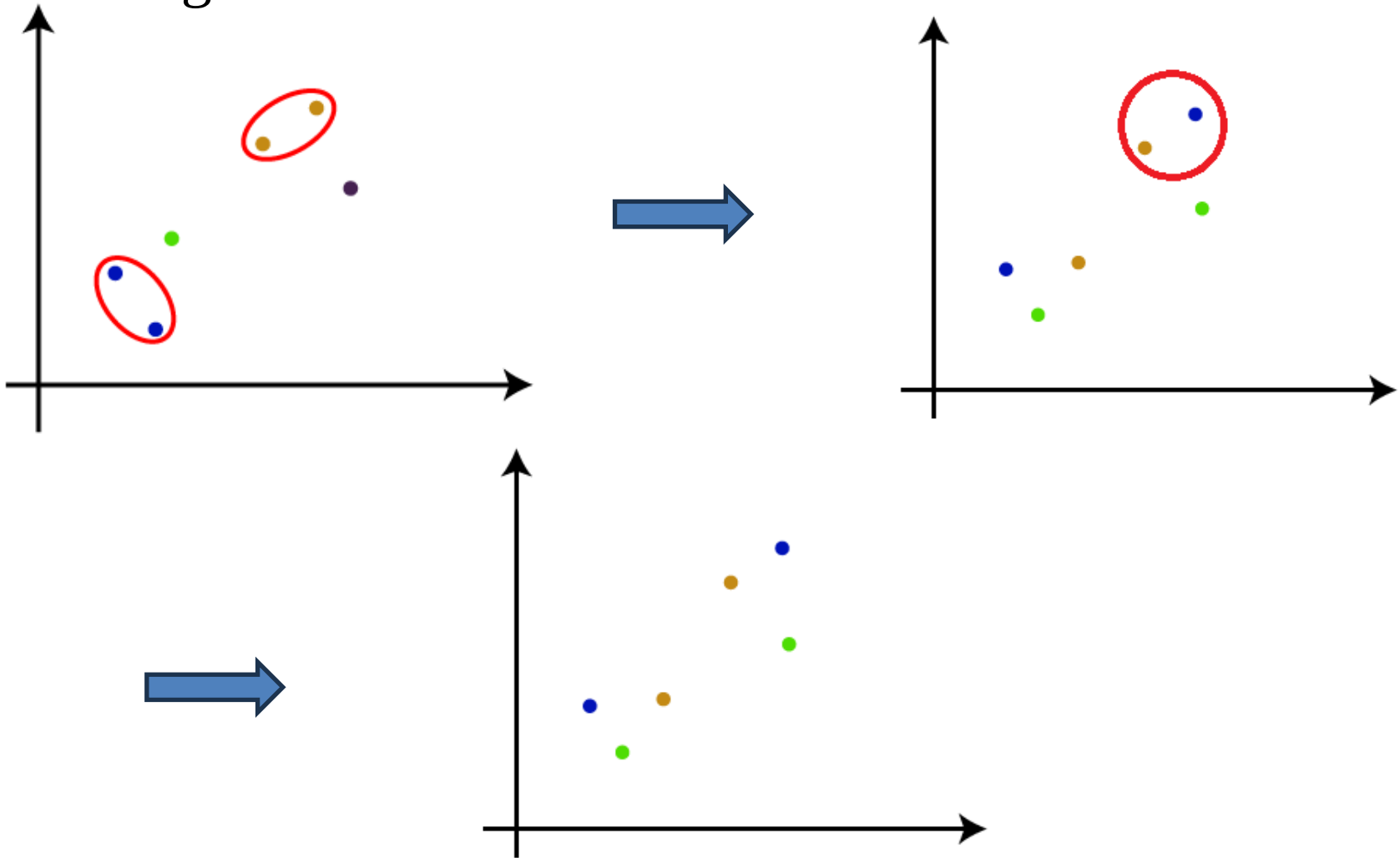
Step-2: *Separate* the datapoints which are not similar into clusters. So, there will now be 2 clusters.



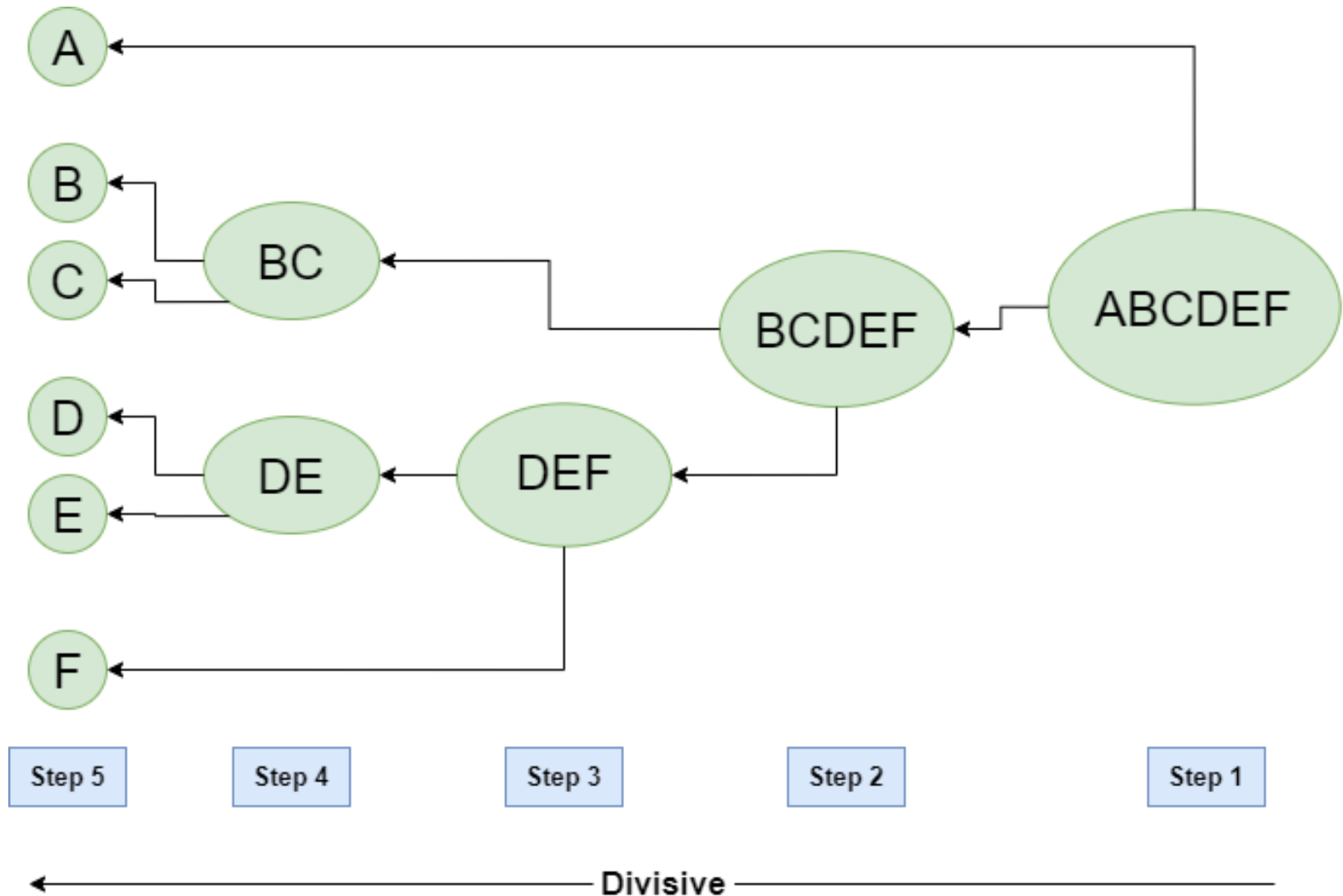
Step-3: *Again, Separate* the datapoints which are not similar into clusters. So, there will now be 3 clusters.



- **Step-4:** Repeat *Step 3* until left with **N** clusters. So, we will get the following clusters. Consider the below images :



Example: Let us assume, we have six data points *A*, *B*, *C*, *D*, *E*, and *F*.

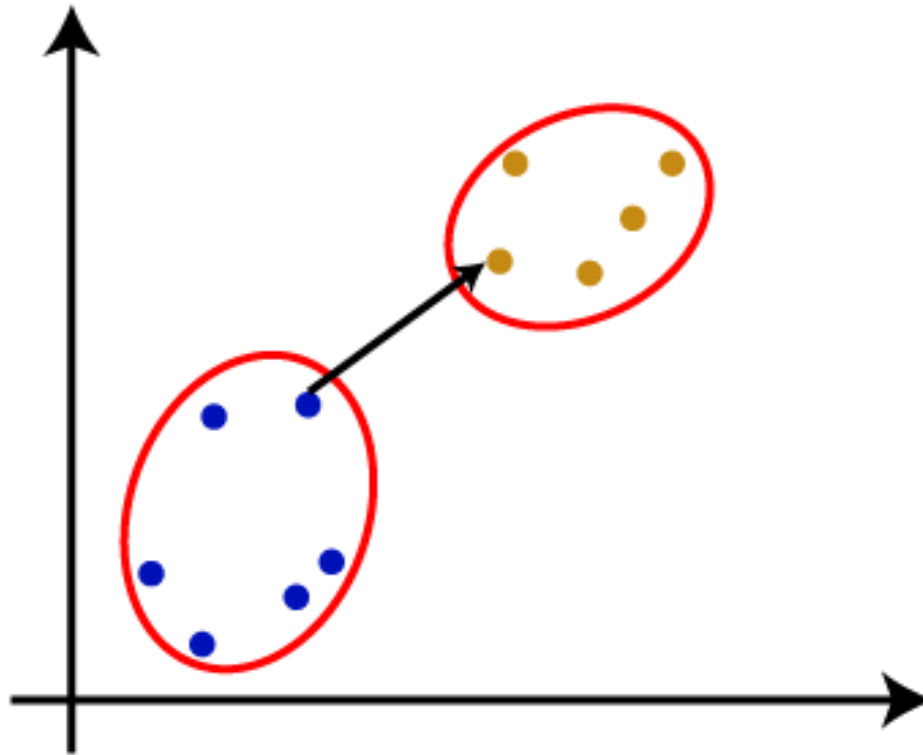


Measure for the distance between two clusters:

- As we have seen, the *closest distance* between the two clusters is crucial for the hierarchical clustering.
- There are various ways to calculate the distance between two clusters. These measures are called *Linkage methods*.
- Some of the popular linkage methods are given below:

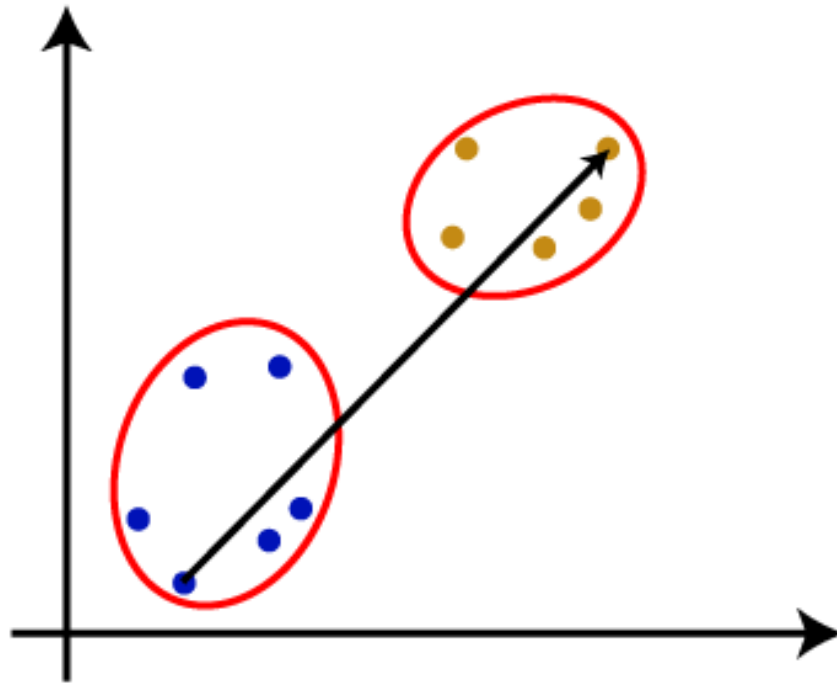
Single Linkage:

It is the *Shortest* Distance between the closest points of the clusters.



Complete Linkage:

It is the *farthest* distance between the two points of two different clusters. It is one of the popular linkage methods.



Centroid Linkage:

It is the linkage method in which the distance between the *centroid* of the clusters is calculated.

