Course: Machine Learning

(Course Code: 20CS3020AA)

Topic: Hierarchical Clustering

Module - 4

Unit - 4

Mr.D.MANI MOHAN
Asst. Professor

AIM

To familiarize students with the concept of Hierarchical clustering.

INSTRUCTIONAL OBJECTIVES

This Session is designed to:

1. Explain about Hierarchical clustering.
2. Demonstrate Types of Hierarchical clustering.
3. Analyze agglomerative (bottom-up) and divisive (top-down).

LEARNING OUTCOMES

At the end of this session, you should be able to:

1. Explain about Hierarchical clustering.
2. Demonstrate Types of Hierarchical clustering.
3. Analyze agglomerative (bottom-up) and divisive (top-down).

WHAT IS HIERARCHICAL CLUSTERING

- Hierarchical clustering is a popular method for grouping objects.
- It creates groups so that objects within a group are similar to each other and different from objects in other groups.
- Clusters are visually represented in a hierarchical tree called a dendrogram.
- Hierarchical clustering is the most popular and widely used method to analyze social network data.
- In this method, nodes are compared with one another based on their similarity.
- Larger groups are built by joining groups of nodes based on their similarity.

Hierarchical Clustering

- A **Hierarchical clustering** method works via grouping data into a tree of clusters. Hierarchical clustering begins by treating every data point as a separate cluster. Then, it repeatedly executes the subsequent steps:

1. Identify the 2 clusters which can be closest together, and
2. Merge the 2 maximum comparable clusters. We need to continue these steps until all the clusters are merged together.

- In Hierarchical Clustering, the aim is to produce a hierarchical series of nested clusters.
- A Dendrogram is a tree-like diagram that statistics the sequences of merges or splits.

THE HIERARCHICAL CLUSTERING TECHNIQUE HAS TWO APPROACHES:

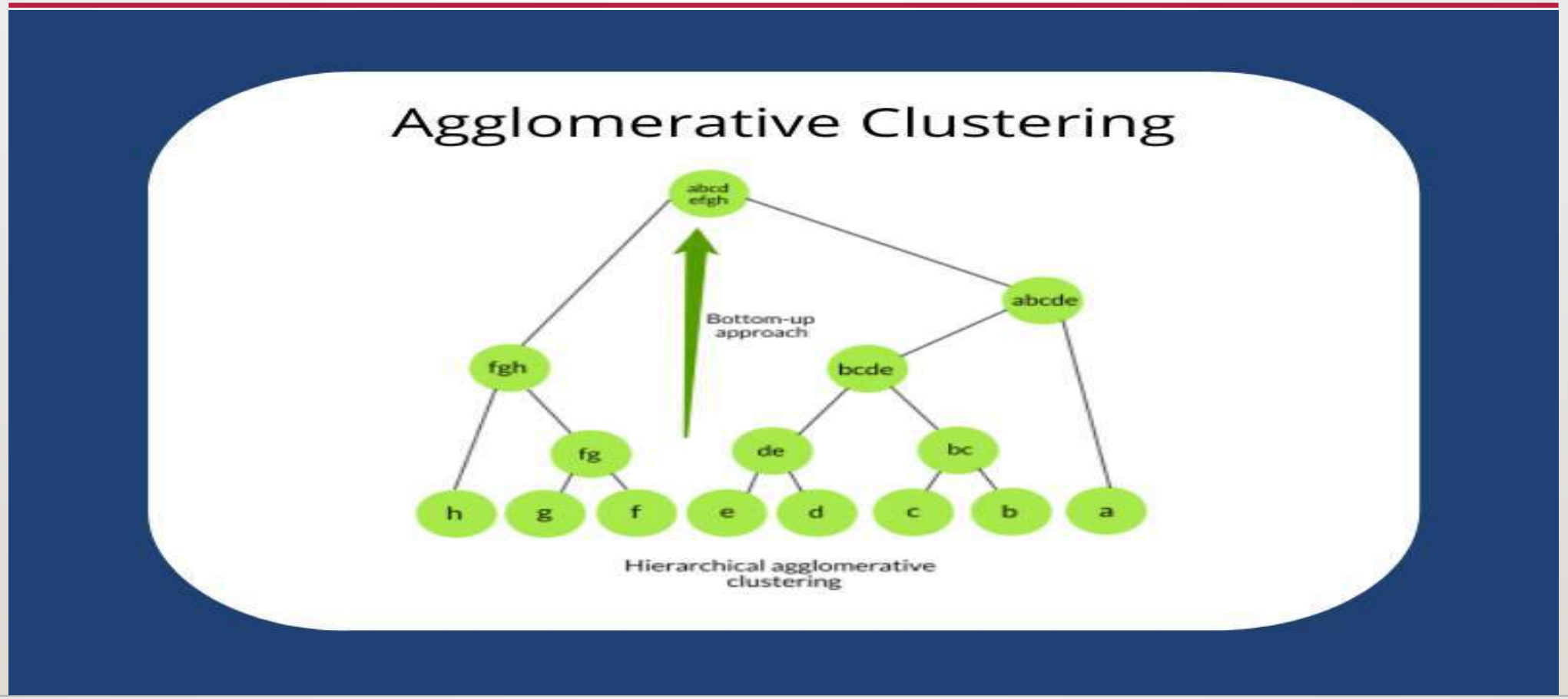
1. **Agglomerative:** Agglomerative is a **bottom-up** approach, in which the algorithm starts with taking all data points as single clusters and merging them until one cluster is left.
2. **Divisive:** Divisive algorithm is the reverse of the agglomerative algorithm as it is a **top-down approach.**

Types of Hierarchical Clustering

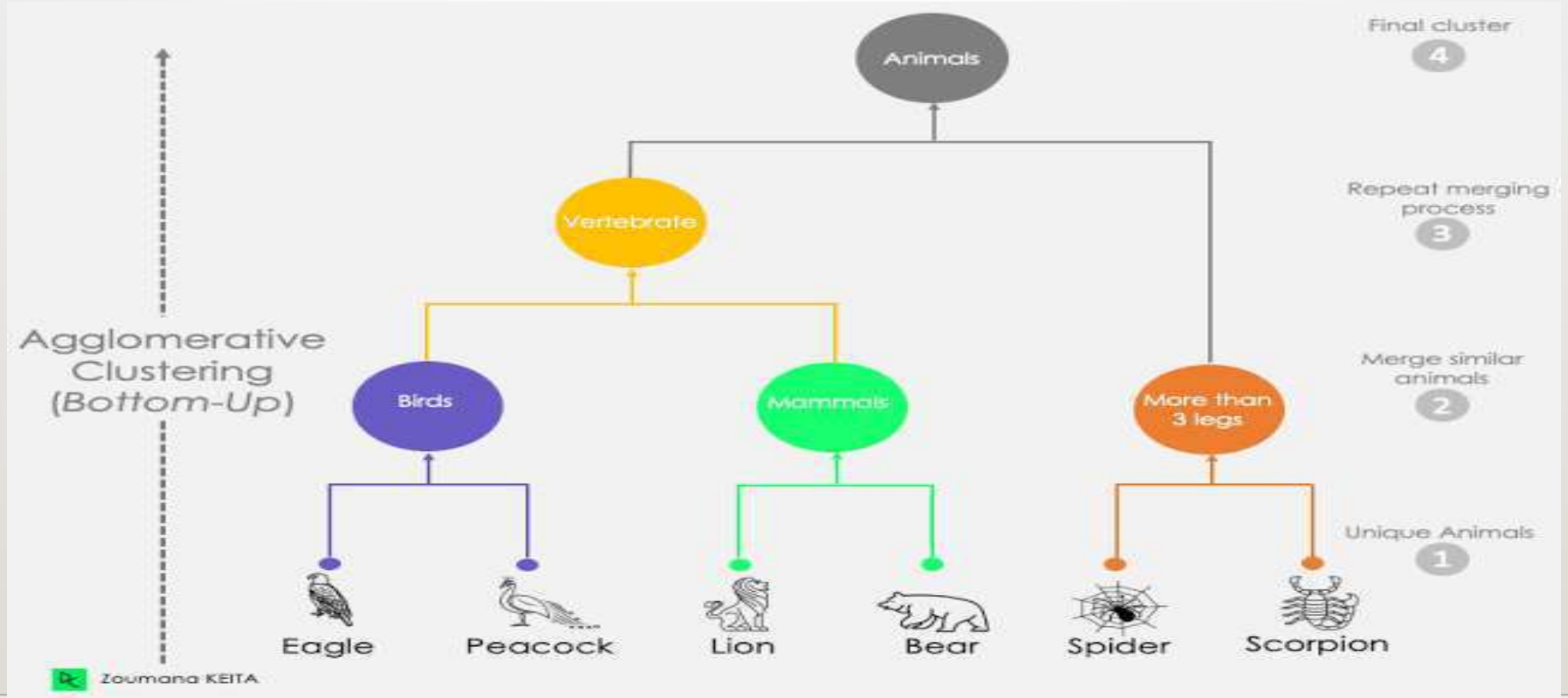
AGGLOMERATIVE CLUSTERING

- Agglomerative clustering is one of the most common types of hierarchical clustering used to group similar objects in clusters.
- Agglomerative clustering is also known as AGNES (Agglomerative Nesting). In agglomerative clustering, each data point act as an individual cluster and at each step, data objects are grouped in a bottom-up method.
- Initially, each data object is in its cluster. At each iteration, the clusters are combined with different clusters until one cluster is formed.

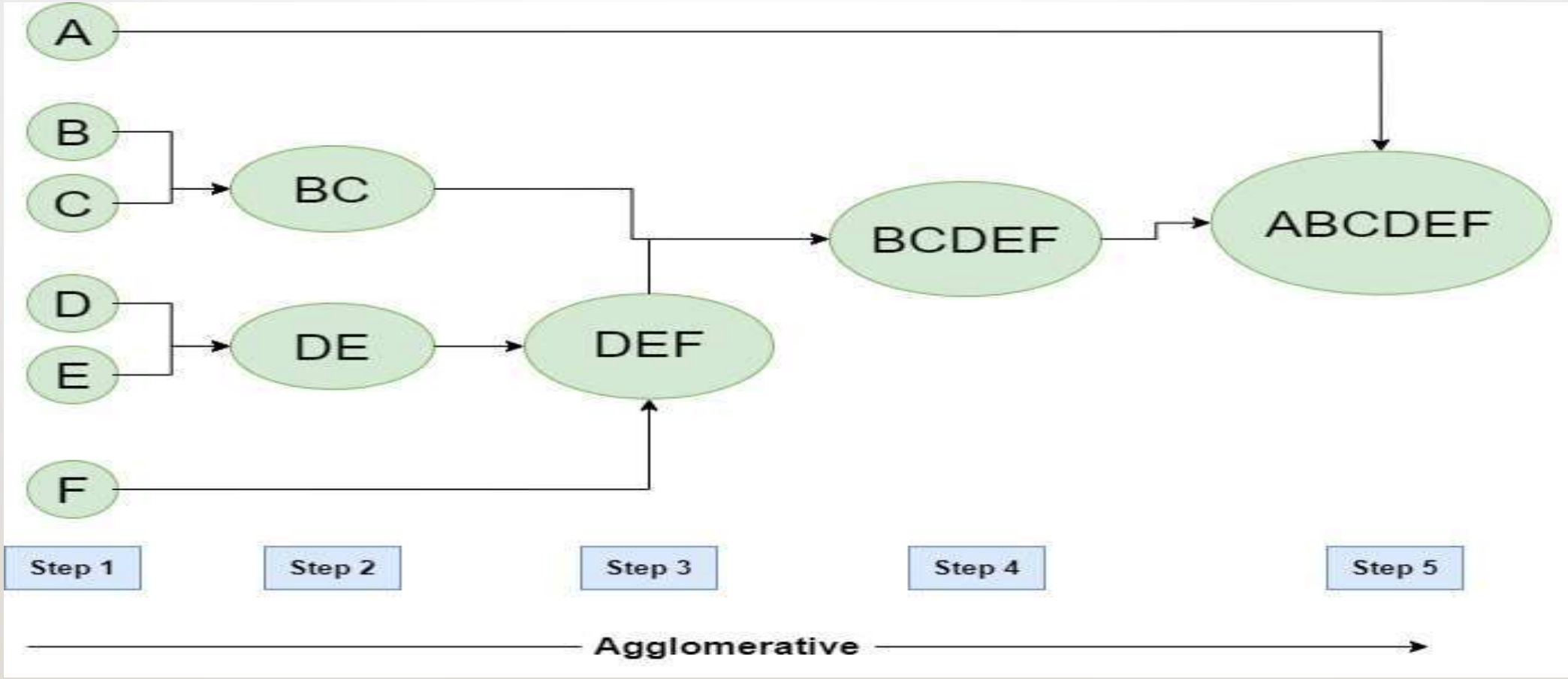
AGGLOMERATIVE CLUSTERING



AGGLOMERATIVE CLUSTERING



AGGLOMERATIVE CLUSTERING



AGGLOMERATIVE CLUSTERING

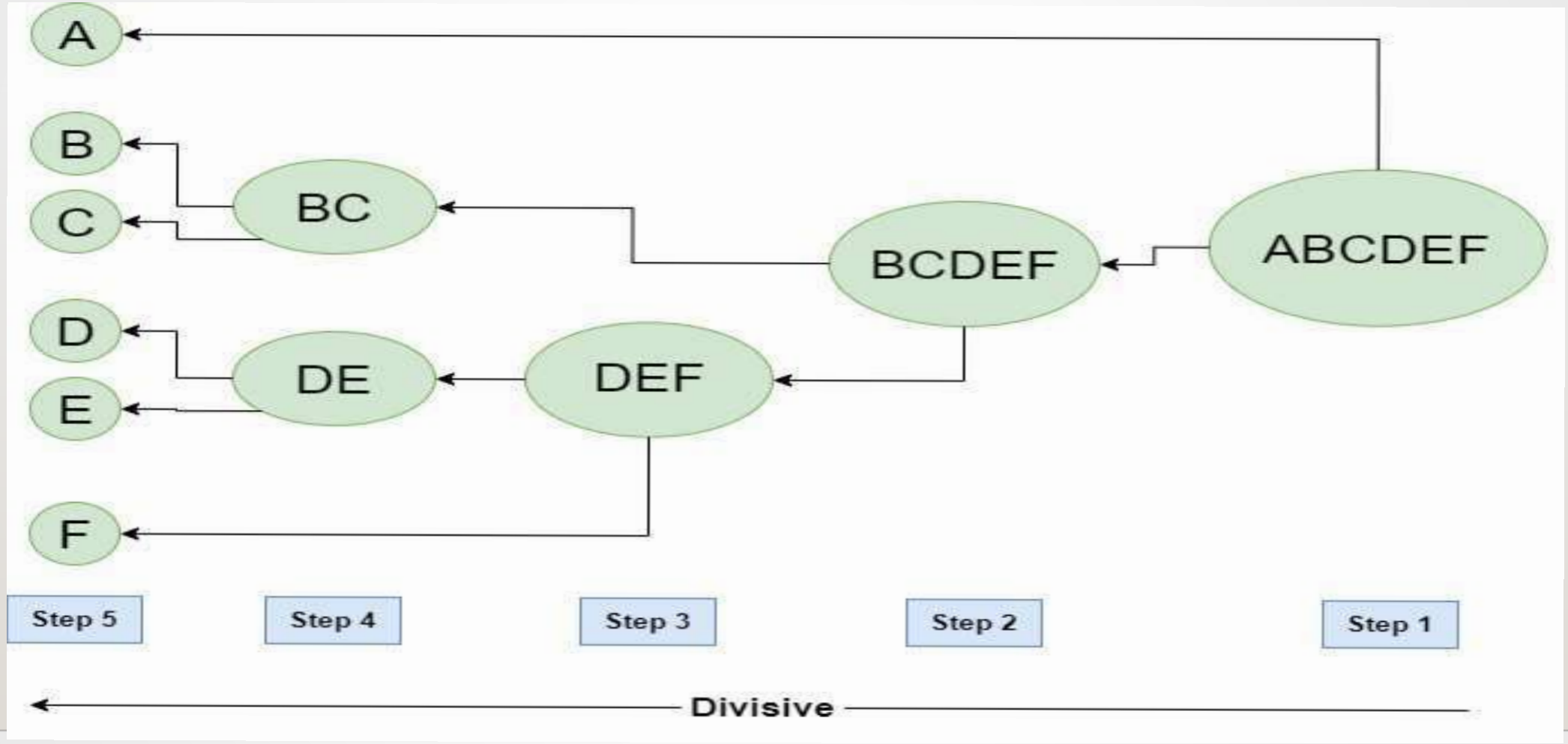
- The algorithm for Agglomerative Hierarchical Clustering is:
- Calculate the similarity of one cluster with all the other clusters (calculate proximity matrix)
- Consider every data point as an individual cluster
- Merge the clusters which are highly similar or close to each other.
- Recalculate the proximity matrix for each cluster
- Repeat Steps 3 and 4 until only a single cluster remains.

DIVISIVE HIERARCHICAL CLUSTERING

- Divisive hierarchical clustering is exactly the opposite of Agglomerative Hierarchical clustering.
- In Divisive Hierarchical clustering, all the data points are considered an individual cluster, and in every iteration, the data points that are not similar are separated from the cluster.
- The separated data points are treated as an individual cluster. Finally, we are left with N clusters.

DIVISIVE HIERARCHICAL CLUSTERING

DIVISIVE CLUSTERING



DIVISIVE CLUSTERING

- This approach starts with all of the objects in the same cluster.
- In the continuous iteration, a cluster is split up into smaller clusters.
- It is down until each object in one cluster or the termination condition holds.
- This method is rigid, i.e., once a merging or splitting is done, it can never be undone.

APPLICATIONS

- Clustering analysis is broadly used in many applications such as market research, pattern recognition, data analysis, and image processing.
- Clustering can also help marketers discover distinct groups in their customer base.
- And they can characterize their customer groups based on the purchasing patterns.
- We don't have to pre-specify any particular number of clusters. ...

- Easy to decide the number of clusters by merely looking at the Dendrogram.

Conclusion

- Hierarchical clustering is a popular method for grouping objects.
- It creates groups so that objects within a group are similar to each other and different from objects in other groups.
- Types of Hierarchical Clustering
- Agglomerative clustering is one of the most common types of hierarchical clustering used to group similar objects in clusters.
- In Divisive Hierarchical clustering, all the data points are considered an individual cluster, and in every iteration, the data points that are not similar are separated from the cluster.

Self-Assessment Questions

1. What are the two types of Hierarchical Clustering?

(a) Top-Down Clustering (Divisive)Boosting

(b) Bottom-Top Clustering (Agglomerative)

(c) Both a and b

(d) Dendrogram

2. Hierarchical clustering should be mainly used for exploration.

(a) TRUE

(b) FALSE

Self-Assessment Questions

3. Which of the following is not clustering method?

(a) Dbscan
(b) Hierarchy
(c) Grid
(d) Project based

4. _____clusters formed in this method forms a tree-type structure based on the hierarchy.

(a) Dbscan
(b) Hierarchy
(c) Grid
(d) Project based

THANK YOU

OUR TEAM