Hierarchical Clustering

WHAT IS HIERARCHICAL CLUSTERING?

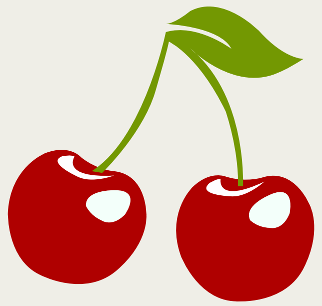
Hierarchical Clustering is an unsupervised machine learning algorithm used to group similar data points into clusters. Unlike K-Means, you don’t need to specify the number of clusters at the beginning.

EXAMPLE:

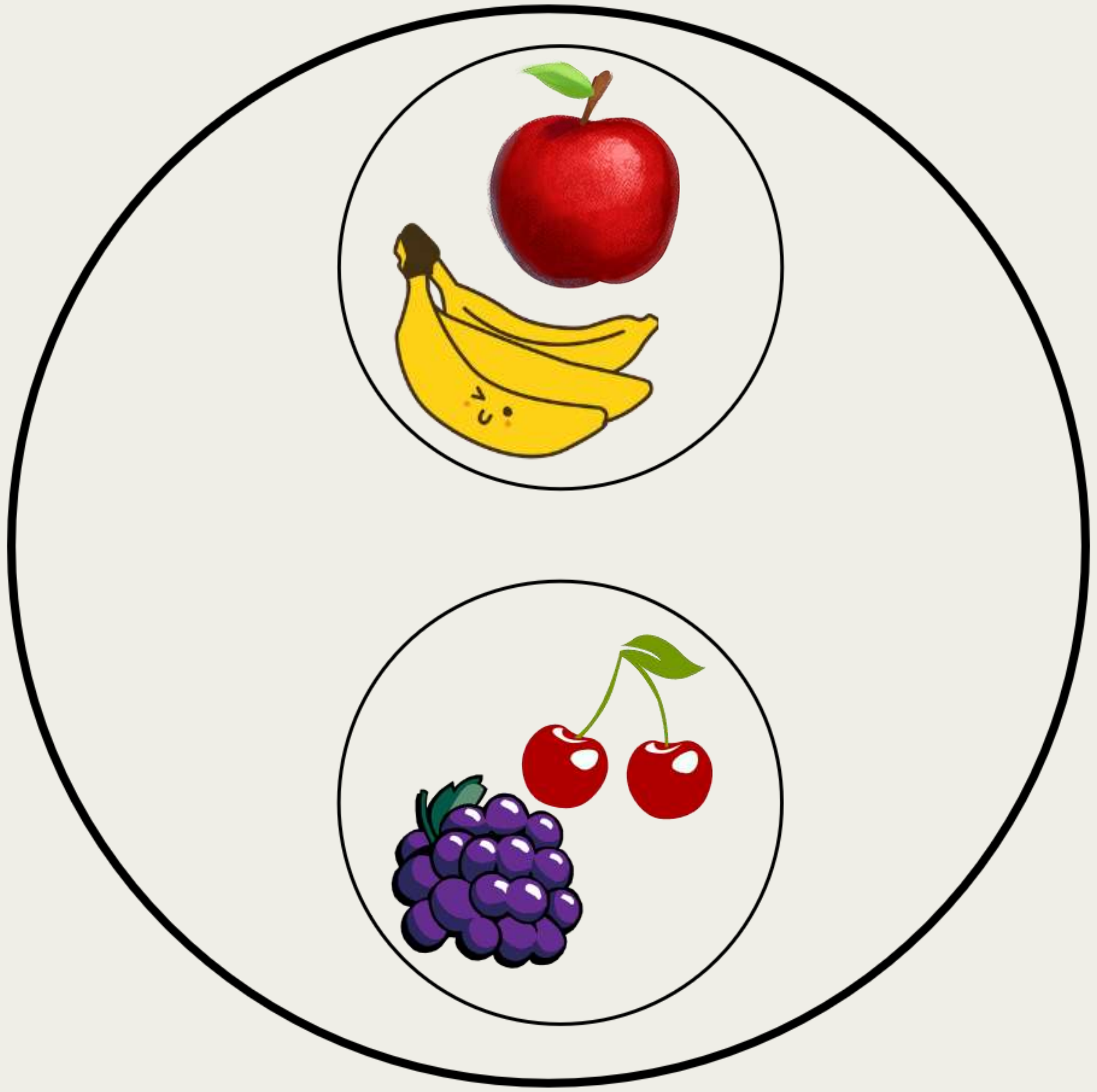
100g

120g

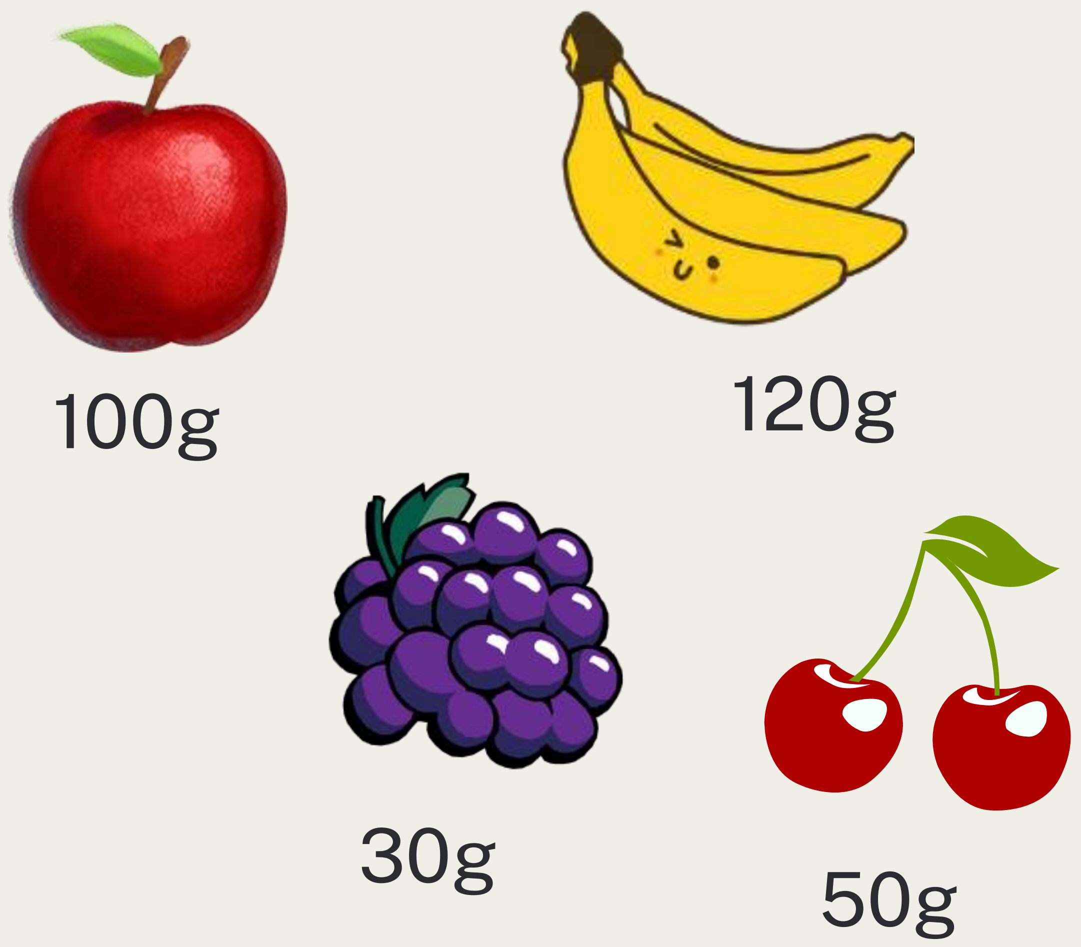
30g

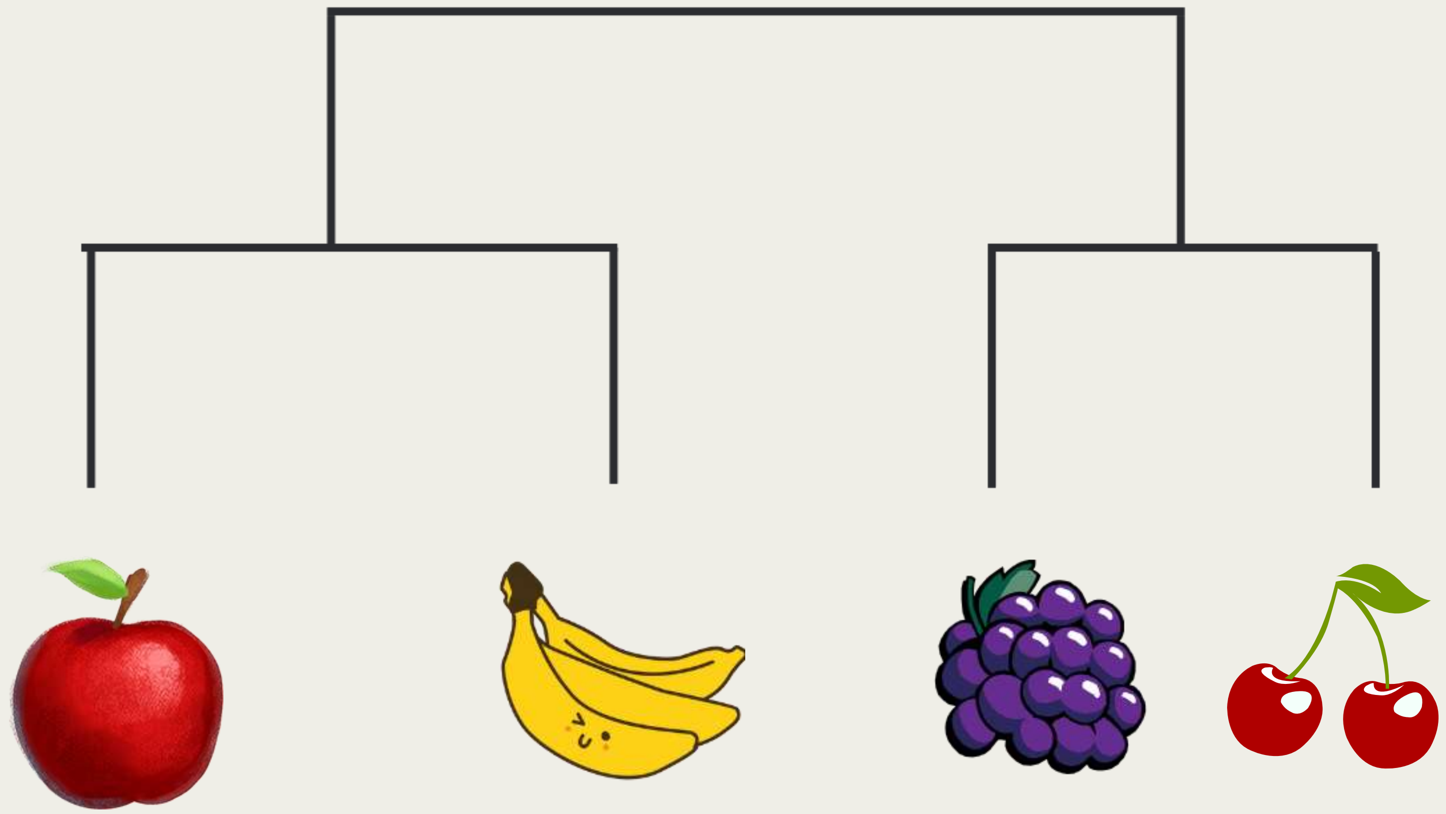


50g



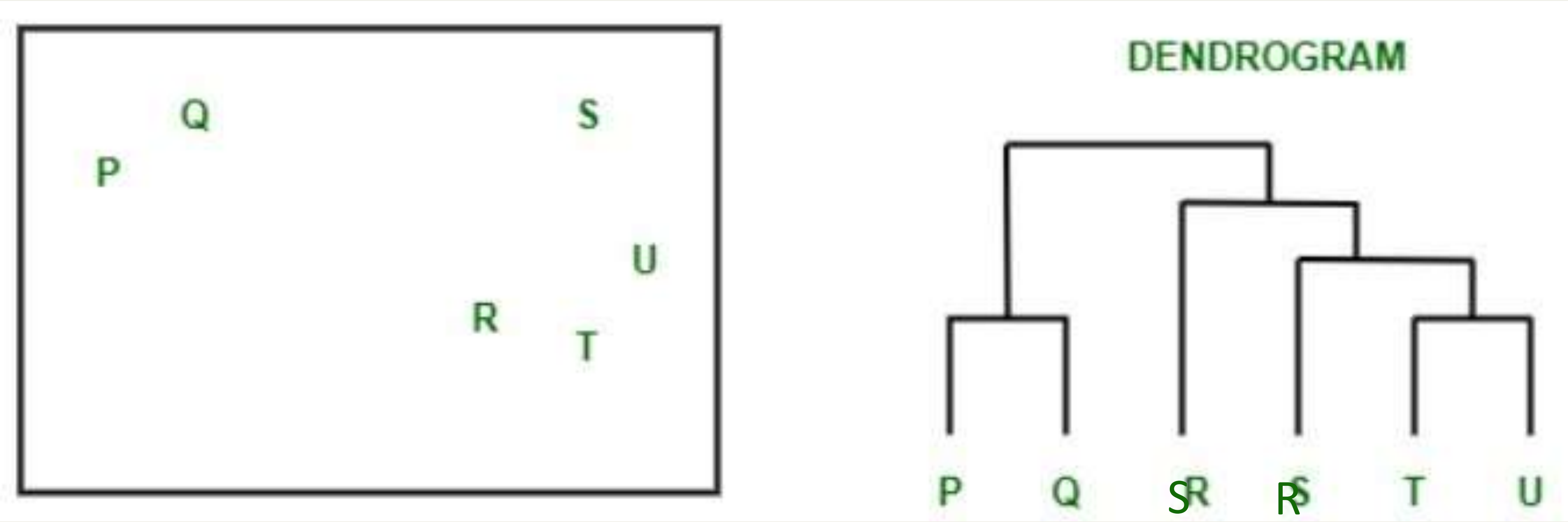
DENDOGRAM





DENDOGRAM - ANOTHER EXAMPLE

A dendrogram is like a family tree for clusters. It shows how individual data points or groups of data merge together. The bottom shows each data point as its own group, and as you move up, similar groups are combined. The lower the merge point, the more similar the groups are. It helps you see how things are grouped step by step. The working of the dendrogram can be explained using the below diagram:



TYPES OF HIERARCHICAL CLUSTERING

Agglomerative Clustering

It is also known as the bottom-up approach or hierarchical agglomerative clustering (HAC).

Divisive clustering

It is also known as a top-down approach.

HIERARCHICAL AGGLOMERATIVE CLUSTERING

Bottom-up algorithms treat each data as a singleton cluster at the outset and then successively agglomerate pairs of clusters until all clusters have been merged into a single cluster that contains all data.

Start with individual points

↓

Calculate distances between clusters

↓

Merge the closest clusters

↓

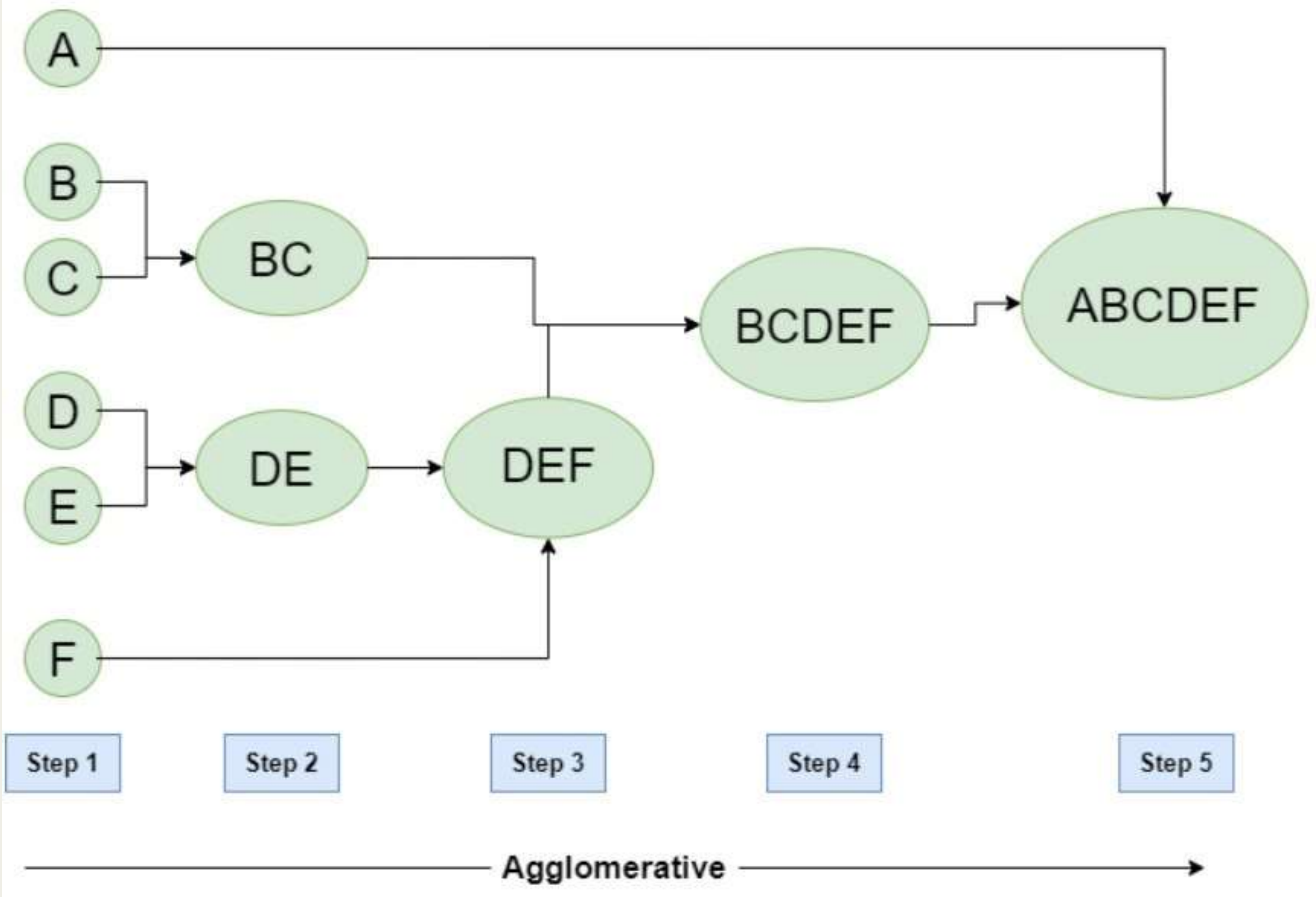
Update distance matrix

↓

Repeat steps 3 and 4

↓

Create a dendrogram



HIERARCHICAL DIVISIVE CLUSTERING

Top-down clustering requires a method for splitting a cluster that contains the whole data and proceeds by splitting clusters recursively until individual data have been split into singleton clusters.

Start with all data points in one cluster

↓

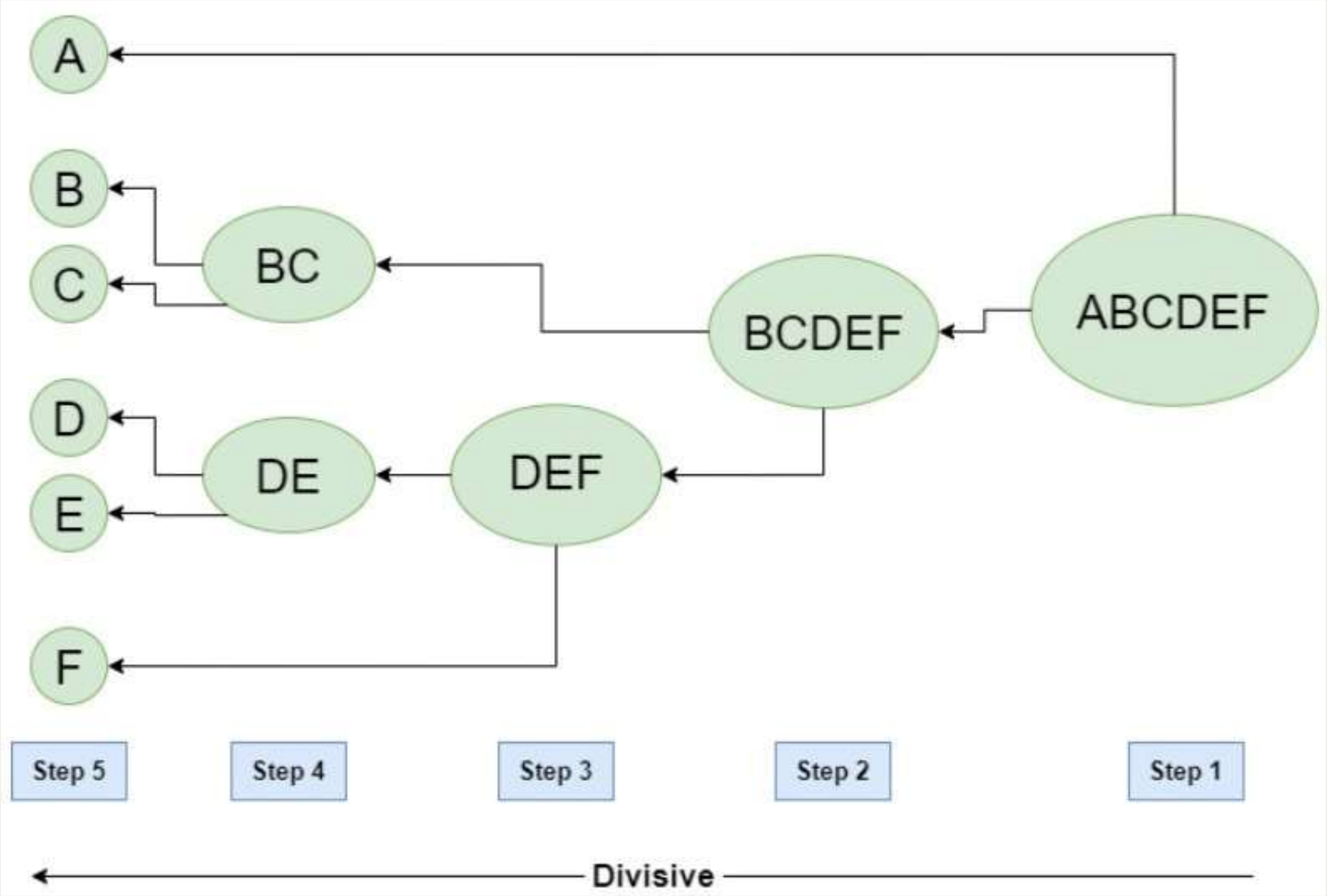
Split the cluster

↓

Repeat the process

↓

Stop when each data point is in its own cluster:



LINKAGE

Single Linkage

Minimum Distance

Take the smallest distance between any pair of points (one from each cluster). 📌 Distance(A, B) = min(5, 6, 7, 8) = 5 Think: “Which two points are closest from each cluster?”

CompleteLinkage

Maximum Distance

Take the largest distance between any pair of points (one from each cluster). 📌 Distance(A, B) = max(5, 6, 7, 8) = 8 Think: “What’s the furthest pair between these two clusters?”

AverageLinkage

Mean Distance

Take the averageof allpairwise distances between points in the two clusters. 📌 Distance(A, B) = (5 + 6 + 7 + 8) / 4 = 6.5 Think: “What’s the average distance between all points in A and B?”

Ward’s Linkage

Minimum Variance

Thisis a bit different – it doesn’t directly use distances between individual points. Instead, it tries to merge clusters in a way that increases the total within-cluster variance as little as possible.

📌 Think of it as: “Merge clusters that lead to the least increase in total squared distances within the new cluster.”

✅ Ward’s method tends to create balanced-sized, compact clusters.

From → To	Distance
1 → 3	5
1 → 4	7
2 → 3	6
2 → 4	8

Assume we have two clusters:

Cluster A = {Point 1, Point 2}

Cluster B = {Point 3, Point 4}

Steps in Agglomerative Hierarchical Clustering

Dataset (3 points):

Point	X	Y
A	2	1
B	4	7
C	5	4

Step 1: Treat Each Point as a Cluster

- Initial clusters: {A}, {B}, {C}

Step 2: Calculate Distances Between Clusters

Use Euclidean distance:

- $\text{Distance}(\text{A, B}) = \sqrt{(2-4)^2 + (1-7)^2} = \sqrt{4+36} = \sqrt{40} \approx 6.32$
- $\text{Distance}(\text{A, C}) = \sqrt{(2-5)^2 + (1-4)^2} = \sqrt{9+9} = \sqrt{18} \approx 4.24$
- $\text{Distance}(\text{B, C}) = \sqrt{(4-5)^2 + (7-4)^2} = \sqrt{1+9} = \sqrt{10} \approx 3.16$

Step 3: Merge the Closest Pair

- Find the closest pair: Min distance between any two clusters.
- Merge (for example): {B}, {C} (if Distance(B, C) is the minimum).
- New clusters after merge: {A}, {B, C}

Step 4: Recalculate Distances

- Use a linkage method to define the distance between clusters.
- **Common linkage methods:**
  - Single linkage: min distance between two clusters
  - Complete linkage: max distance
  - Average linkage: mean distance
  - Ward's method: minimize variance
- **Assume Single Linkage:**

$$\text{Distance}(A, B, C) = \min(\text{Distance}(A, B), \text{Distance}(A, C))$$

Step 5: Merge Again

- Merge remaining clusters based on recalculated distances.
- Final cluster: {A, B, C}