Lesson 13
Unsupervised Learning

(Hierarchical Clustering)

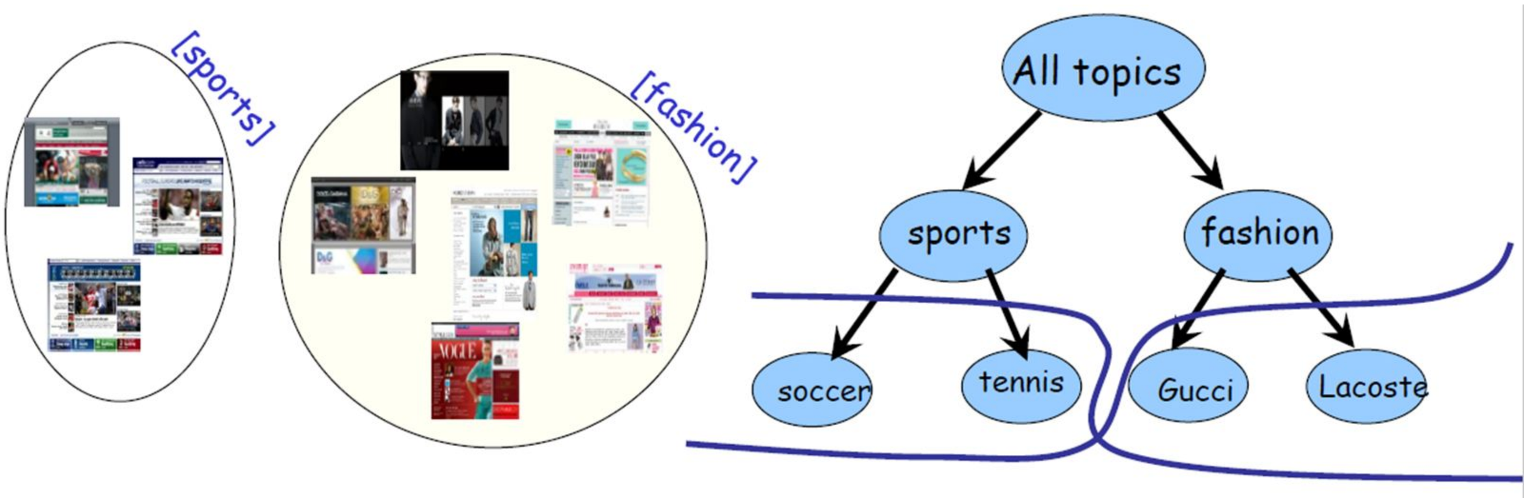
Credit to(Bishop: 9-9.2)

Hierarchical Clustering

Hierarchical clustering is a method of cluster analysis used in data mining. It seeks to build a hierarchy of clusters in a step-by-step manner.

Hierarchical algorithms: Create a hierarchical decomposition of the set of objects using some criterion (focus of this class)

Build a tree-based hierarchical from a set of documents.



Different users might care about different levels of granularity.

There are two main types of hierarchical clustering:

1- Agglomerative (Bottom-Up Approach): (in this class)

- Initial Step: Starts by treating each data point as a separate cluster. So, if there are N data points, you begin with N clusters.
- Clustering Process: In each step, the algorithm merges the two clusters that are **closest to each other** until all the clusters are merged into one big cluster containing all data points.
- Dendrogram: The result can be represented in a tree-like structure called a dendrogram, which shows the arrangement of the clusters and their proximity.

2- Divisive (Top-Down Approach):

- Initial Step: Begins with all data points in a single cluster.
- Clustering Process: At each step, the algorithm splits the cluster until each cluster contains only one data point.
- Top-Down Splitting: This is less common compared to agglomerative clustering and is computationally more intensive.

Algorithm of Agglomerative Hierarchical Clustering :

1. **Initialization**:
   - Treat each data point as a separate cluster. Thus, if you have N data points, you start with N clusters, each containing just one data point.
2. **Compute Distance Matrix:**
   - Calculate the distance between each pair of clusters. Common distance metrics include Euclidean, Manhattan, and Cosine distances. The choice of distance metric can significantly affect the outcome of the clustering.
   - This results in an N×N distance matrix, where the distance between a cluster and itself is zero.
3. **Find the Closest Clusters:**
   - Identify the two clusters that are closest to each other based on the distance matrix.
4. **Merge Clusters:**
   - Combine the two closest clusters into a single cluster.
   - This step reduces the total number of clusters by one.
5. **Update Distance Matrix:**
   - Recalculate the distances between the new cluster and all the existing clusters.
   - The method of recalculating the distance depends on the **linkage criterion used**.
6. **Repeat** :
   - • Repeat steps 3 to 5 until all data points are merged into **a single cluster**.

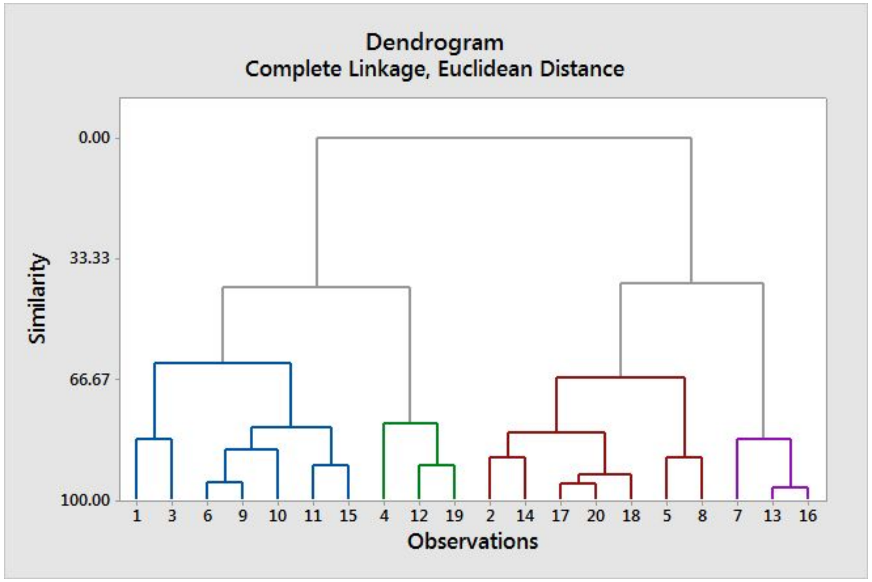
Important function

- Hierarchical cluster analysis on a set of dissimilarities
- Important function used here is a function that computes and returns the distance matrix computed by using the specified distance measure to compute the distances between the rows of a data matrix.
- By default, it is Euclidean distance.
- Mathematically Euclidean distance between two vectors p = (p1, p2,…, pn) and q = (q1, q2,…, qn) are two points in Euclidean. n-space is given by

$$d(p,q) = d(q,p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$

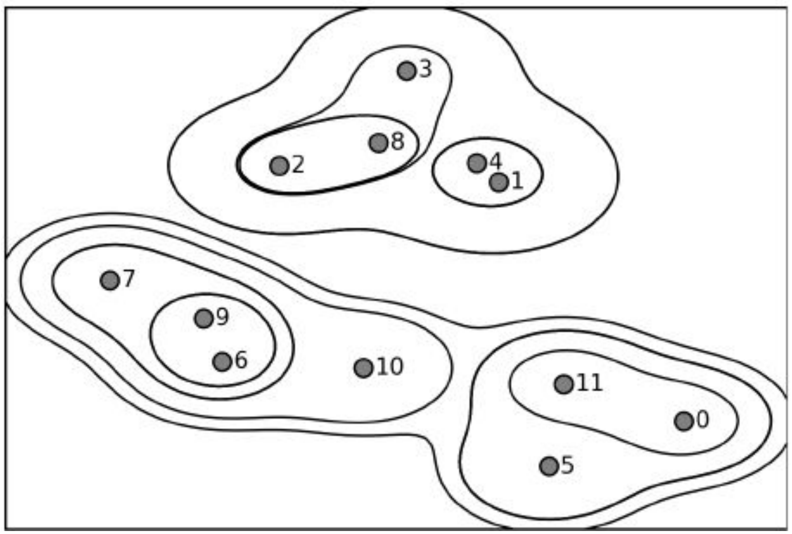
Cluster dendrogram

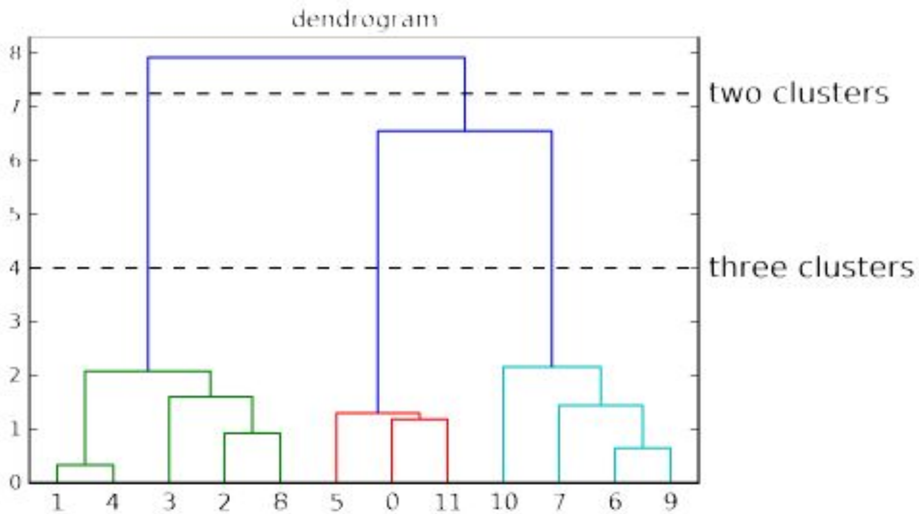
- The **dendrogram** is a tree diagram that displays the groups that are formed by clustering observations at each step and their similarity levels.
- The **similarity** level is measured along **the vertical axis** (alternately, you can display the distance level), and the different **observations** are listed along **the horizontal axis**.



Hierarchical Clustering and Dendrograms

- The figure below shows an overlay of all possible clustering shown in Figure **agglomerative_ algorithm**, providing some insight into how each cluster breaks up into smaller clusters





- **Finding number of cluster using Dendogram :**
  - Look for the Longest Vertical Line: The length of this vertical line indicates a substantial distance between the points it connects.

How do we compute distance between clusters

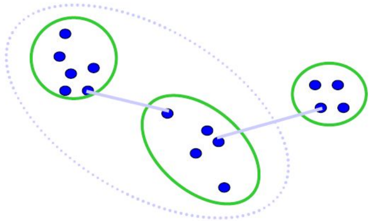
Single Link: Considering the distance between one cluster and another to be equal to the shortest distance from any member of one cluster to any member of the other cluster

Complete Link: The distance between one cluster and another is considered to be equal to the greatest distance from any member of one cluster to any member of the other cluster

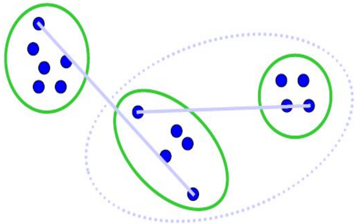
Average correlation: considering the distance between one cluster to another as equal to the average distance from any member of one cluster to any member of the other cluster

Computing distance between clusters

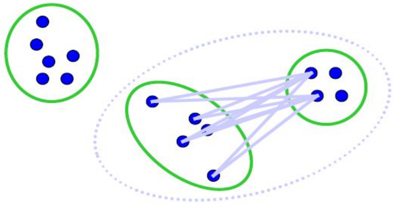
- Have a distance measure on pairs of objects.
- $d(x, y)$: Distance between x and y



- Single linkage: $$\text{dist}(A, B) = \min_{x \in A, x' \in B} d(x, x')$$



- Complete linkage: $$\text{dist}(A, B) = \max_{x \in A, x' \in B} d(x, x')$$



- Average linkage: $$\text{dist}(A, B) = \operatorname*{average}_{x \in A, x' \in B} d(x, x')$$

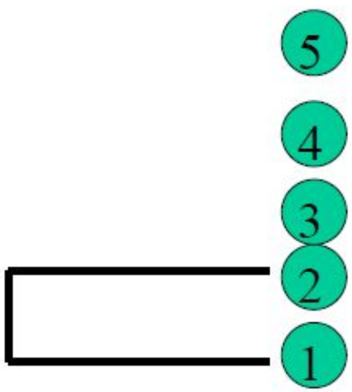
Example: single link

- Clustering starts by computing a distance between every pair of units that you want to cluster.
- A distance matrix will be symmetric (because the distance between x and y is the same as the distance between y and x)
- and will have zeroes on the diagonal (because every item is distance zero from itself). The table below is an example of a distance matrix.
- Only the lower triangle is shown, because the upper triangle can be filled in by reflection.

$$
\begin{array}{c|c}
 & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\
\begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} &
\begin{bmatrix}
0 & & & & \\
2 & 0 & & & \\
6 & 3 & 0 & & \\
10 & 9 & 7 & 0 & \\
9 & 8 & 5 & 4 & 0
\end{bmatrix}
\end{array}
$$

Example: single link

$$
\begin{array}{c|ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & 0 & & & & \\
2 & 2 & 0 & & & \\
3 & 6 & 3 & 0 & & \\
4 & 10 & 9 & 7 & 0 & \\
5 & 9 & 8 & 5 & 4 & 0
\end{array}
$$



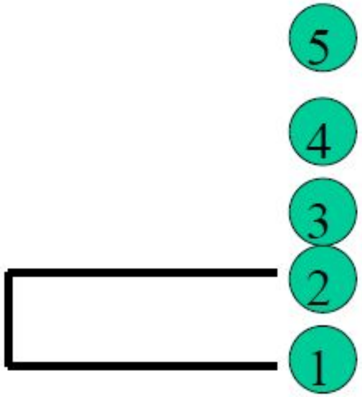
Example: single link

$$
\begin{array}{c|ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & 0 & & & & \\
2 & 2 & 0 & & & \\
3 & 6 & 3 & 0 & & \\
4 & 10 & 9 & 7 & 0 & \\
5 & 9 & 8 & 5 & 4 & 0
\end{array}
\quad \Rightarrow \quad
\begin{array}{c|cccc}
 & (1,2) & 3 & 4 & 5 \\
\hline
(1,2) & 0 & & & \\
3 & 3 & 0 & & \\
4 & 9 & 7 & 0 & \\
5 & 8 & 5 & 4 & 0
\end{array}
$$

$$d_{(1,2),3} = \min\{d_{1,3}, d_{2,3}\} = \min\{6,3\} = 3$$

$$d_{(1,2),4} = \min\{d_{1,4}, d_{2,4}\} = \min\{10,9\} = 9$$

$$d_{(1,2),5} = \min\{d_{1,5}, d_{2,5}\} = \min\{9,8\} = 8$$

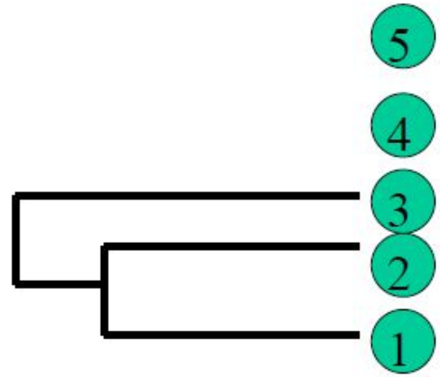


Example: single link

$$
\begin{array}{c|ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & 0 & & & & \\
2 & 2 & 0 & & & \\
3 & 6 & 3 & 0 & & \\
4 & 10 & 9 & 7 & 0 & \\
5 & 9 & 8 & 5 & 4 & 0
\end{array}
\Rightarrow
\begin{array}{c|cccc}
 & (1,2) & 3 & 4 & 5 \\
\hline
(1,2) & 0 & & & \\
3 & 3 & 0 & & \\
4 & 9 & 7 & 0 & \\
5 & 8 & 5 & 4 & 0
\end{array}
\Rightarrow
\begin{array}{c|ccc}
 & (1,2,3) & 4 & 5 \\
\hline
(1,2,3) & 0 & & \\
4 & 7 & 0 & \\
5 & 5 & 4 & 0
\end{array}
$$

$$d_{(1,2,3),4} = \min\{d_{(1,2),4}, d_{3,4}\} = \min\{9,7\} = 7$$

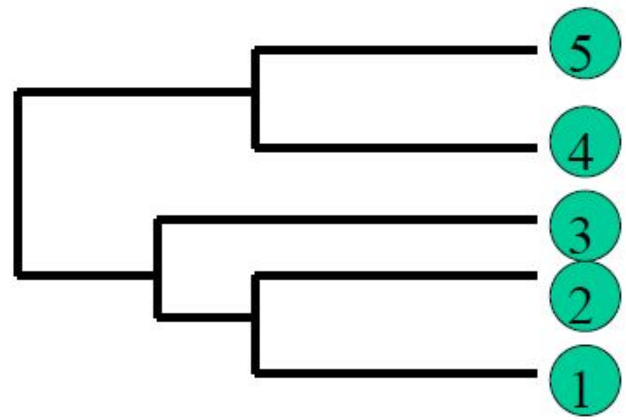
$$d_{(1,2,3),5} = \min\{d_{(1,2),5}, d_{3,5}\} = \min\{8,5\} = 5$$



Example: single link

$$
\begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & 0 & & & & \\ 2 & 2 & 0 & & & \\ 3 & 6 & 3 & 0 & & \\ 4 & 10 & 9 & 7 & 0 & \\ 5 & 9 & 8 & 5 & 4 & 0 \end{array}
\Rightarrow
\begin{array}{c|cccc} & (1,2) & 3 & 4 & 5 \\ \hline (1,2) & 0 & & & \\ 3 & 3 & 0 & & \\ 4 & 9 & 7 & 0 & \\ 5 & 8 & 5 & 4 & 0 \end{array}
\Rightarrow
\begin{array}{c|ccc} & (1,2,3) & 4 & 5 \\ \hline (1,2,3) & 0 & & \\ 4 & 7 & 0 & \\ 5 & 5 & 4 & 0 \end{array}
$$

$$d_{(1,2,3),(4,5)} = \min\{d_{(1,2,3),4}, d_{(1,2,3),5}\} = 5$$



Example: A Step-by-Step

The Dataset

Let’s consider a simple dataset containing five points in a 2D space:

Point	x	y
A	1	2
B	2	3
C	3	1
D	5	4
E	6	5

Example: A Step-by-Step

Step 1: Calculate Pairwise Distances

We will calculate the Euclidean distance between each pair of points using the formula:

$$d(P_i, P_j) = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$$

Pair	Distance
A, B	$d(A,B) = \sqrt{(2-1)^2 + (3-2)^2} = \sqrt{2} \approx 1.41$
A, C	$d(A,C) = \sqrt{(3-1)^2 + (1-2)^2} = \sqrt{5} \approx 2.24$
A, D	$d(A,D) = \sqrt{(5-1)^2 + (4-2)^2} = \sqrt{20} \approx 4.47$
A, E	$d(A,E) = \sqrt{(6-1)^2 + (5-2)^2} = \sqrt{34} \approx 5.83$
B, C	$d(B,C) = \sqrt{(3-2)^2 + (1-3)^2} = \sqrt{5} \approx 2.24$
B, D	$d(B,D) = \sqrt{(5-2)^2 + (4-3)^2} = \sqrt{10} \approx 3.16$
B, E	$d(B,E) = \sqrt{(6-2)^2 + (5-3)^2} = \sqrt{20} \approx 4.47$
C, D	$d(C,D) = \sqrt{(5-3)^2 + (4-1)^2} = \sqrt{13} \approx 3.61$
C, E	$d(C,E) = \sqrt{(6-3)^2 + (5-1)^2} = \sqrt{25} = 5$
D, E	$d(D,E) = \sqrt{(6-5)^2 + (5-4)^2} = \sqrt{2} \approx 1.41$

	A	B	C	D	E
A	0				
B	1.41	0			
C	2.24	2.24	0		
D	4.47	3.16	3.64	0	
E	5.83	4.47	5	1.41	0

Example: A Step-by-Step (Single Linkage)

Single Linkage

- Definition: The distance between two clusters is defined as the minimum distance between points in the two clusters.

Merging Process:

1. **Merge A and B** (distance = 1.41).
2. New clusters: {AB}, C, D, E. Update the distance matrix.
3. Calculate distances from cluster {AB} to remaining points:

- d(AB,C)=min(d(A,C),d(B,C))=min(2.24,2.24)=2.24
- d(AB,D)=min(d(A,D),d(B,D))=min(4.47,3.16)=3.16
- d(AB,E)=min(d(A,E),d(B,E))=min(5.83,4.47)=4.47

4-Merge {AB} and C (distance = 2.24).
5-New clusters: {ABC}, D, E. Update distances:

d(ABC,D)=min(d(AB,D),d(C,D))=min(3.16,3.61)=3.16
d(ABC,E)=min(d(AB,E),d(C,E))=min(4.47,5.00)=4.47

Pair	Distance
{AB}, C	2.24
{AB}, D	3.16
{AB}, E	4.47
C, D	3.61
C, E	5.00
D, E	1.41

Pair	Distance
{ABC}, D	3.16
{ABC}, E	4.47
D, E	1.41

Example: A Step-by-Step (Single Linkage)

Single Linkage

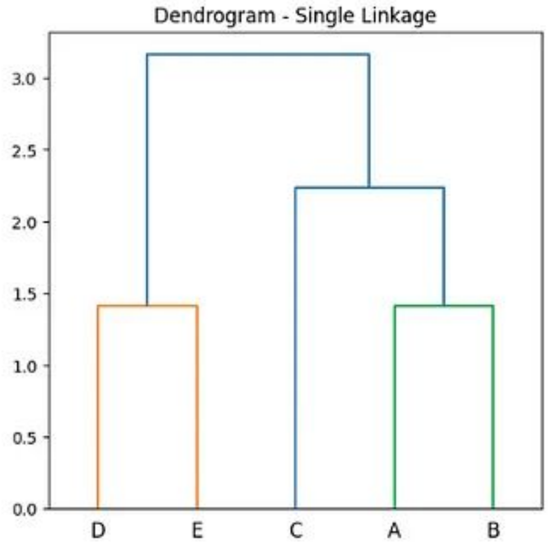
- Definition: The distance between two clusters is defined as the minimum distance between points in the two clusters.

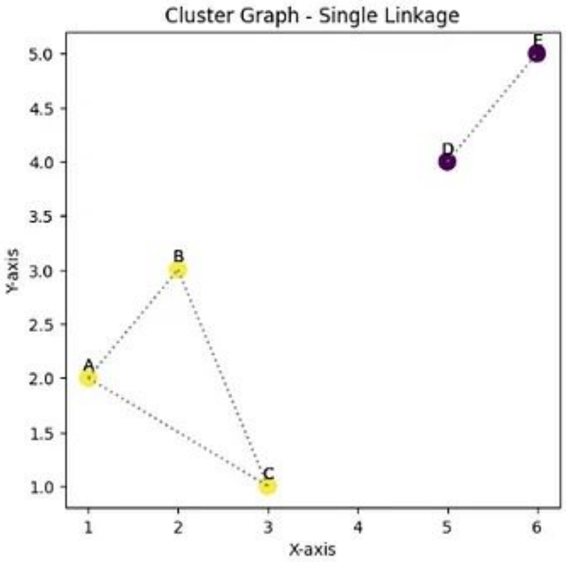
6-Merge D and E (distance = 1.41).

7- New clusters: {ABC}, {DE}.

8- Merge {ABC} and {DE} (distance = max distance between D and E).

Final Cluster Structure: {A, B, C}, {D, E}





Example: A Step-by-Step (Complete Linkage)

2. Complete Linkage

Definition: The distance between two clusters is defined as the maximum distance between points in the two clusters.

Merging Process:

1. Merge A and B (distance = 1.41).
2. New clusters: {AB}, C, D, E. Update distances.
3. Calculate distances from cluster {AB} to remaining points:

d(AB,C)=max(d(A,C),d(B,C))=max(2.24,2.24)=2.24

d(AB,D)=max(d(A,D),d(B,D))=max(4.47,3.16)=4.47

d(AB,E)=max(d(A,E),d(B,E))=max(5.83,4.47)=5.83

Pair	Distance
{AB}, C	2.24
{AB}, D	4.47
{AB}, E	5.83
C, D	3.61
C, E	5.00
D, E	1.41

4.Merge {AB} and C (distance = 2.24).

5.New clusters: {ABC}, D, E. Update distances:

d(ABC,D)=max(d(AB,D),d(C,D))=max(4.47,3.61)=4.47

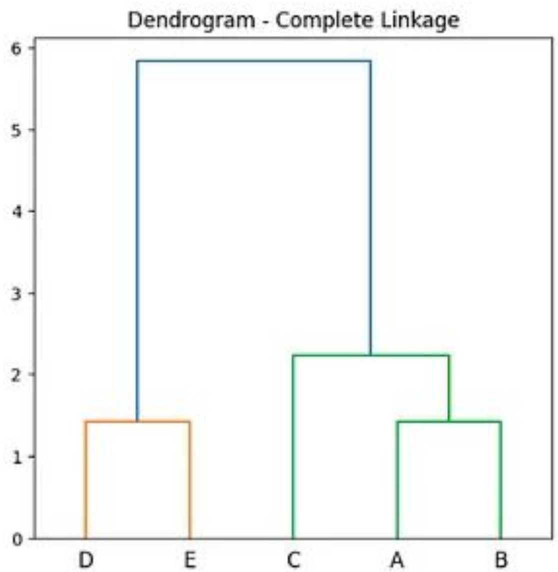
d(ABC,E)=max(d(AB,E),d(C,E))=max(5.83,5.00)=5.83

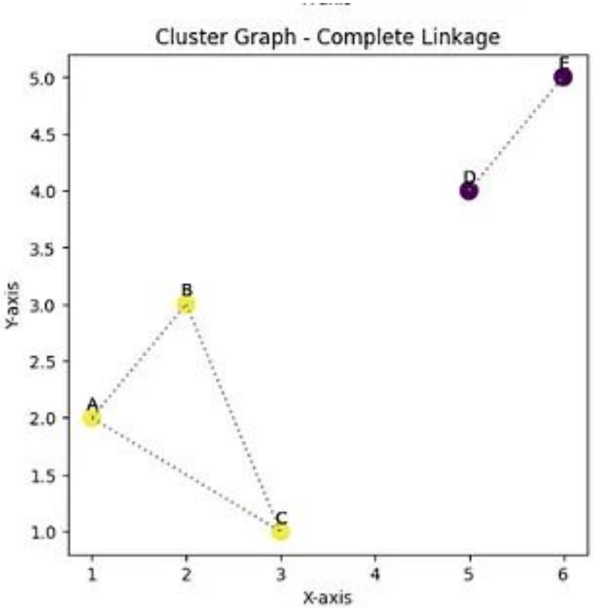
Pair	Distance
{ABC}, D	4.47
{ABC}, E	5.83
D, E	1.41

Example: A Step-by-Step (Complete Linkage)

2. Complete Linkage
Definition: The distance between two clusters is defined as the maximum distance between points in the two clusters.

6-Merge D and E (distance = 1.41).
7- New clusters: {ABC}, {DE}.
8- Merge {ABC} and {DE} (distance = max distance between D and E).
Final Cluster Structure: {A, B, C}, {D, E}





Example: A Step-by-Step (Average Linkage)

3. Average Linkage
Definition: The distance between two clusters is defined as the average distance between points in the two clusters.

Merging Process:
1.Merge A and B (distance = 1.41).
2.New clusters: {AB}, C, D, E. Update distances.
3.Calculate distances from cluster {AB} to remaining points:

- $d(AB, C) = \frac{d(A,C)+d(B,C)}{2} = \frac{2.24+2.24}{2} = 2.24$
- $d(AB, D) = \frac{d(A,D)+d(B,D)}{2} = \frac{4.47+3.16}{2} \approx 3.82$
- $d(AB, E) = \frac{d(A,E)+d(B,E)}{2} = \frac{5.83+4.47}{2} \approx 5.15$

Pair	Distance
{AB}, C	2.24
{AB}, D	3.82
{AB}, E	5.15
C, D	3.61
C, E	5.00
D, E	1.41

4- Merge {AB} and C (distance = 2.24).

5-New clusters: {ABC}, D, E. Update distances:

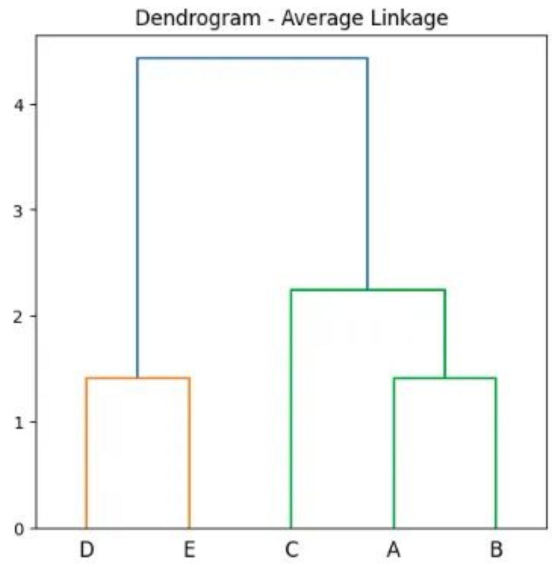
- $d(ABC, D) = \frac{d(AB,D)+d(C,D)}{2} = \frac{3.82+3.61}{2} \approx 3.72$
- $d(ABC, E) = \frac{d(AB,E)+d(C,E)}{2} = \frac{5.15+5.00}{2} = 5.08$

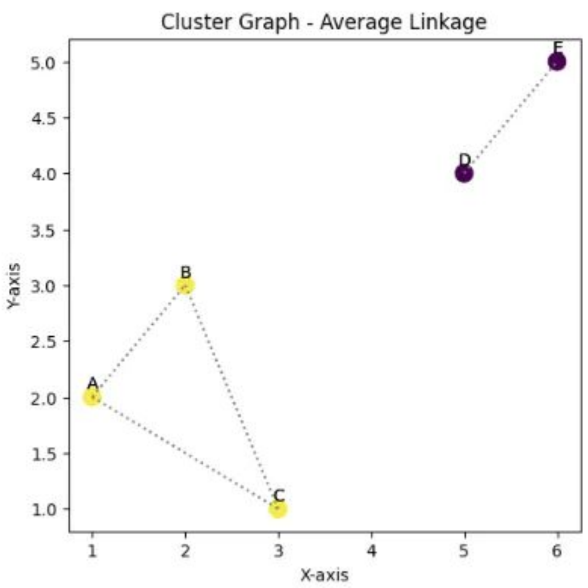
Pair	Distance
{ABC}, D	3.72
{ABC}, E	5.08
D, E	1.41

Example: A Step-by-Step (Average Linkage)

3. Average Linkage
Definition: The distance between two clusters is defined as the average distance between points in the two clusters.

6- Merge D and E (distance = 1.41).
7- New clusters: {ABC}, {DE}.
8- Merge {ABC} and {DE} (distance = average of distances).
Final Cluster Structure: {A, B, C}, {D, E}





Space and Time Complexity of Hierarchical clustering Technique:

Space complexity: The space required for the Hierarchical clustering Technique is very high when the number of data points are high as we need to store the similarity matrix in the RAM.

Space Complexity is $O(n^2)$, where n is # of data (similarity matrix)

Time complexity: Since we've to perform n iterations and in each iteration, we need to update the similarity matrix and restore the matrix, the time complexity is also very high. The time complexity is the order of the cube of n.

Time Complexity is, naively, $O(N^3)$, where n is # of data (similarity matrix)

Clustering methods: Comparison

	Hierarchical	**K-means**
Running time	naively, $O(N^3)$	fastest (each iteration is linear)
Assumptions	requires a similarity / distance measure	strong assumptions
Input parameters	none	K (number of clusters)
Clusters	subjective (only a tree is returned)	exactly K clusters