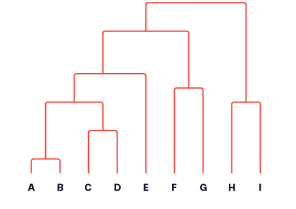
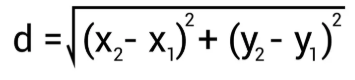
Dendrogram:

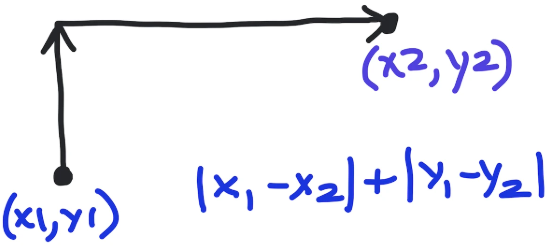
* Hierarchical clustering is usually accompanied by dendrogram. It indicates both the similarity and the order that the clusters are formed.
* The cluster which has the shortest branch was the first formed cluster. Likewise, the second shortest branch is called as cluster 2 and so on.
* For similarity measure is arbitrarily chosen but most cases we use **Euclidian distance.** Here the similarity between data points (rows) are found.



* Euclidean distance formula



* Also, we have Manhattan distance



* The clusters are formed by using various techniques.
  + **Single Linkage** (Minimum distance)
  + **Complete Linkage** (Maximum distance)
  + **Average Linkage** (Mean distance)
  + **Centroid Linkage** (Distance between centroids)
  + **Ward’s Method** (Minimizes variance within clusters)

Example (Using complete linkage)

**Complete Linkage Hierarchical Clustering Example**

**Dataset**

We have four data points, each with three features:

| **Point** | **Feature 1 (X)** | **Feature 2 (Y)** | **Feature 3 (Z)** |
| --- | --- | --- | --- |
| A | 2.0 | 3.0 | 5.0 |
| B | 3.0 | 5.0 | 8.0 |
| C | 5.0 | 8.0 | 12.0 |
| D | 6.0 | 9.0 | 14.0 |

**Step 1: Compute the Distance Matrix**

Using Euclidean distance formula:

d(A,B)=(x2−x1)2+(y2−y1)2+(z2−z1)2d(A, B) = \sqrt{(x\_2 - x\_1)^2 + (y\_2 - y\_1)^2 + (z\_2 - z\_1)^2}

|  | **A** | **B** | **C** | **D** |
| --- | --- | --- | --- | --- |
| A | 0 | 3.74 | 9.11 | 11.53 |
| B | 3.74 | 0 | 5.39 | 7.81 |
| C | 9.11 | 5.39 | 0 | 2.45 |
| D | 11.53 | 7.81 | 2.45 | 0 |

**Step 2: Perform Clustering Using Complete Linkage**

Complete linkage defines the distance between clusters as the **maximum** distance between any two points in the clusters.

**First Merge: C and D (Minimum Distance = 2.45)**

New distance calculations:

* d({C,D},A)=max⁡(d(C,A),d(D,A))=11.53d(\{C,D\}, A) = \max(d(C, A), d(D, A)) = 11.53
* d({C,D},B)=max⁡(d(C,B),d(D,B))=7.81d(\{C,D\}, B) = \max(d(C, B), d(D, B)) = 7.81

|  | **A** | **B** | **{C,D}** |
| --- | --- | --- | --- |
| A | 0 | 3.74 | 11.53 |
| B | 3.74 | 0 | 7.81 |
| {C,D} | 11.53 | 7.81 | 0 |

**Second Merge: A and B (Minimum Distance = 3.74)**

New distance calculation:

* d({A,B},{C,D})=max⁡(d(A,{C,D}),d(B,{C,D}))=11.53d(\{A,B\}, \{C,D\}) = \max(d(A, \{C,D\}), d(B, \{C,D\})) = 11.53

|  | **{A,B}** | **{C,D}** |
| --- | --- | --- |
| {A,B} | 0 | 11.53 |
| {C,D} | 11.53 | 0 |

**Final Merge: {A,B} and {C,D}**

Final cluster distance = **11.53**.

**Step 3: Dendrogram Representation**

┌────────── {A,B} ───────────┐

│ │

│ │

└────────── {C,D} ───────────┘

**Summary of Merging Steps**

| **Merge Step** | **Clusters Merged** | **Distance** |
| --- | --- | --- |
| **1st Merge** | {C} & {D} | 2.45 |
| **2nd Merge** | {A} & {B} | 3.74 |
| **Final Merge** | {A,B} & {C,D} | 11.53 |

**Conclusion**

* **Complete Linkage** uses the **maximum distance** between clusters to determine merges.
* It forms **compact, well-separated clusters**.
* It avoids elongated chain-like clusters, unlike **Single Linkage**.
* The final clustering structure is determined at distance **11.53**.

This process can be visualized using a **dendrogram**, which helps in deciding the optimal number of clusters.

**End of Document**