

Hierarchical Clustering Tutorial

September 17, 2025

1 Introduction

Hierarchical clustering builds a multi-resolution view of data by successively merging or splitting clusters without pre-specifying their number. Agglomerative (bottom-up) methods start with singletons and iteratively merge the closest clusters using linkage criteria such as single, complete, average, or Ward's linkage. The resulting dendrogram captures nested structure, allowing practitioners to choose cluster counts post hoc by cutting the tree at different heights.

2 Theory and Formulas

2.1 Linkage Functions

For two clusters A and B with elements \mathbf{x}_i and \mathbf{x}_j , popular linkage definitions include

$$\text{single}(A, B) = \min_{\mathbf{x}_i \in A, \mathbf{x}_j \in B} \|\mathbf{x}_i - \mathbf{x}_j\|_2, \quad (1)$$

$$\text{complete}(A, B) = \max_{\mathbf{x}_i \in A, \mathbf{x}_j \in B} \|\mathbf{x}_i - \mathbf{x}_j\|_2, \quad (2)$$

$$\text{average}(A, B) = \frac{1}{|A||B|} \sum_{\mathbf{x}_i \in A} \sum_{\mathbf{x}_j \in B} \|\mathbf{x}_i - \mathbf{x}_j\|_2, \quad (3)$$

$$\text{Ward}(A, B) = \frac{|A||B|}{|A| + |B|} \|\boldsymbol{\mu}_A - \boldsymbol{\mu}_B\|_2^2, \quad (4)$$

where $\boldsymbol{\mu}_A$ and $\boldsymbol{\mu}_B$ are centroids. Ward's linkage minimizes the increase in within-cluster variance at each merge.

2.2 Agglomerative Algorithm

A generic agglomerative procedure is:

1. Initialize each point as a singleton cluster and compute all pairwise distances.
2. Repeatedly merge the pair of clusters with minimal linkage distance.
3. Update the distance matrix using the chosen linkage rule.
4. Continue until all samples form one cluster or a stopping criterion is met.

The merge history forms a hierarchy encoded in the dendrogram's branching structure. Cutting the dendrogram at height h yields clusters with maximum within-cluster dissimilarity bounded by h .

2.3 Cophenetic Distance

The cophenetic distance between points i and j is the dendrogram height at which they merge. Comparing the cophenetic correlation with the original distance matrix helps assess how faithfully the hierarchy preserves pairwise relationships.

3 Applications and Tips

- **Taxonomy and phylogeny:** dendrograms naturally express hierarchical relationships among species or document categories.
- **Market basket analysis:** hierarchical clustering on products uncovers nested groupings useful for recommendation and shelf planning.
- **Feature engineering:** dendrogram cuts at different heights provide multi-scale cluster labels or grouping constraints for other algorithms.
- **Best practices:** standardize features, choose linkage consistent with data geometry, and watch for chaining (single linkage) or sensitivity to outliers (complete linkage).

4 Python Practice

The script `gen_clustering_hierarchical_clustering_figures.py` generates synthetic blobs, computes an agglomerative clustering, and saves both a dendrogram and a 2D scatter plot colored by cluster labels obtained from cutting the tree.

Listing 1: Excerpt from `gen_clustering_hierarchical_clustering_figures.py`

```
1 from scipy.cluster.hierarchy import dendrogram, linkage, fcluster
2
3 linkage_matrix = linkage(points, method="ward")
4 labels = fcluster(linkage_matrix, t=3, criterion="maxclust")
5
6 # Plot dendrogram and colored scatter of assigned clusters
```

5 Result

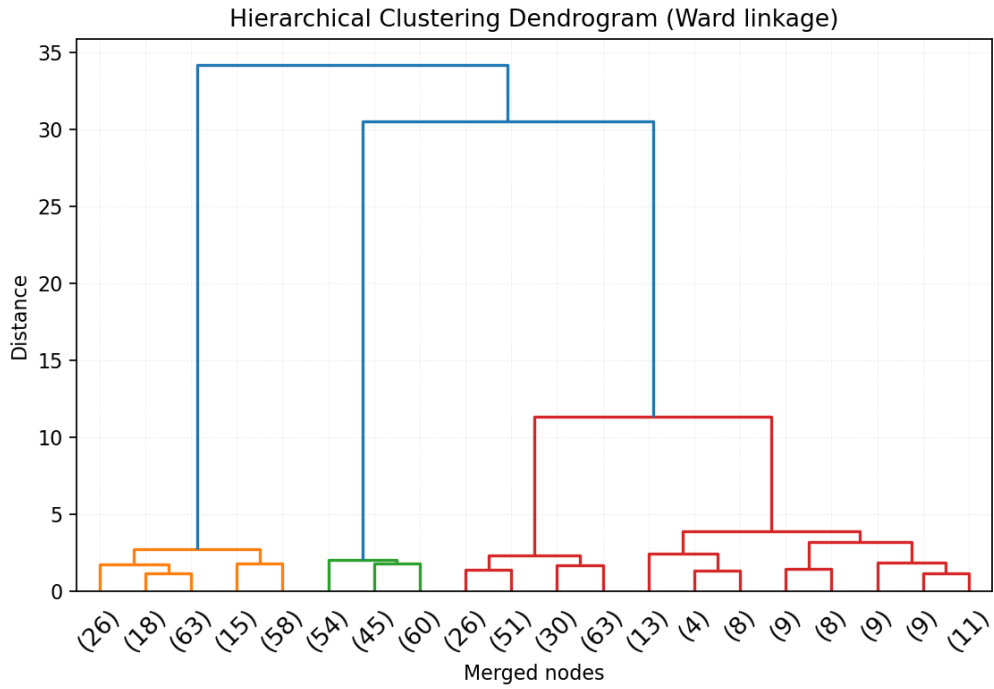


Figure 1: Dendrogram produced by Ward linkage on synthetic blobs

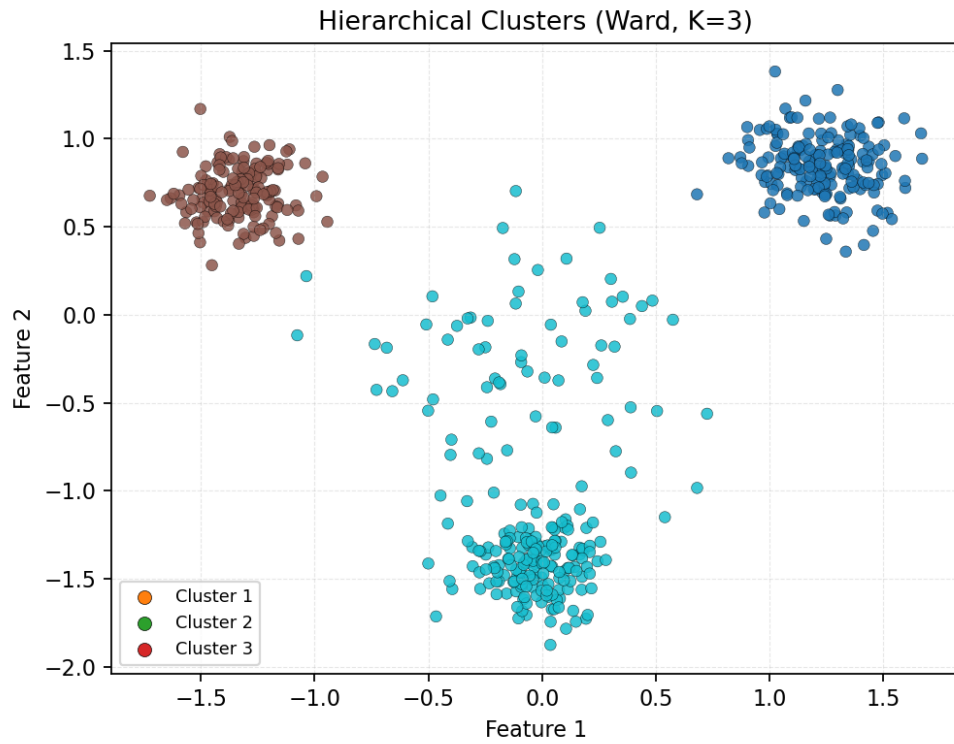


Figure 2: Cluster assignments obtained by cutting the dendrogram at three clusters

6 Summary

Hierarchical clustering reveals nested structure without committing to a fixed number of clusters upfront. Linkage choices control behavior: Ward favors compact clusters, average balances chaining and tightness, while single and complete emphasize connectivity extremes. The example demonstrates how dendrogram visualization and cluster cuts complement one another in exploratory analysis.