

Minkowski distance family

- $r = 1$ (L1): Manhattan; Hamming for binary.
- $r = 2$ (L2): Euclidean.
- $r \rightarrow \infty$ (L ∞): Chebyshev (max component difference).

Hierarchical clustering

- Produces a nested tree of clusters; visualized via a dendrogram.
- Typical applications: gene clustering, phylogeny, taxonomies.

Agglomerative procedure

1. Compute pairwise similarity/distance matrix.
2. Initialize: each point is a cluster.
3. Iteratively merge the closest two clusters.
4. Update the matrix using a linkage rule.
5. Repeat until one cluster remains.

Linkage (distance update) options

Linkage	Definition	Behavior
Single (Min)	min pairwise distance between clusters	Chaining
Complete (Max)	max pairwise distance	Compact, outlier-sensitive
Average	mean of all pairwise distances (UPGMA)	Balanced
Centroid	distance between centroids	Can invert dendrogram
Ward's	minimizes increase in within-cluster SSE	Compact/spherical

Running example: gene expression (5 genes $\times$ 4 conditions)

Gene	wt	mutant_1	mutant_2	mutant_3
At4g35770	1.5	3	3	1.5
At1g30720	4	7.5	7.5	5
At4g27450	1.5	1	1	1.5
At2g34930	10	25	23	15
At2g05540	1	1	2	1

- Similarity used: Pearson correlation (pattern-oriented, scale-invariant).

Pairwise correlations (given)

	At4g35770	At1g30720	At4g27450	At2g34930	At2g05540
At4g35770	1	0.9733	-1	0.9493	0.5774
At1g30720	0.9733	1	-0.9733	0.9909	0.562
At4g27450	-1	-0.9733	1	-0.9493	-0.5774
At2g34930	0.9493	0.9909	-0.9493	1	0.4528
At2g05540	0.5774	0.562	-0.5774	0.4528	1

Agglomerative merges (treat “closest” as highest correlation; or use distance = $1 - r$)

- Merge 1: At2g34930 with At1g30720 ($r = 0.9909$).
- Update via “max correlation” linkage.
- Merge 2: {At2g34930, At1g30720} with At4g35770 ($r = 0.9733$).
- Update: cluster vs At4g27450 = -0.9493 ; vs At2g05540 = 0.5774 .
- Merge 3: cluster with At2g05540 ($r = 0.5774$).
- Final: remaining At4g27450 joins last (strong negative correlation to the main group).
- Interpretation: a positively correlated group {At2g34930, At1g30720, At4g35770, At2g05540}; At4g27450 shows anti-pattern.

Programming resources

- Scikit-learn: <https://scikit-learn.org/stable/>
- Example dendrogram for agglomerative clustering:
<https://scikit->

[learn.org/stable/auto_examples/cluster/plot_agglomerative_dendrogram.html#sphx-glr-auto-examples-cluster-plot-agglomerative-dendrogram-py](https://www.learn.org/stable/auto_examples/cluster/plot_agglomerative_dendrogram.html#sphx-glr-auto-examples-cluster-plot-agglomerative-dendrogram-py)

Advanced topic: Mahalanobis distance

- Definition: $d_M(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)}$, Σ = covariance matrix.
- Purpose: accounts for feature scales and correlations (distribution-aware).
- Practical notes: estimate Σ and use Σ^{-1} ; regularize if singular/ill-conditioned (e.g., shrinkage).

Normalization note (brief)

- Min-max normalization rescales features but doesn't model correlations; Mahalanobis incorporates both variance and covariance.