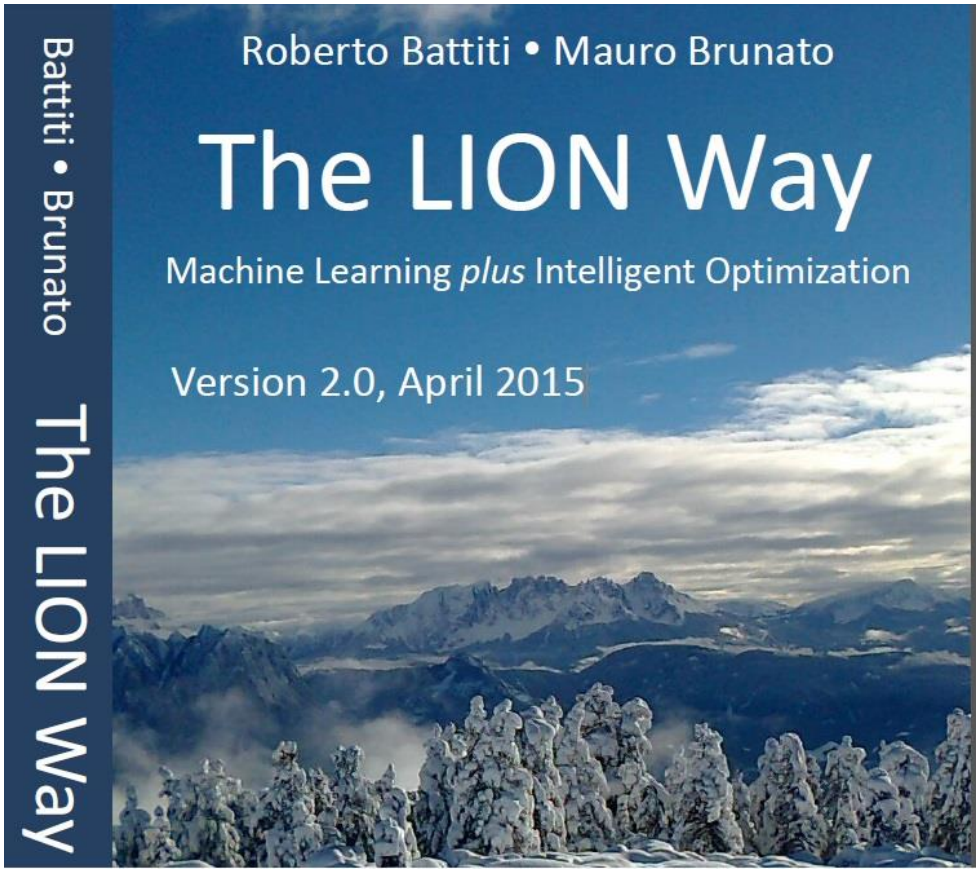


ROBERTO BATTITI, MAURO BRUNATO.
The LION Way: Machine Learning plus *Intelligent Optimization*.
LIONlab, University of Trento, Italy,
Apr 2015

http://intelligent-optimization.org/LIONbook

Bottom-up (agglomerative) clustering

Birds of a feather flock together.

Agglomerative clustering: definition

- **Hierarchical algorithms** find successive clusters by **merging** previously established smaller clusters,
  - Begin with each element as a separate cluster.
  - At each step the most similar clusters are merged. Note: This requires a measure of similarity between two clusters. (or between a cluster and a single element).

Distance between clusters

- A distance between clusters can be derived from a distance between elements

Three main choices:

$$\overline{\delta}_{ave}(C,D) = \frac{\sum_{x\in C,\, y\in D}\delta(x,y)}{|C|\cdot|D|};$$

$$\overline{\delta}_{min}(C,D) = \min_{x\in C,\, y\in D}\delta(x,y);$$

$$\overline{\delta}_{max}(C,D) = \max_{x\in C,\, y\in D}\delta(x,y).$$

C,D are two clusters, δ(x,y) is the distance between the elements x and y.

Merging algorithm

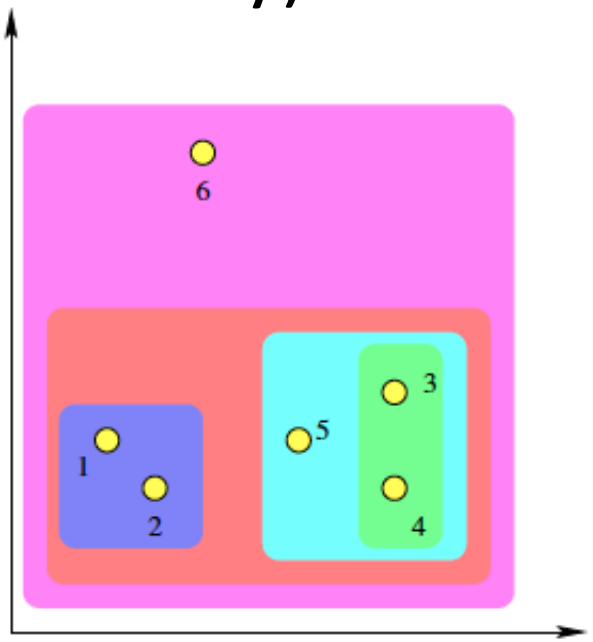
- Given a set of clusters *C,* proceed as follows:

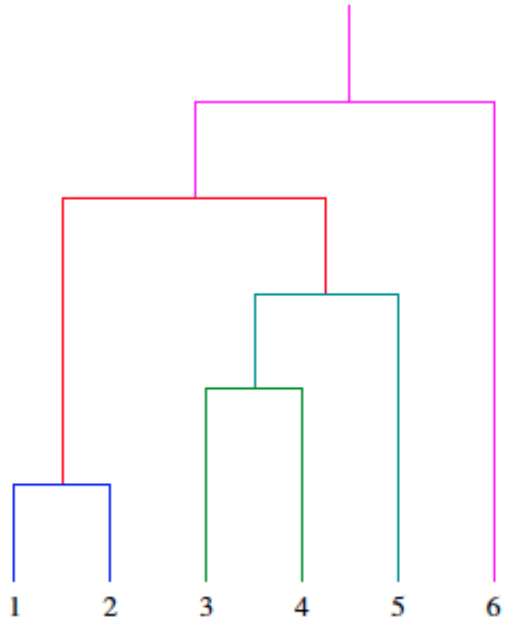
1. Find clusters C and D in *C* with minimum distance $\overline{\delta}^* = \min_{C \neq D} \overline{\delta}(C, D)$
2. substitute C and D with their **union** C **U** D, and register δ* as the distance for which the specific merging occurred;

until a single cluster containing all entities is obtained

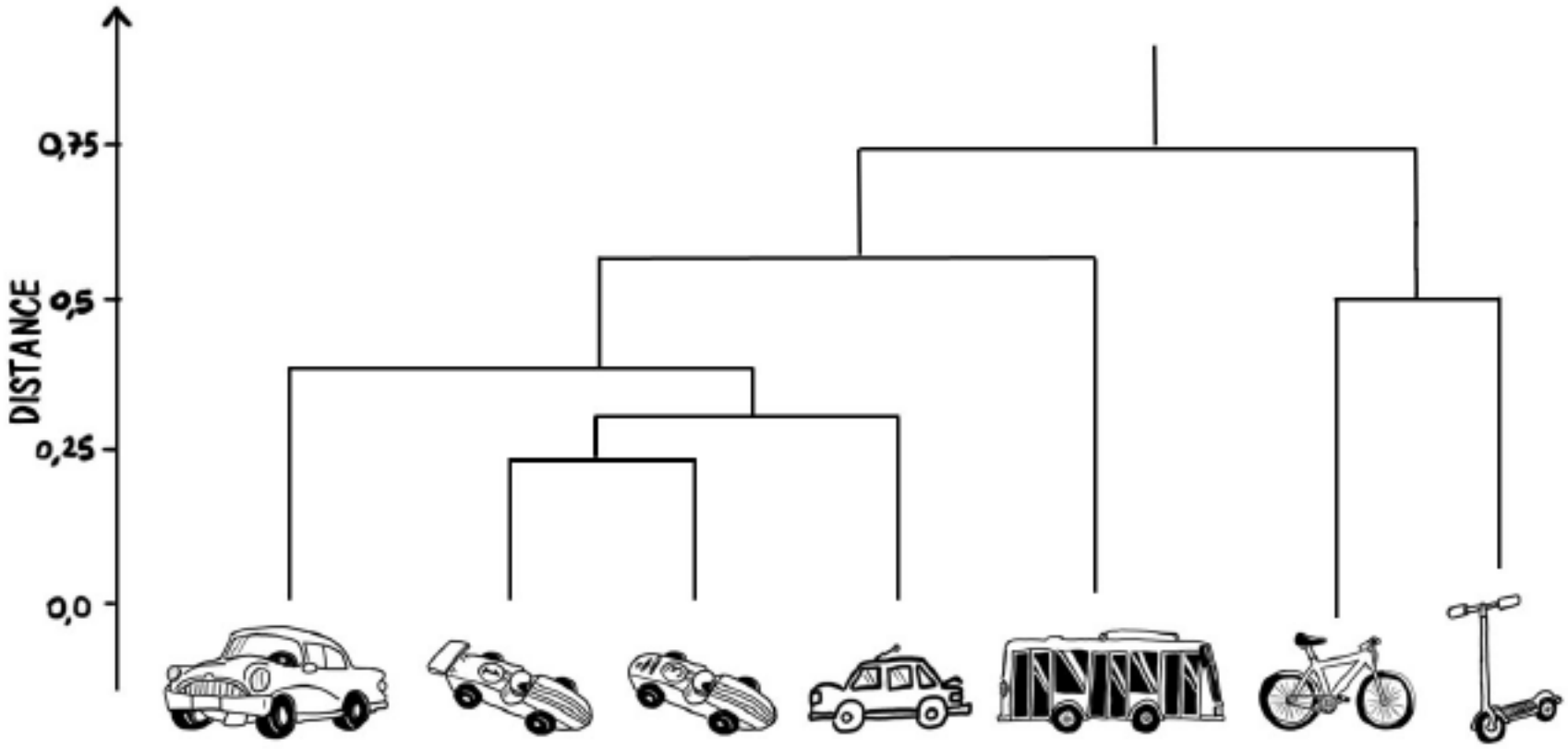
Dendrogram

- Hierarchical merging can be visualized through dendrograms, where the original entities are at the bottom and each merging action is represented with a horizontal line connecting the two fused clusters (at a level given by the dissimilarity)

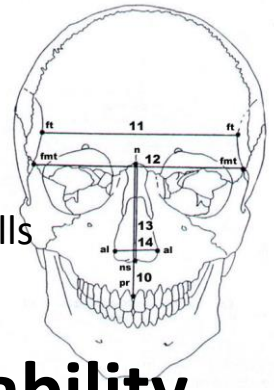




A dendrogram for vehicles



Mahalanobis distance



Prompted by the problem of identifying similarities of skulls based on measurements in 1927

Given a point x, how do we define the **probability for the point to belong to a cluster**?

For a symmetric spherical distribution:

- Evaluate the standard deviation σ of the distances of the sample points
- Evaluate its distance from the average of points (center of mass)
- Define the normalized distance $|\mathbf{x}-\boldsymbol{\mu}|/\sigma$
- Derive the probability of the test point belonging to the set from the normal distribution

Mahalanobis distance (2)

- If the distribution is highly *non*-spherical the probability of the test point belonging to the set will depend also on the **direction** of the vector ($\mathbf{x}$- $\boldsymbol{\mu}$)

Mahalanobis distance (3)

In the case on the left we can use the Euclidean distance as a dissimilarity measure, while in the other case we need to refer to the **Mahalanobis distance**, because the data are distributed in an ellipsoidal shape.

Mahalanobis distance (4)

- The ellipsoid that best represents the set's probability distribution can be estimated by building the **covariance matrix** of the samples
- The center p of the cluster is the mean value:

$$\bar{\boldsymbol{p}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i.$$

- Let the covariance matrix components be defined as:

$$S_{ij} = \frac{1}{n}\sum_{k=1}^{n}(p_{ki} - \bar{p}_i)(p_{kj} - \bar{p}_j), \qquad i, j = 1, .., D.$$

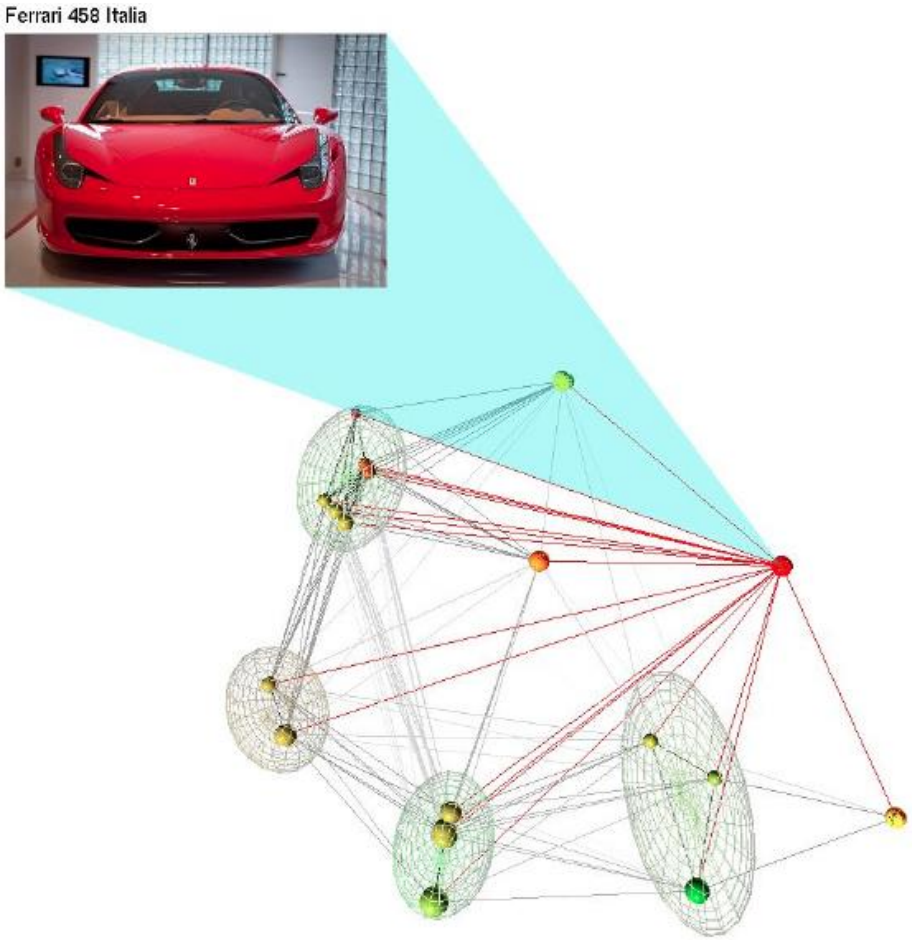
Mahanalobis distance (5)

- The Mahalanobis distance of a vector x from a set of values with mean and covariance matrix S is defined as:

$$D_M(x) = \sqrt{(x-\mu)^T S^{-1}(x-\mu)}.$$

- It is the distance of the test point from the center of mass divided by the width of the ellipsoid in the direction of the test point

Clustering visualization



Gist

- Agglomerative clustering builds a **tree** (a hierarchical organization) containing bigger and bigger clusters
- It is a “bottom up” method: first nearby points are merged, then similar sets are merged, until a single set is obtained.
- The number of clusters is not specified at the beginning
- a proper number can be obtained by cutting the tree (a.k.a. **dendrogram**) at an appropriate level of similarity