HIERARCHICAL CLUSTERING

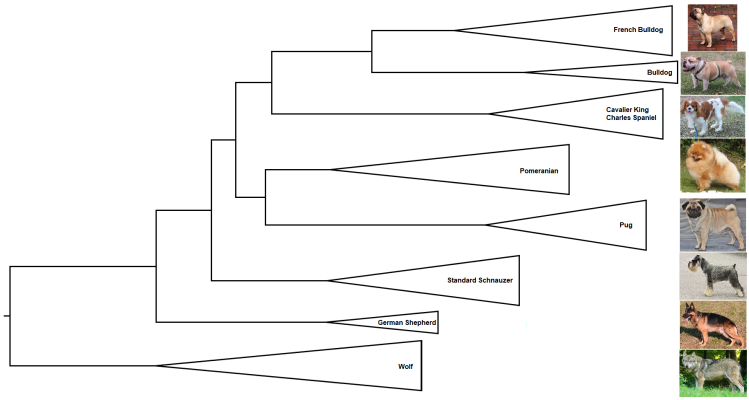
Marco Bressan

Università degli Studi di Milano

April 21, 2021

Hierarchical clustering

Instead of partitioning the input set X, we compute a **hierarchy**.



Wlements that are "similar" should be close in the tree.

Agglomerative clustering

One very popular approach is **agglomerative** clustering.

As usual suppose we have a **distance** $d : X \times X \mapsto \mathbb{R}_0^+$.

Algorithm 1: AgglomerativeClustering(X, d)

---

for *each* $x \in X$ **do** create a cluster $C_x = \{x\}$;
while *there are at least* 2 *clusters left* **do**
    find the two "closest" clusters C_i and C_j;
    create a cluster $C_i \cup C_j$ and set it as parent of C_i and C_j
end
return the resulting tree;

Agglomerative clustering

The algorithm is not well-defined until we specify what "closest clusters" means.

Single-linkage a.k.a. **nearest neighbor**

$$d(C, C') = \min_{x \in C} \min_{y \in C'} d(x, y)$$

Complete-linkage a.k.a. **farthest neighbor**

$$d(C, C') = \max_{x \in C} \max_{y \in C'} d(x, y)$$

Ward

$$d(C, C') = \sum_{\boldsymbol{x} \in C \cup C'} \|\boldsymbol{x} - \boldsymbol{\mu}(C \cup C')\|_2^2 - \left(\sum_{\boldsymbol{x} \in C} \|\boldsymbol{x} - \boldsymbol{\mu}(C)\|_2^2 + \sum_{\boldsymbol{x} \in C'} \|\boldsymbol{x} - \boldsymbol{\mu}(C')\|_2^2 \right)$$

Example: Ward clustering of European languages

```
langs <- read.csv("langfreq.csv", row.names = 1)
D <- dist(langs)
hc <- hclust(D, method = "ward.D2")
plot(as.dendrogram(hc))
cutree(…)  # try
```

Example: Ward clustering of European languages

```
langs <- read.csv("langfreq.csv", row.names = 1)
D <- dist(langs)
hc <- hclust(D, method = "ward.D2")
plot(as.dendrogram(hc))
cutree(…)  # try
```