Guided Tour of Machine Learning in Finance

Week 3: Unsupervised Learning

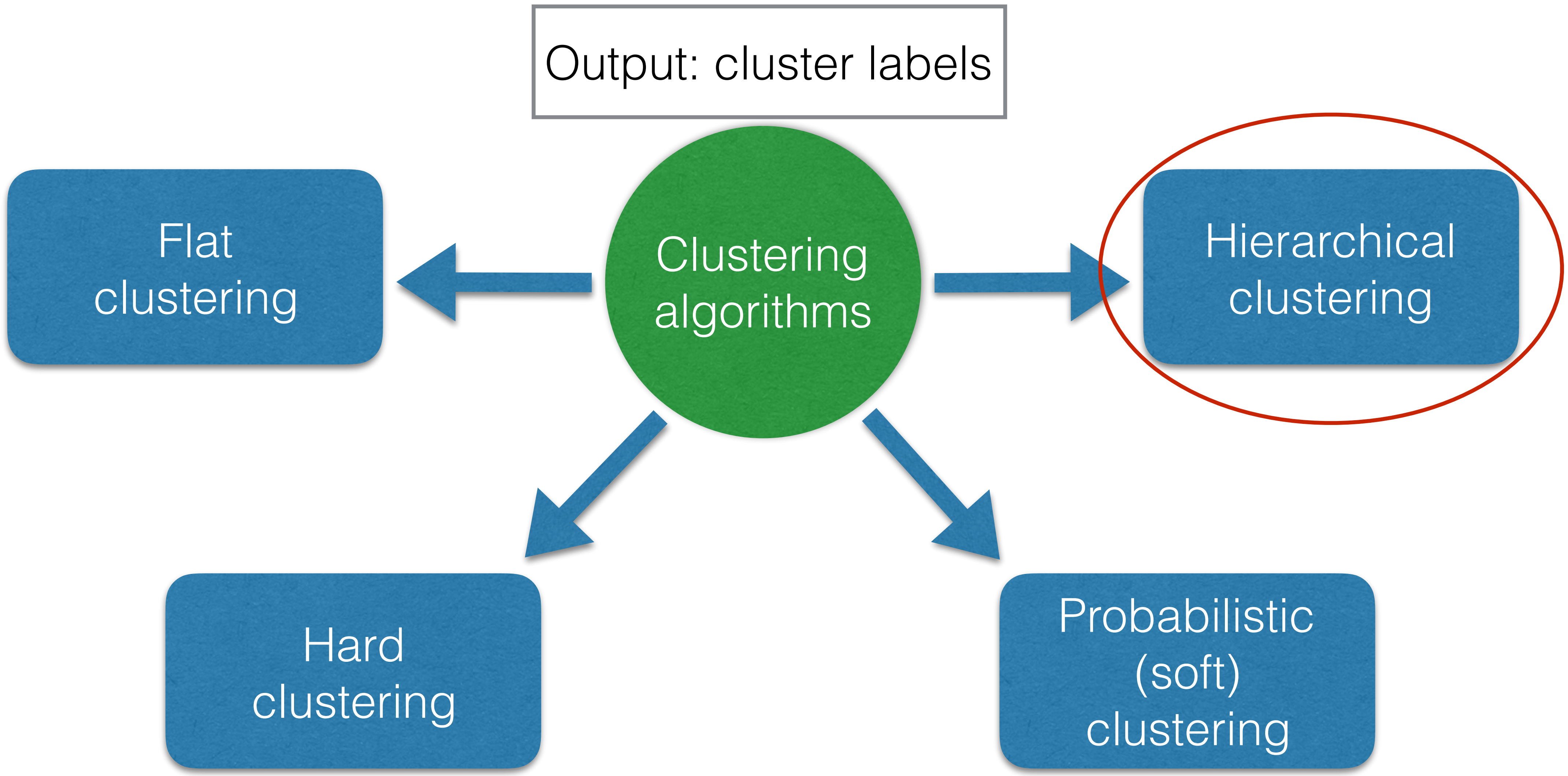
Hierarchical clustering algorithms

Igor Halperin

NYU Tandon School of Engineering, 2017

Types of clustering

Now consider hierarchical clustering methods



Why hierarchical clustering?

- Many complex physical, biological and social systems have nested hierarchical structures (sub-clusters within clusters, and so on)
- The hierarchical structure of interactions affects the dynamics of complex systems
- A quantitative description of hierarchies of the system is a key step in the modeling of complex systems, see P. Anderson (The Nobel Prize in Physics, 1977), “The More is Different”, Science 177 (1972), pp. 393-396.

Why hierarchical clustering?

- Many complex physical, biological and social systems have nested hierarchical structures (sub-clusters within clusters, and so on)
- The hierarchical structure of interactions affects the dynamics of complex systems
- A quantitative description of hierarchies of the system is a key step in the modeling of complex systems, see P. Anderson (The Nobel Prize in Physics, 1977), “The More is Different”, Science 177 (1972), pp. 393-396.

“In closing, I offer two examples from economics of what I hope to have said. Marx said that the quantitative differences become qualitative ones, but a dialogue in Paris in 1920’s sums it up even more clearly:

Why hierarchical clustering?

- Many complex physical, biological and social systems have nested hierarchical structures (sub-clusters within clusters, and so on)
- The hierarchical structure of interactions affects the dynamics of complex systems
- A quantitative description of hierarchies of the system is a key step in the modeling of complex systems, see P. Anderson (The Nobel Prize in Physics, 1977), "The More is Different", Science 177 (1972), pp. 393-396.

"In closing, I offer two examples from economics of what I hope to have said. Marx said that the quantitative differences become qualitative ones, but a dialogue in Paris in 1920's sums it up even more clearly:

FITZGERALD: The rich are different from us.

HEMINGWAY: Yes, they have more money."

Hierarchical structures in finance

- Financial markets have hierarchical features (global indexes vs sector indexes vs geographic-based indexes).
- There are multi-factor asset pricing models with a hierarchical factor structure
- We follow a ML paradigm and try to learn a hierarchical structure directly from data!
- A hierarchical factor model with independent factors at different levels of the hierarchy can be extracted from a cluster found by a ML algorithm

Agglomerative hierarchical clustering

1. Start with each data point forming its own cluster
2. Update clusters by adding to each point its closest neighbor
3. Merge the most “similar” cluster
4. Continue until convergence, or until a needed number of clusters is found

This is a “bottom-up” approach to clustering. Mergers are made in a greedy manner. Complexity in general for agglomerative clustering is $O(N^2 \log N)$, but is made $O(N^2)$ by two special cases: single linkage and complete linkage algorithms.

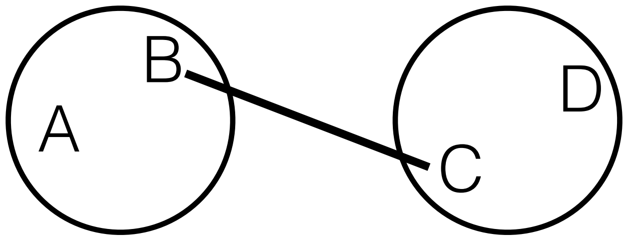
When N is large, one can first use K-means (complexity $O(KND)$), and then apply a hierarchical clustering to K-means centroids.

Choice of a metric

Metric	Definition
Euclidean distance	$\left\Vert X^{(1)} - X^{(2)} \right\Vert_2 = \sqrt{\sum_i \left(X_i^{(1)} - X_i^{(2)} \right)^2}$
Squared euclidean distance	$\left\Vert X^{(1)} - X^{(2)} \right\Vert_2^2 = \sum_i \left(X_i^{(1)} - X_i^{(2)} \right)^2$
Manhattan distance	$\left\Vert X^{(1)} - X^{(2)} \right\Vert_1 = \sum_i \left\vert X_i^{(1)} - X_i^{(2)} \right\vert$
Maximum distance	$\left\Vert X^{(1)} - X^{(2)} \right\Vert_\infty = \max_i \left\vert X_i^{(1)} - X_i^{(2)} \right\vert$
Mahalanobis distance	$\left\Vert X^{(1)} - X^{(2)} \right\Vert_2 = \sqrt{\left(X^{(1)} - X^{(2)} \right)^T \mathbf{S} \left(X^{(1)} - X^{(2)} \right)}$

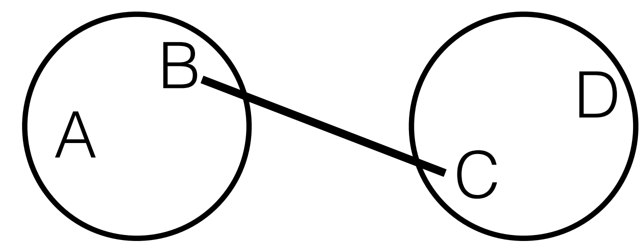
Linkage methods

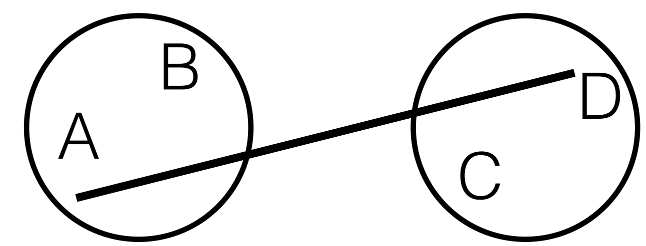
Name	Definition
Single linkage clustering	$\min \{ d(a,b) : a \in A, b \in B \}$



Linkage methods

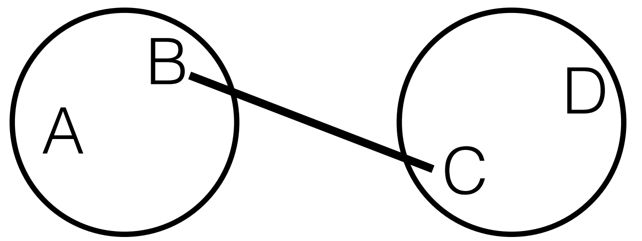
Name	Definition
Single linkage clustering	$\min\left\{d(a,b) : a \in A, b \in B\right\}$
Complete linkage clustering	$\max\left\{d(a,b) : a \in A, b \in B\right\}$

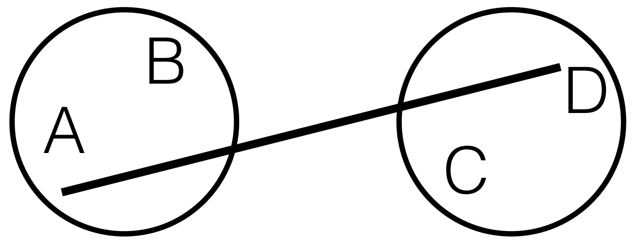


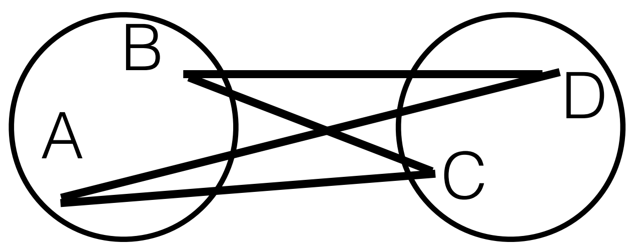


Linkage methods

Name	Definition
Single linkage clustering	$\min\left\{d(a,b) : a \in A, b \in B\right\}$
Complete linkage clustering	$\max\left\{d(a,b) : a \in A, b \in B\right\}$
Average linkage clustering	$\frac{1}{\lvert A \rvert \lvert B \rvert} \sum_{a \in A} \sum_{b \in B} d(a,b)$

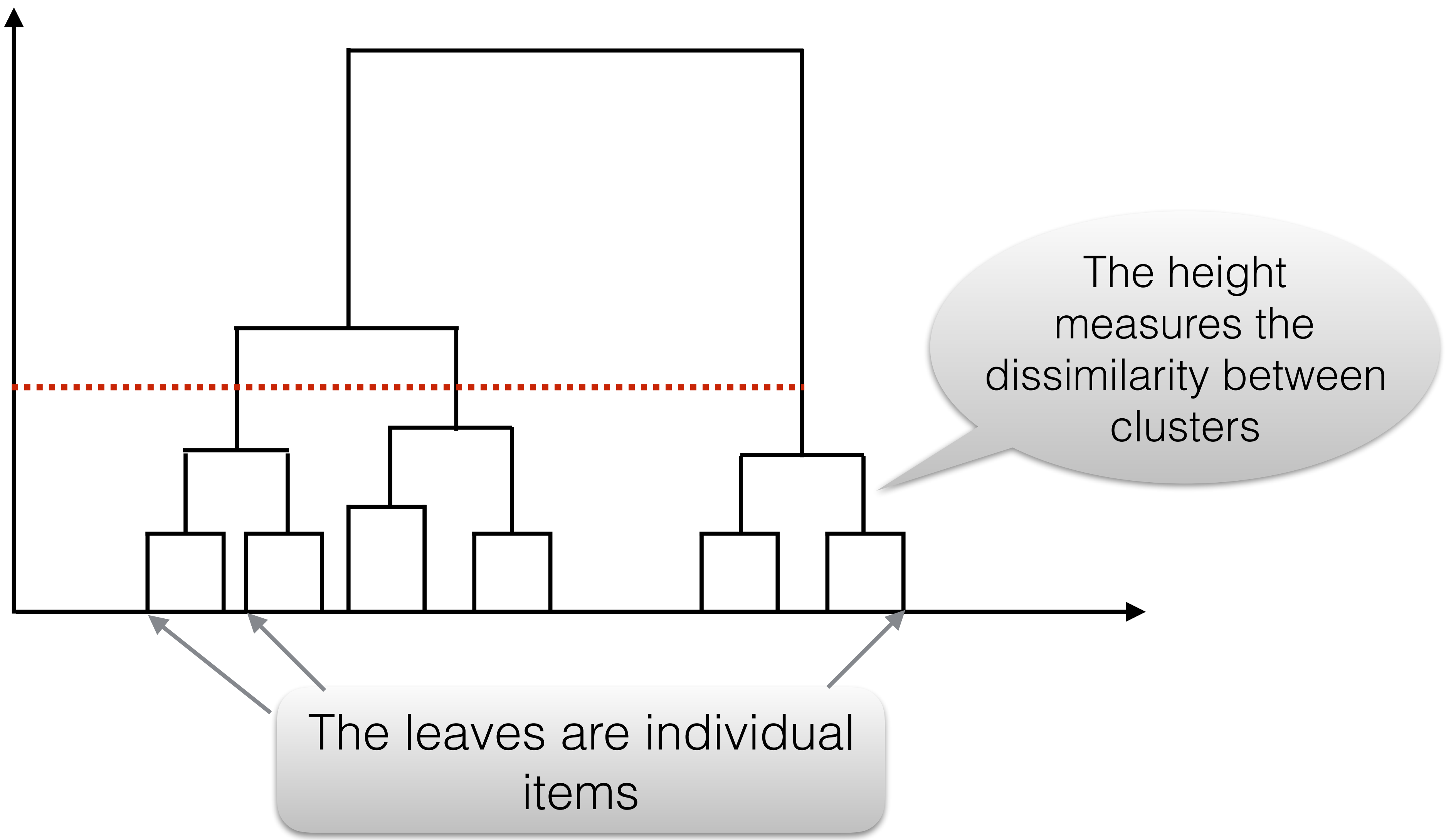






Dendrogram

Dendrogram is a binary tree representing the process of merging sub-clusters in a hierarchical cluster



Control question

Select all correct answers

1. Hierarchical clustering produces hierarchies of probability distributions for all cluster and sub-cluster labels.
2. Agglomerative clustering is a bottom-up type of hierarchical clustering that merges sub-clusters in a greedy manner.
3. While complexity of general agglomerative clustering methods is $O(N^2 \log N)$, it can be reduced to $O(N^2)$ for Single-Linkage and Complete Linkage methods.
4. For Average Linkage clustering, complexity is $O(N \log N)$, though so far it was observed only using GPU architectures.

Correct answers: 2,3