Clustering
Hierarchical Methods

Huiping Cao
Hierarchical Methods

- Use distance matrix as clustering criteria. This method does not require the number of clusters $k$ as an input, but needs a termination condition.

- **Agglomerative** hierarchical clustering; bottom-up

- **Divisive** hierarchical clustering; top-down
AGNES (Agglomerative Nesting)

- Introduced in Kaufmann and Rousseeuw (1990)
- **Start** by letting each object form its own clusters
- **Iteratively merges** clusters
  - Finds two clusters that are closest to each other
- **Termination**: all nodes belong to the same cluster
- **Analysis**: Each iteration merges two clusters. This method requires at most $n$ iterations.
DIANA (Divisive Analysis)

- Introduced in Kaufmann and Rousseeuw (1990)
- **Start** by placing all objects in one cluster
- **Iteratively divides** clusters
- **Eventually** (1) each node forms a cluster on its own, or (2) objects within a cluster are sufficiently similar to each other.
DIANA (Divisive Analysis)

- Inverse order of AGNES

- Practically, use heuristics in partitioning because there are $2^{n-1} - 1$ possible ways to partition a set of $n$ objects into two exclusive subsets.

- There are many more agglomerative methods than divisive methods.
A user can specify the desired number of clusters as a termination condition.
Dendrogram: Shows How the Clusters are Merged

- Decompose data objects into several levels of nested partitioning (tree of clusters), called a **dendrogram**.
- A clustering of the data objects is obtained by cutting the **dendrogram at the desired level**, then each connected component forms a cluster.
R: AGNES, DIANA

> library(cluster)
> help(agnes)
> help(diana)
> help(mona)

CRAN document:
https://cran.r-project.org/web/packages(cluster)/cluster.pdf

The method mona returns a list representing a divisive hierarchical clustering of a dataset with binary variables only.
Hierarchical Clustering Methods

- Major weakness of agglomerative clustering methods
  - Do not scale well: time complexity of at least $O(n^2)$, where $n$ is the number of total objects

- Further improvement
  - BIRCH (1996): uses CF-tree and incrementally adjusts the quality of sub-clusters
  - ROCK (1999): clustering categorical data by neighbor and link analysis
  - CHAMELEON (1999): hierarchical clustering using dynamic modeling
BIRCH (1996)


- Incrementally construct a **CF (Clustering Feature) tree**, a hierarchical data structure for multiphase clustering
Clustering Feature Vector in BIRCH

- **Clustering feature (CF)** \( CF = (N, LS, SS) \)
  - \( N \): Number of data points
  - \( LS = \sum_{i=1}^{N} X_i \), linear sum of all the points
  - \( SS = \sum_{i=1}^{N} X_i^2 \), square sum of all the points

- Easy to calculate centroid, radius, and diameter
  - \( x_0 = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{LS}{N} \)
  - \( R = \sqrt{\frac{\sum_{i=1}^{N} (x_i-x_0)^2}{N}} = \sqrt{\frac{N \cdot SS - 2LS^2 + LS^2}{N^2}} = \sqrt{\frac{N \cdot SS - LS^2}{N^2}} \)
  - \( D = \sqrt{\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (x_i-x_j)^2}{N(N-1)}} = \sqrt{\frac{2N \cdot SS - 2LS^2}{N(N-1)}} \)

- Additivity theorem
  - \( CF_1 + CF_2 = (N_1 + N_2, LS_1 + LS_2, SS_1 + SS_2) \)
BIRCH (1996)

- **Scales linearly**: finds a good clustering with a single scan and improves the quality with a few additional scans
  - $O(n)$: where $n$ is the number of the objects to be clustered

- **Weaknesses**
  - handles only numeric data, and sensitive to the order of the data record.
  - node size is limited by the memory size
  - not good quality when clusters are not in spherical