Hierarchical clustering

Ting Huang

Adapted from the slides of
David M. Blei from Princeton University

http://www.cs.princeton.edu/courses/archive/spr08/cos424/slides/clustering-2.pdf

Hierarchical clustering

- Hierarchical clustering is a widely used data analysis tool.

Hierarchical clustering

- Hierarchical clustering is a widely used data analysis tool.
- The idea is to build a binary tree of the data that successively merges similar groups of points

Hierarchical clustering

- Hierarchical clustering is a widely used data analysis tool.
- The idea is to build a binary tree of the data that successively merges similar groups of points
- Visualizing this tree provides a useful summary of the data

Agglomerative clustering

- We will talk about *agglomerative clustering*.

Agglomerative clustering

- We will talk about *agglomerative clustering*.
- Algorithm:

Agglomerative clustering

- We will talk about *agglomerative clustering*.
- Algorithm:
  1. Place each data point into its own singleton group

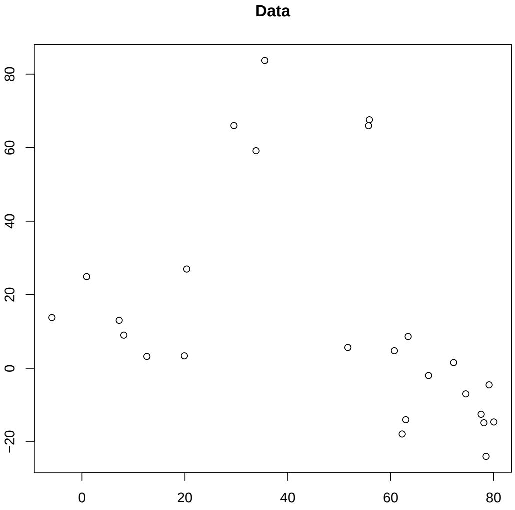
Agglomerative clustering

- We will talk about *agglomerative clustering*.
- Algorithm:
  1. Place each data point into its own singleton group
  2. Repeat: iteratively merge the two closest groups

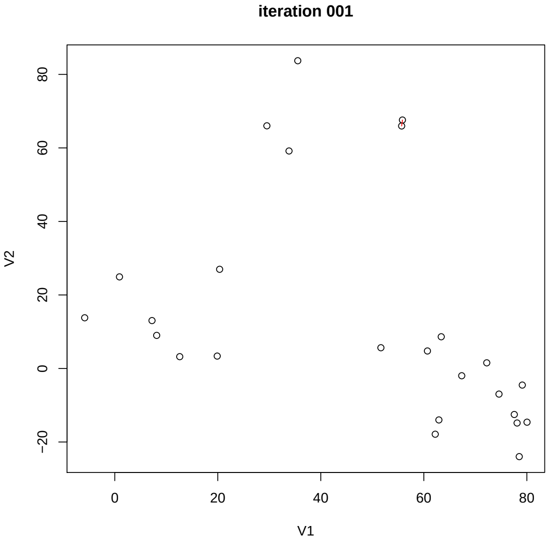
Agglomerative clustering

- We will talk about *agglomerative clustering*.
- Algorithm:
  1. Place each data point into its own singleton group
  2. Repeat: iteratively merge the two closest groups
  3. Until: all the data are merged into a single cluster

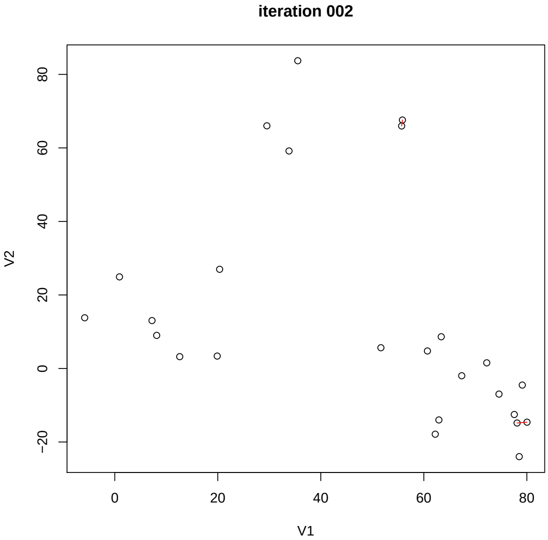
Example



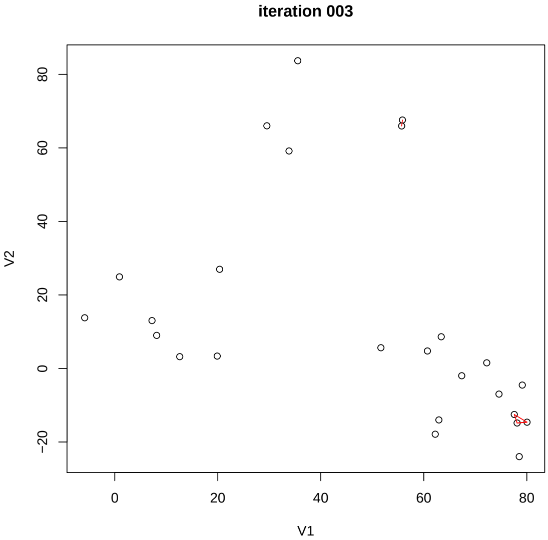
Example



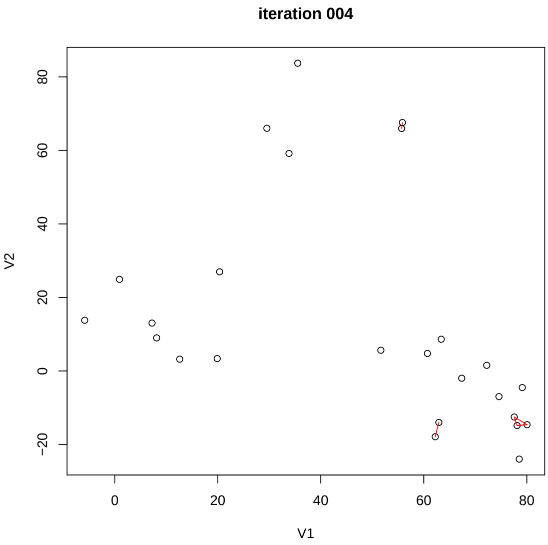
Example



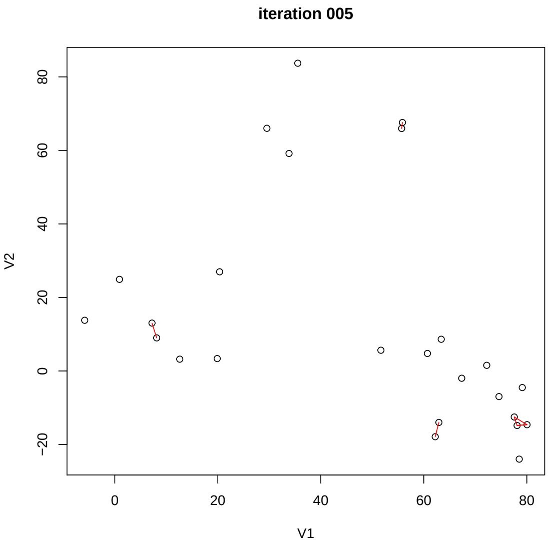
Example



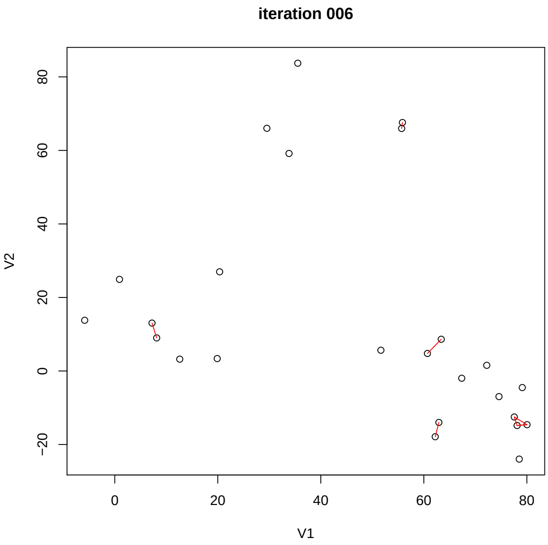
Example



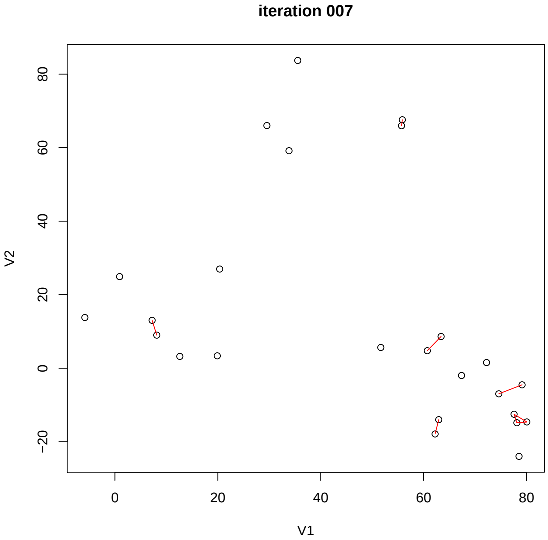
Example



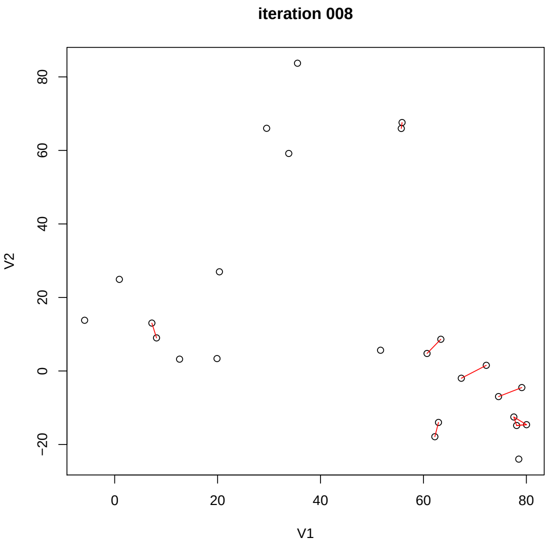
Example



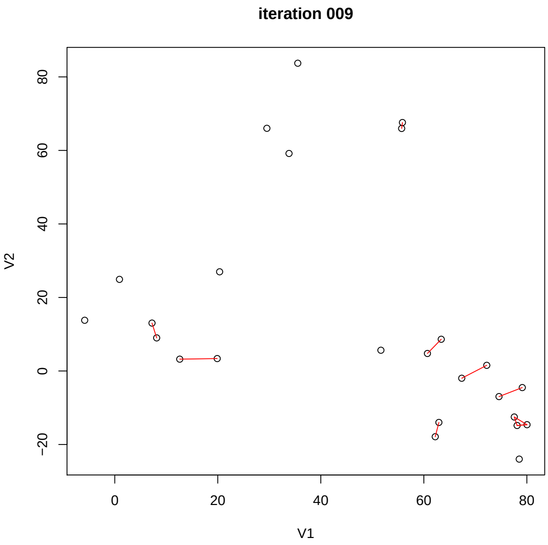
Example



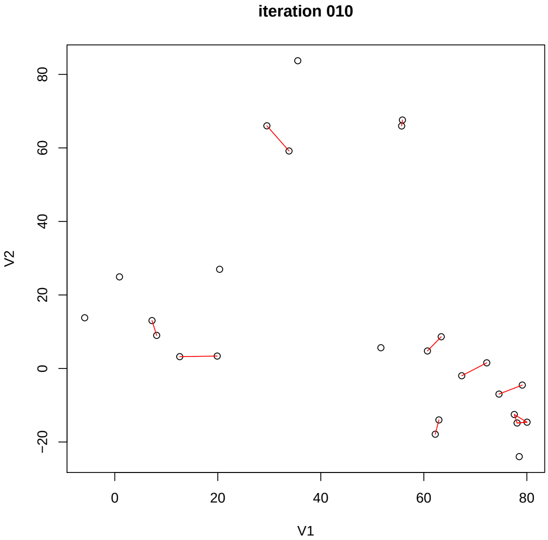
Example



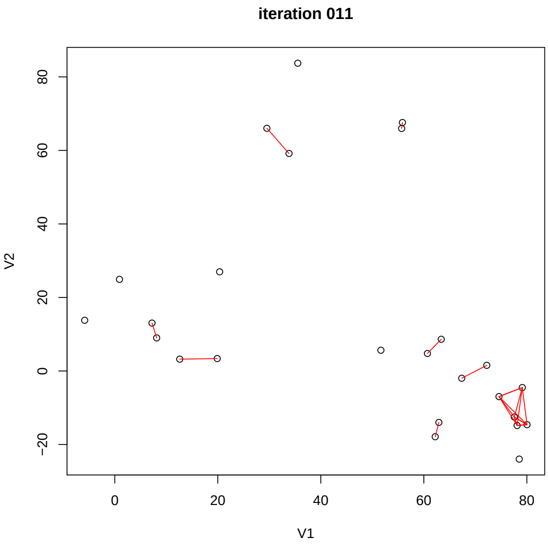
Example



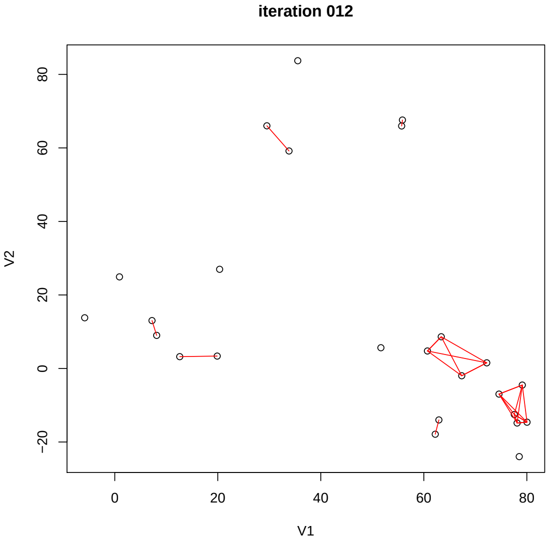
Example



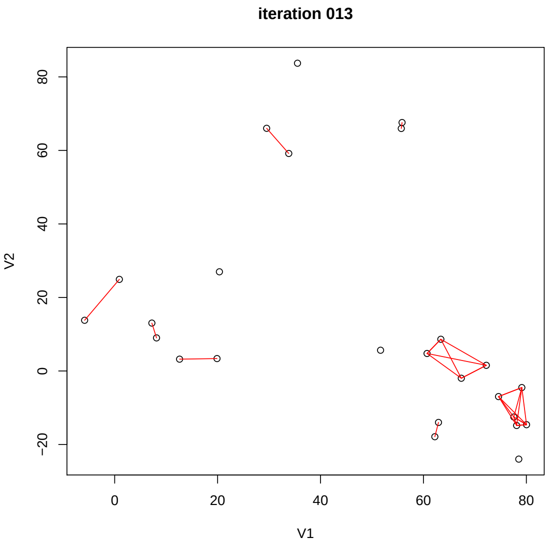
Example



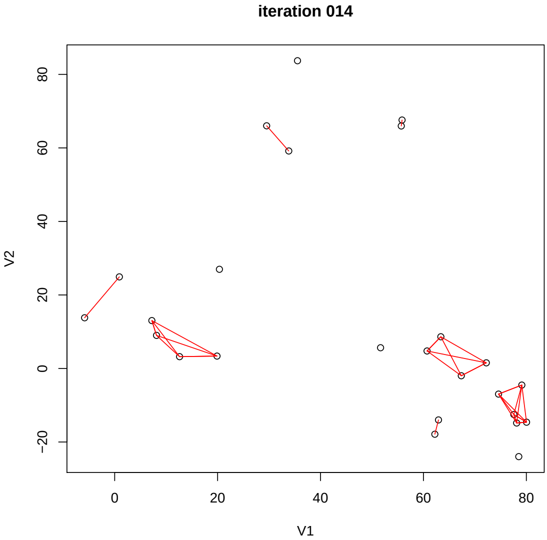
Example



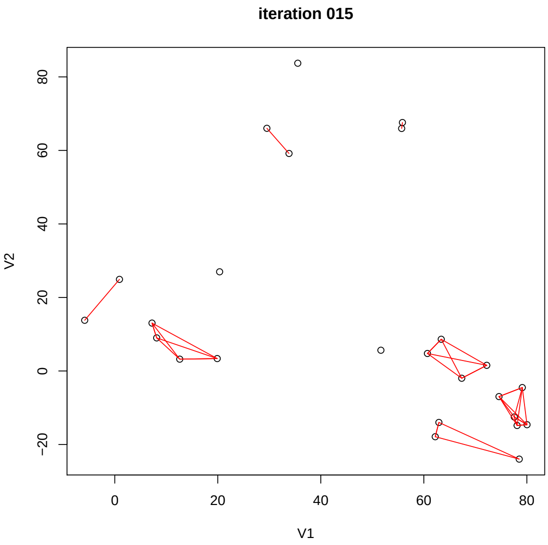
Example



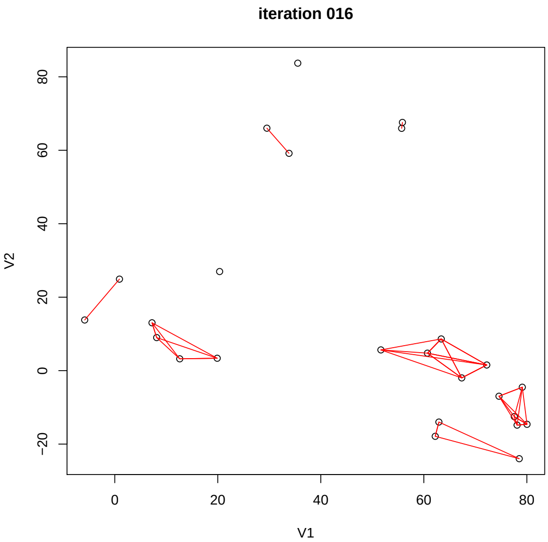
Example



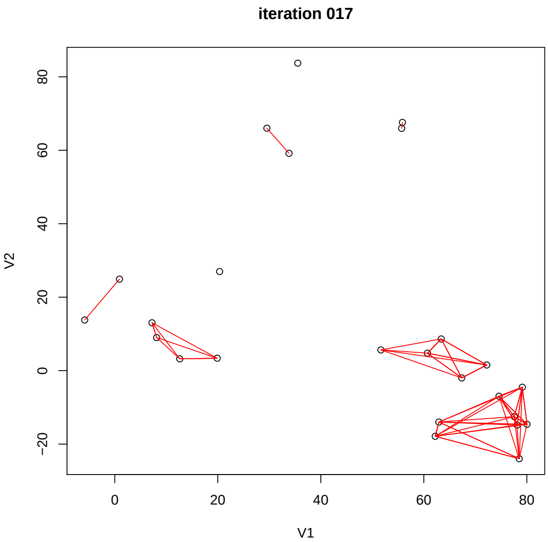
Example



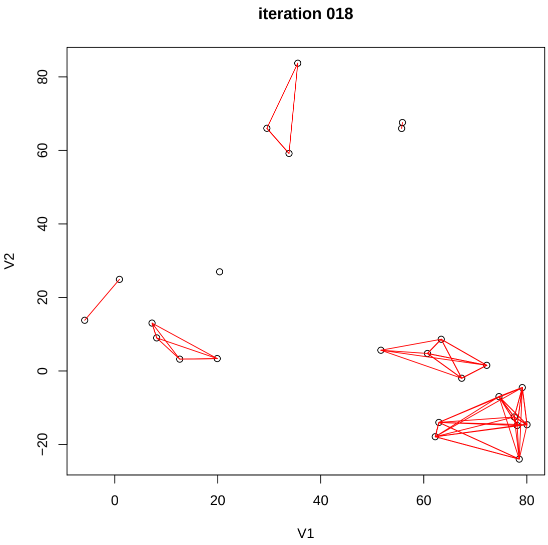
Example



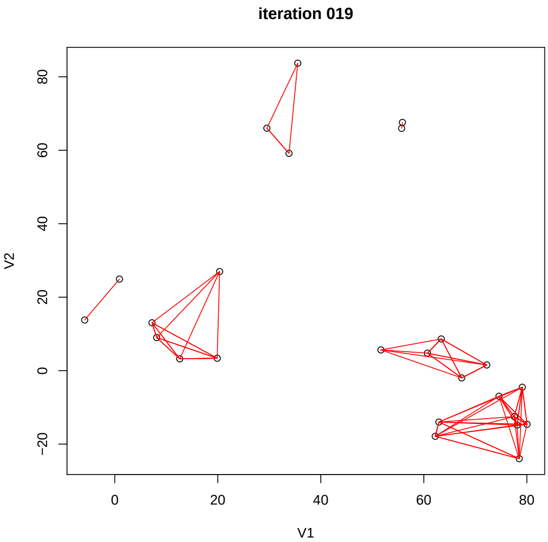
Example



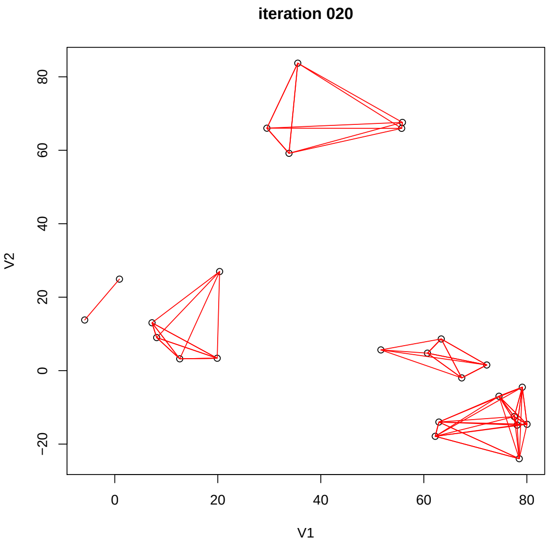
Example



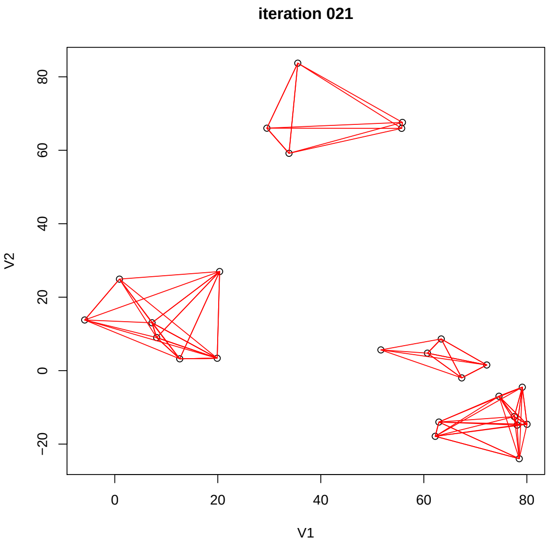
Example



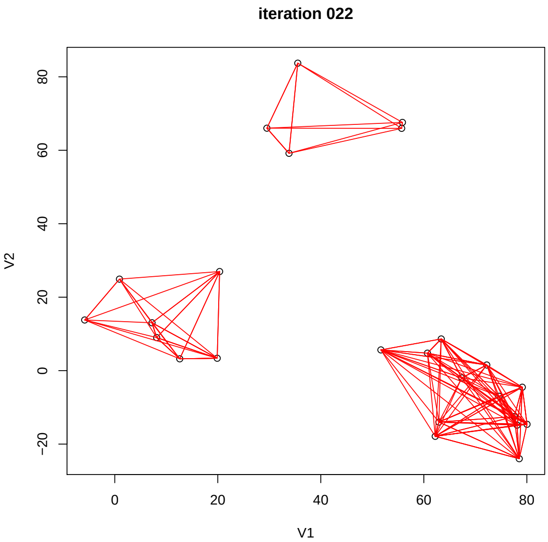
Example



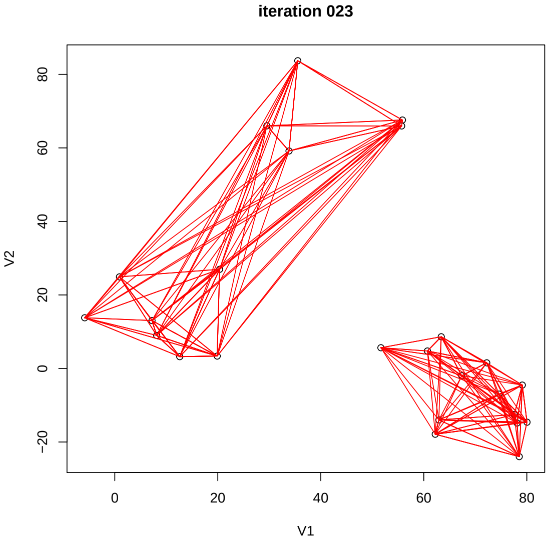
Example



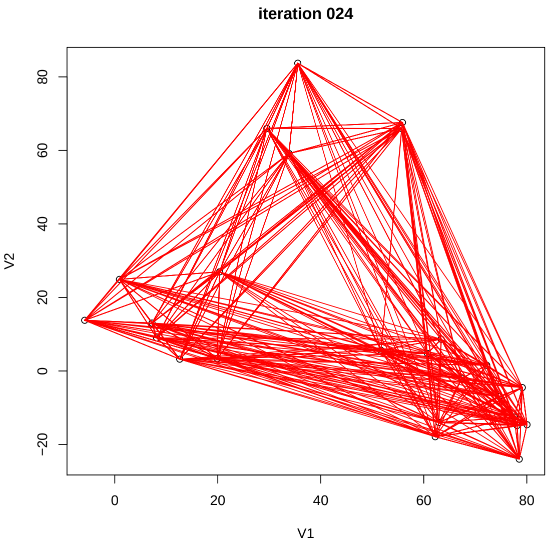
Example



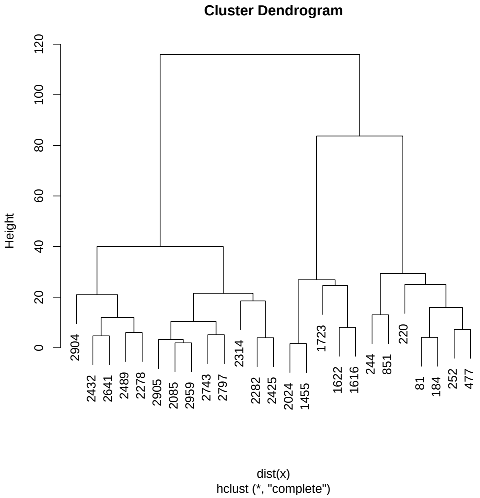
Example



Example



Dendrogram of example data



Groups that merge at high values relative to the merger values of their subgroups are candidates for natural clusters. (Tibshirani et al., 2001)

Group similarity

- Given a distance measure between points, the user has many choices for how to define intergroup similarity.

Group similarity

- Given a distance measure between points, the user has many choices for how to define intergroup similarity.
- Three most popular choices

Group similarity

- Given a distance measure between points, the user has many choices for how to define intergroup similarity.
- Three most popular choices
  - *Single-linkage*: the similarity of the closest pair

$$d_{SL}(G, H) = \min_{i \in G, j \in H} d_{i,j}$$

Group similarity

- Given a distance measure between points, the user has many choices for how to define intergroup similarity.
- Three most popular choices
  - *Single-linkage*: the similarity of the closest pair

$$d_{SL}(G,H) = \min_{i \in G, j \in H} d_{i,j}$$

  - *Complete linkage*: the similarity of the furthest pair

$$d_{CL}(G,H) = \max_{i \in G, j \in H} d_{i,j}$$

Group similarity

- Given a distance measure between points, the user has many choices for how to define intergroup similarity.
- Three most popular choices
  - *Single-linkage*: the similarity of the closest pair

$$d_{SL}(G, H) = \min_{i \in G, j \in H} d_{i,j}$$

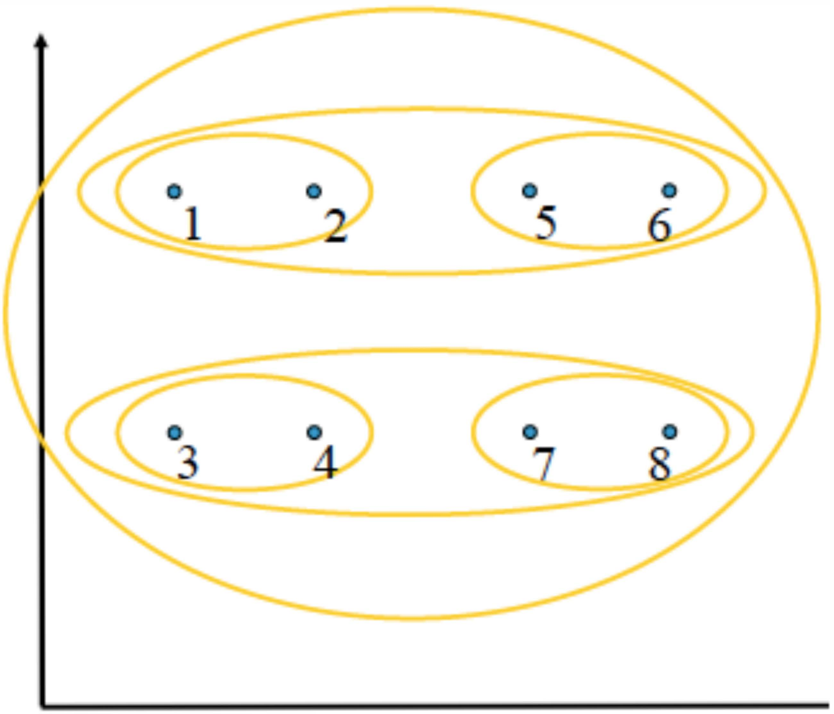
  - *Complete linkage*: the similarity of the furthest pair

$$d_{CL}(G, H) = \max_{i \in G, j \in H} d_{i,j}$$

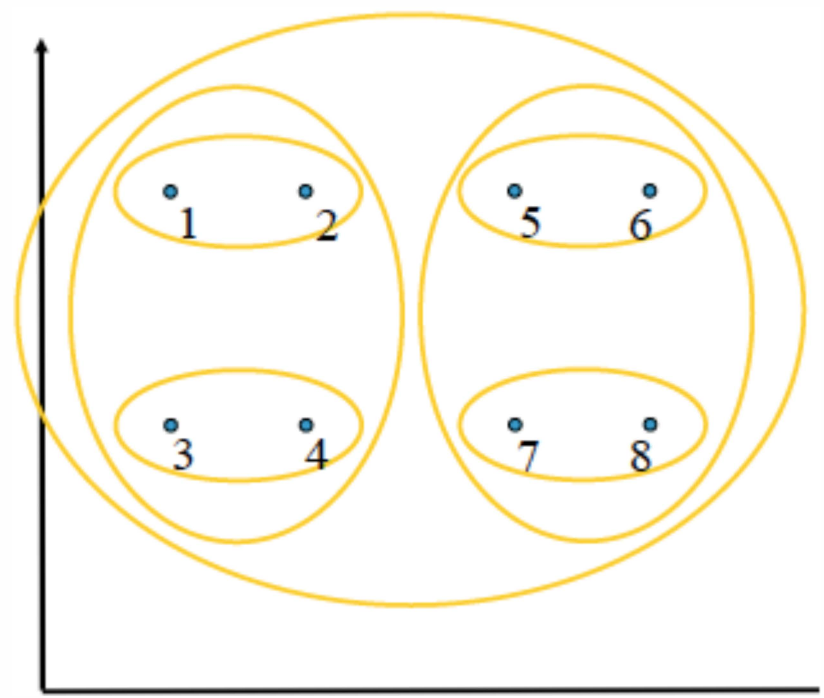
  - *Group average*: the average similarity between groups

$$d_{GA} = \frac{1}{N_G N_H} \sum_{i \in G} \sum_{j \in H} d_{i,j}$$

Closest pair
(single-link clustering)

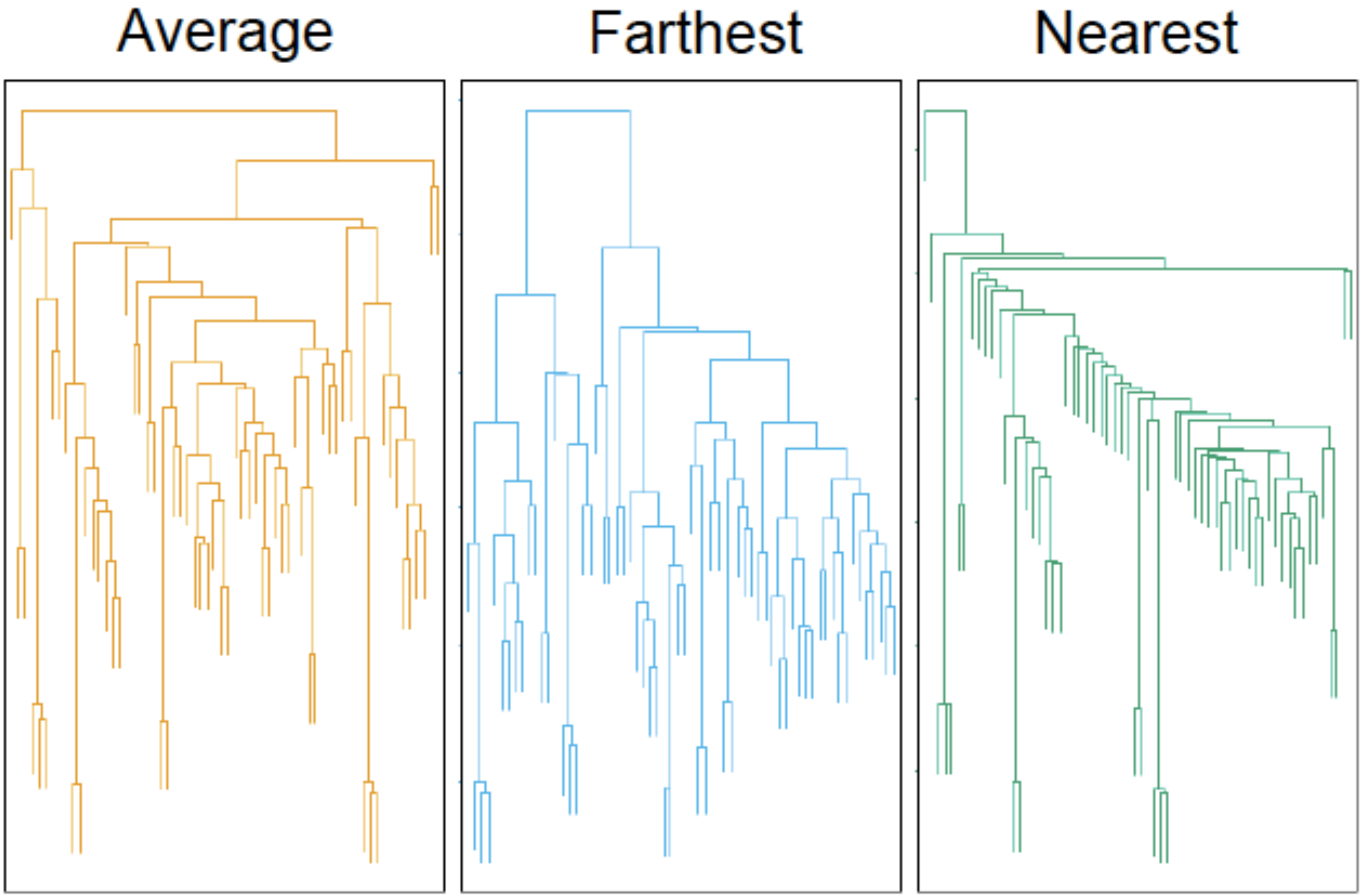


Farthest pair
(complete-link clustering)



[Pictures from Thorsten Joachims]

Slide from Lu Wang's machine learning course



Slide from Lu Wang's machine learning course
Mouse tumor data from [Hastie *et al.*]

Properties of intergroup similarity

- Single linkage can produce "chaining," where a sequence of close observations in different groups cause early merges of those groups

Properties of intergroup similarity

- Single linkage can produce "chaining," where a sequence of close observations in different groups cause early merges of those groups
- Complete linkage has the opposite problem. It might not merge close groups because of outlier members that are far apart.

Properties of intergroup similarity

- Single linkage can produce "chaining," where a sequence of close observations in different groups cause early merges of those groups
- Complete linkage has the opposite problem. It might not merge close groups because of outlier members that are far apart.
- Group average represents a natural compromise, but depends on the scale of the similarities. Applying a monotone transformation to the similarities can change the results.