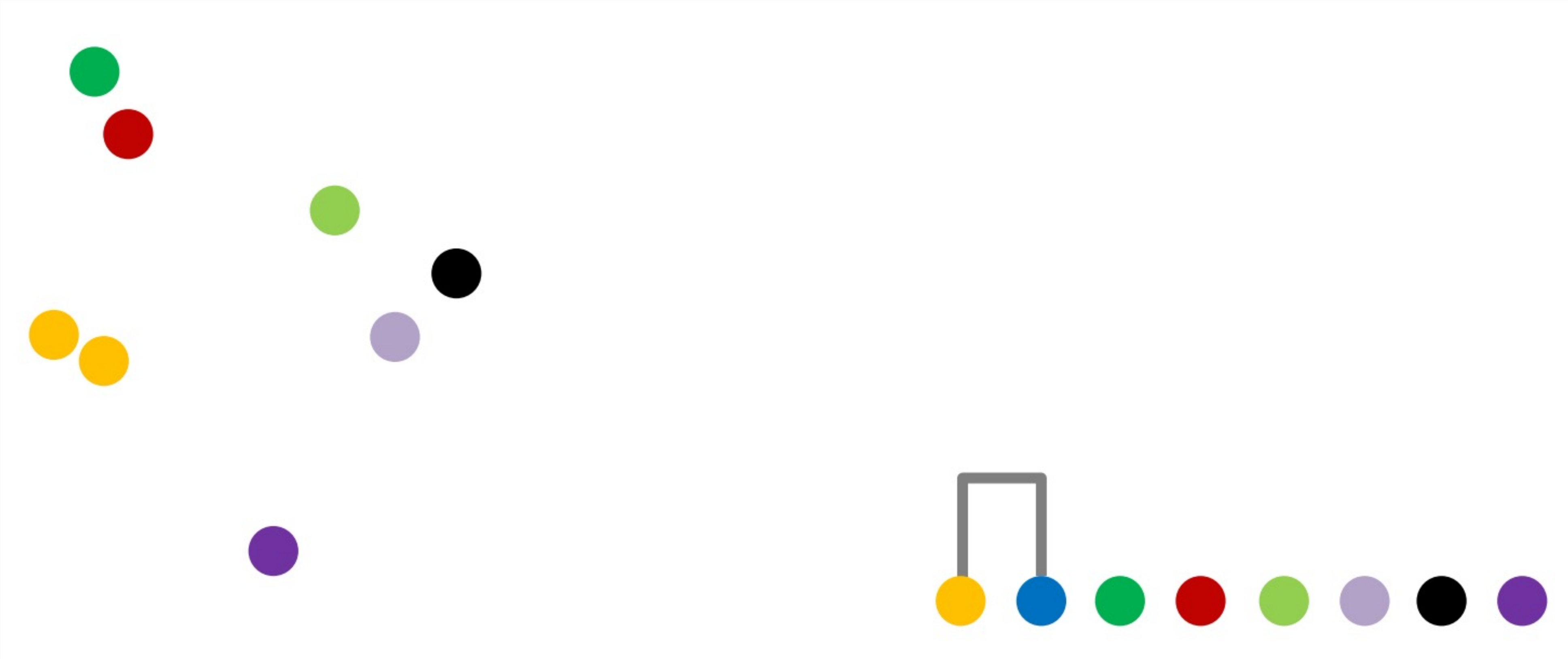
Parameters parameters

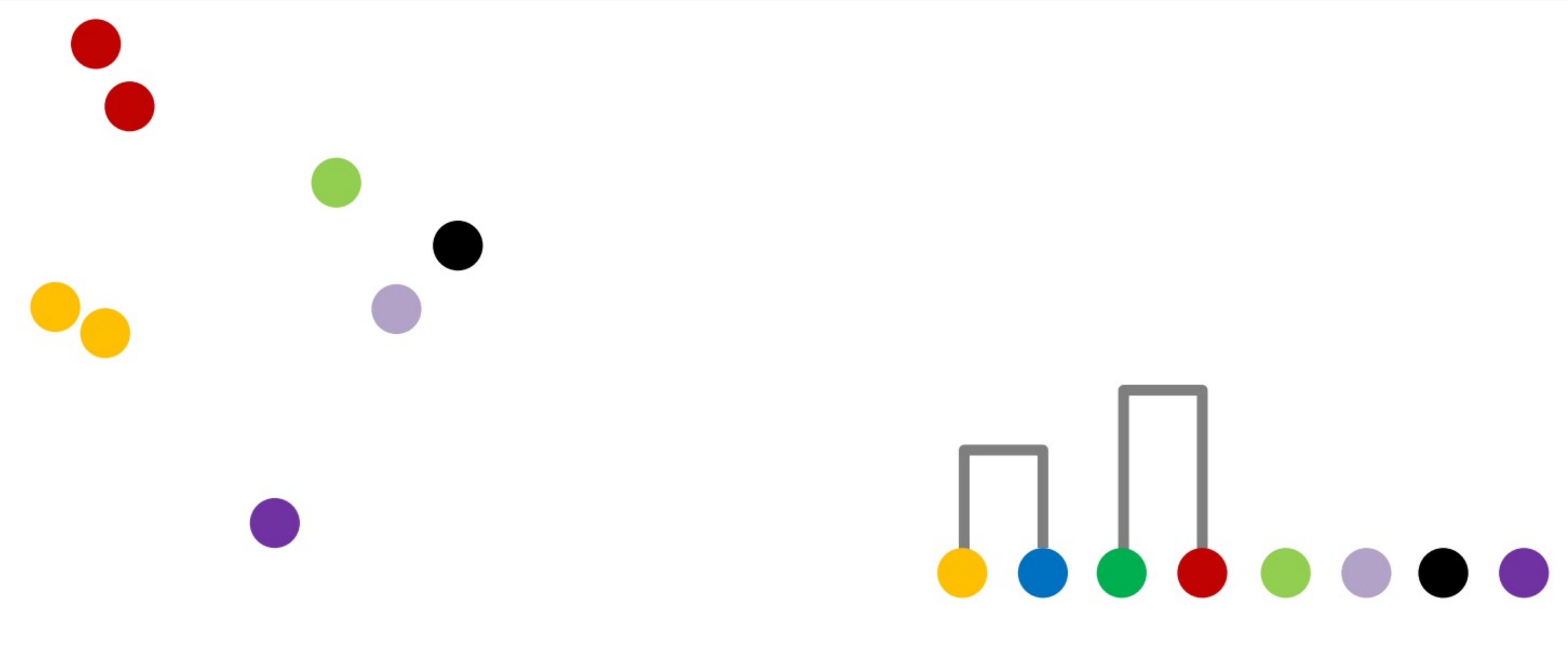
- For K means we need K and result depends on initialization
- For mean shift we need the window size and a lot of computation
- Hierarchical Clustering keeps a history of all possible cluster assignments

Hierarchical Clustering

Hierarchical Clustering

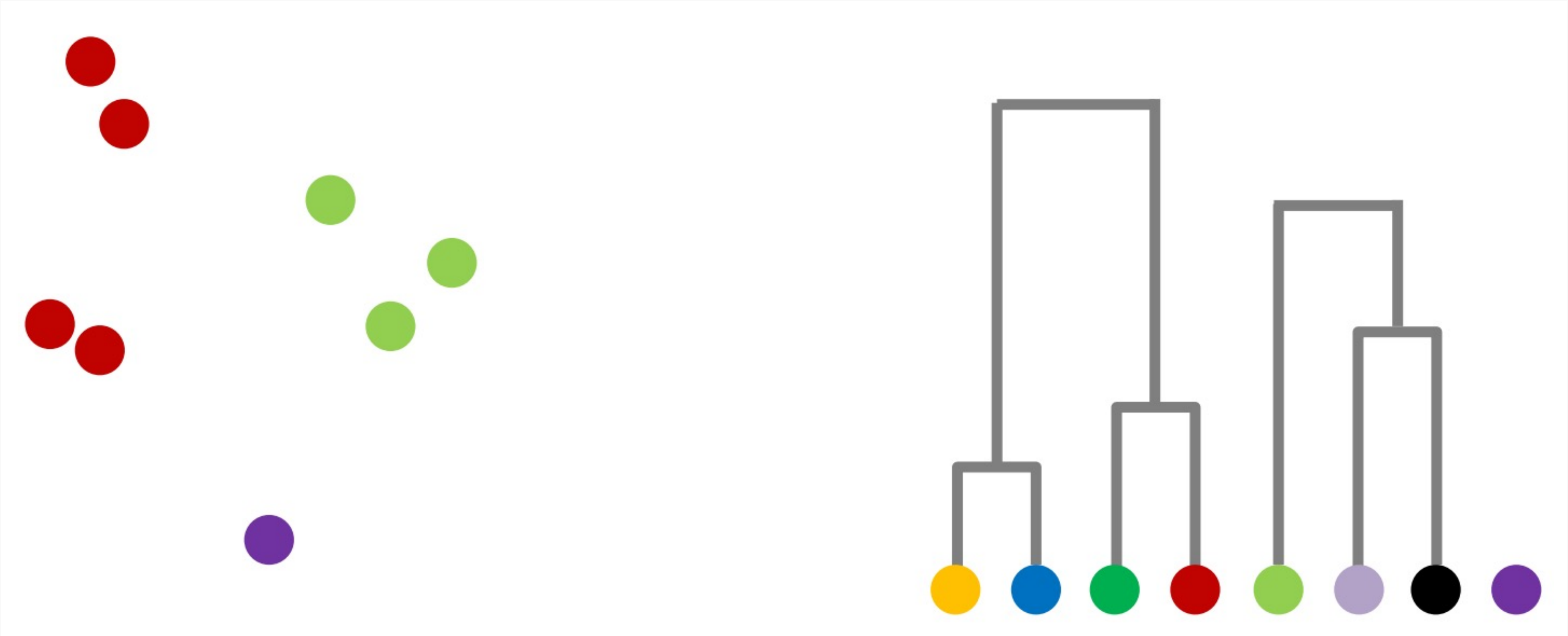


Hierarchical Clustering

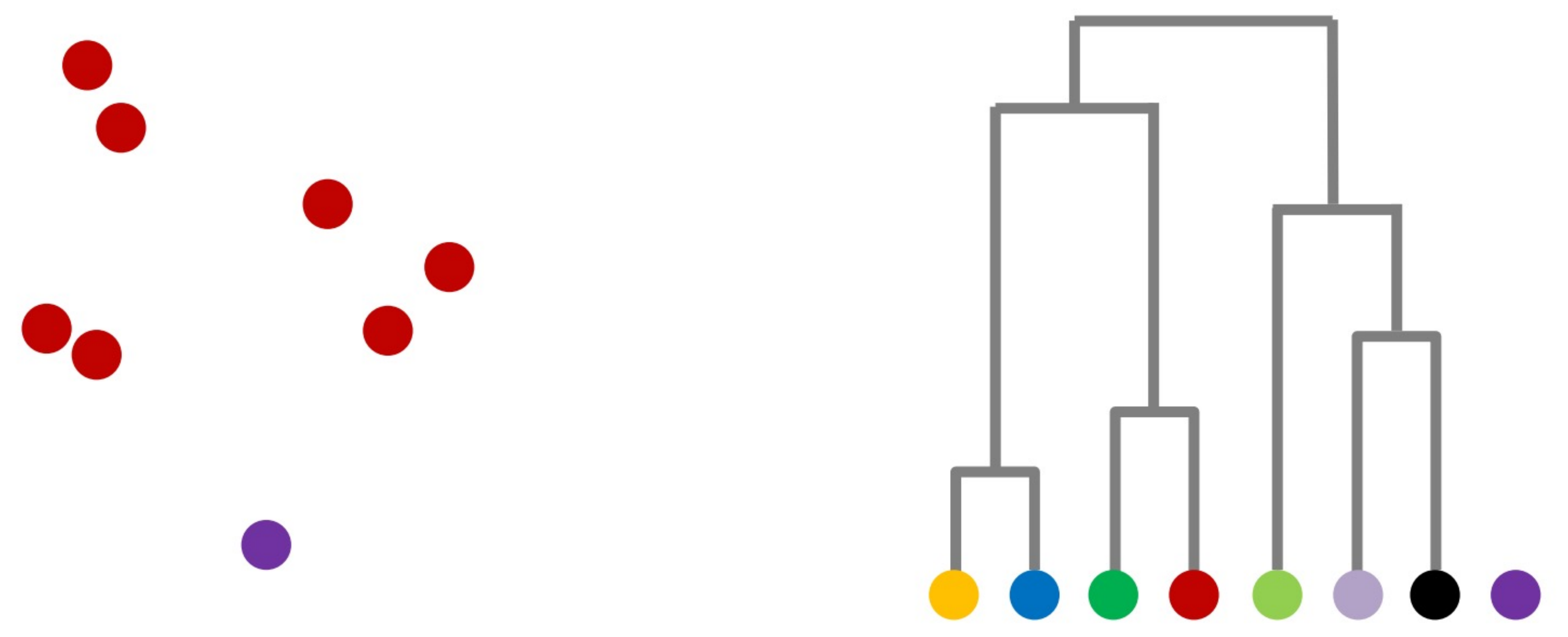


Hierarchical Clustering

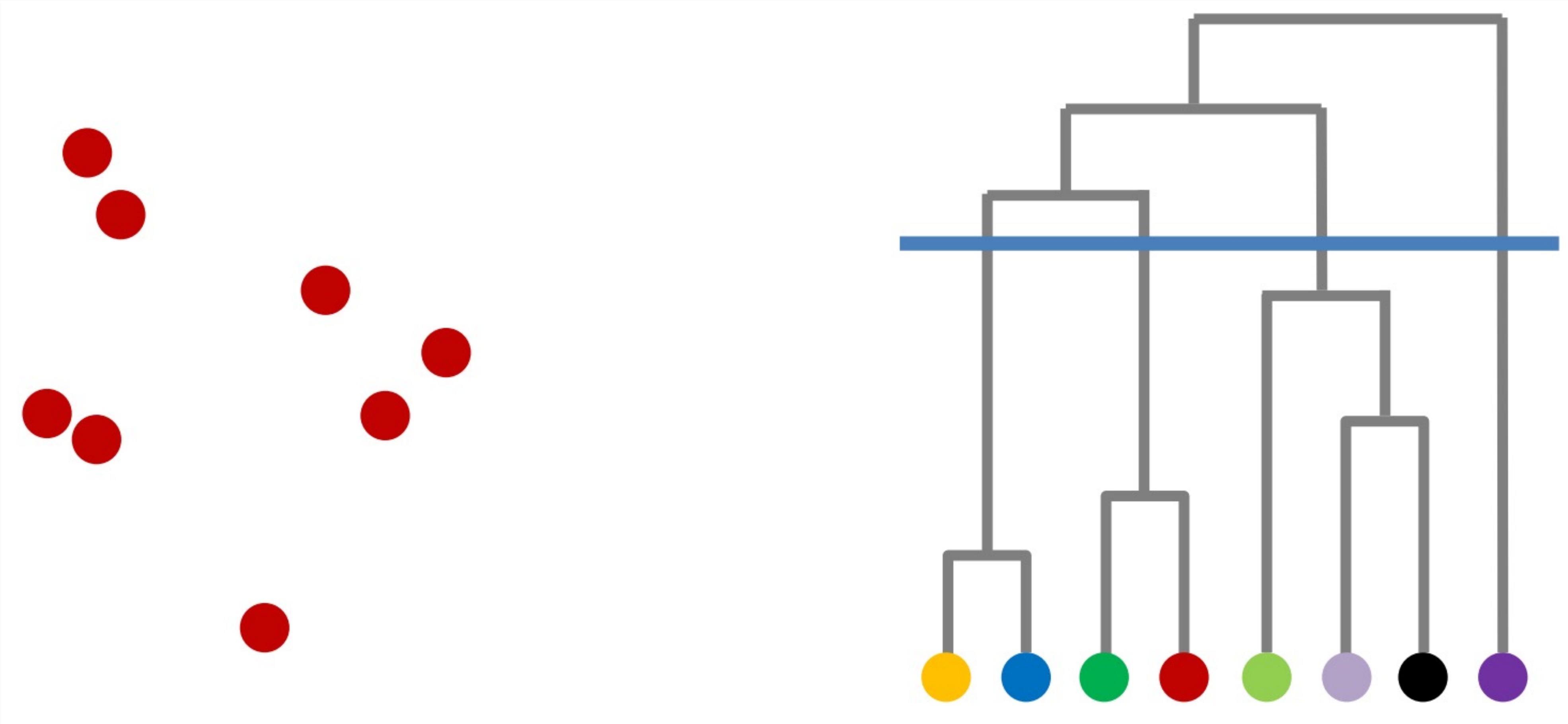
Hierarchical Clustering



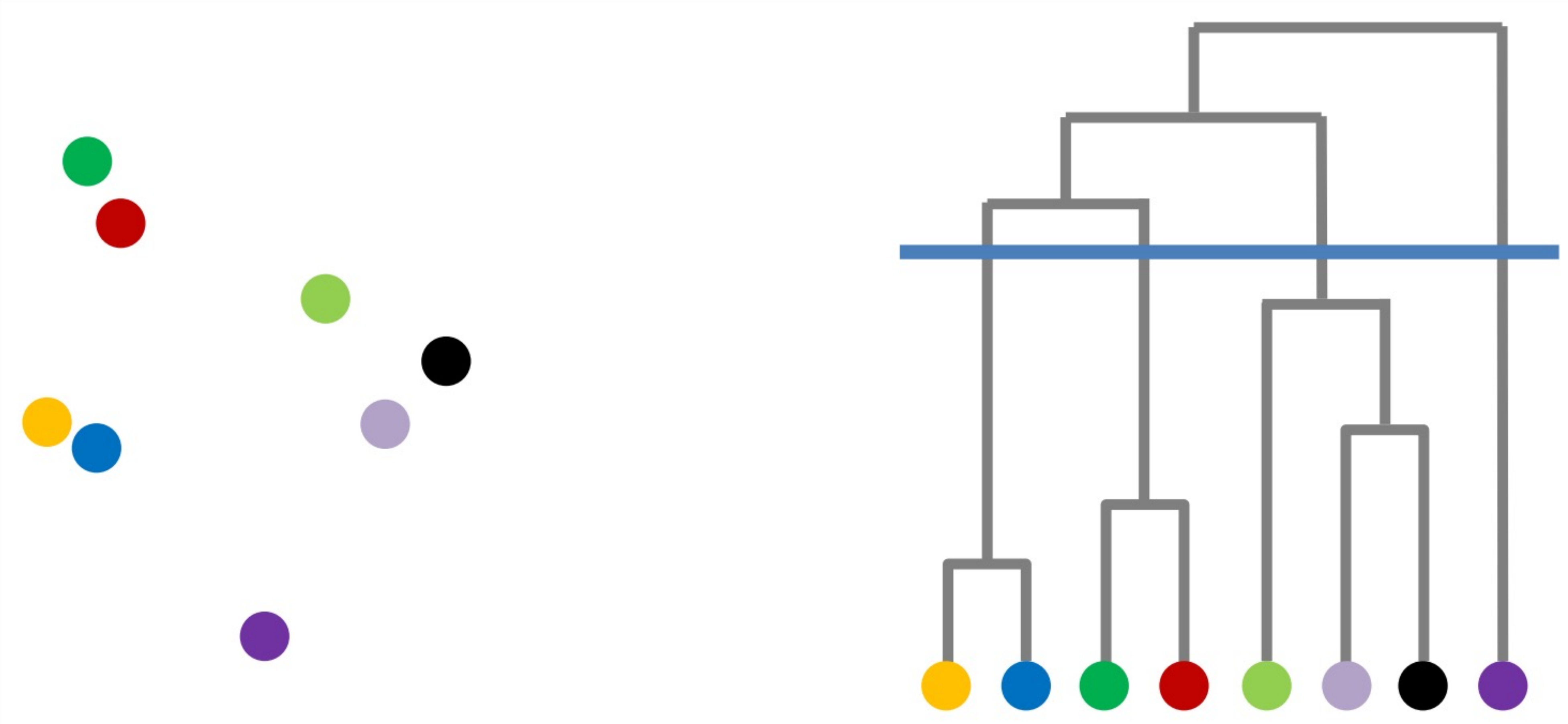
Hierarchical Clustering



Hierarchical Clustering



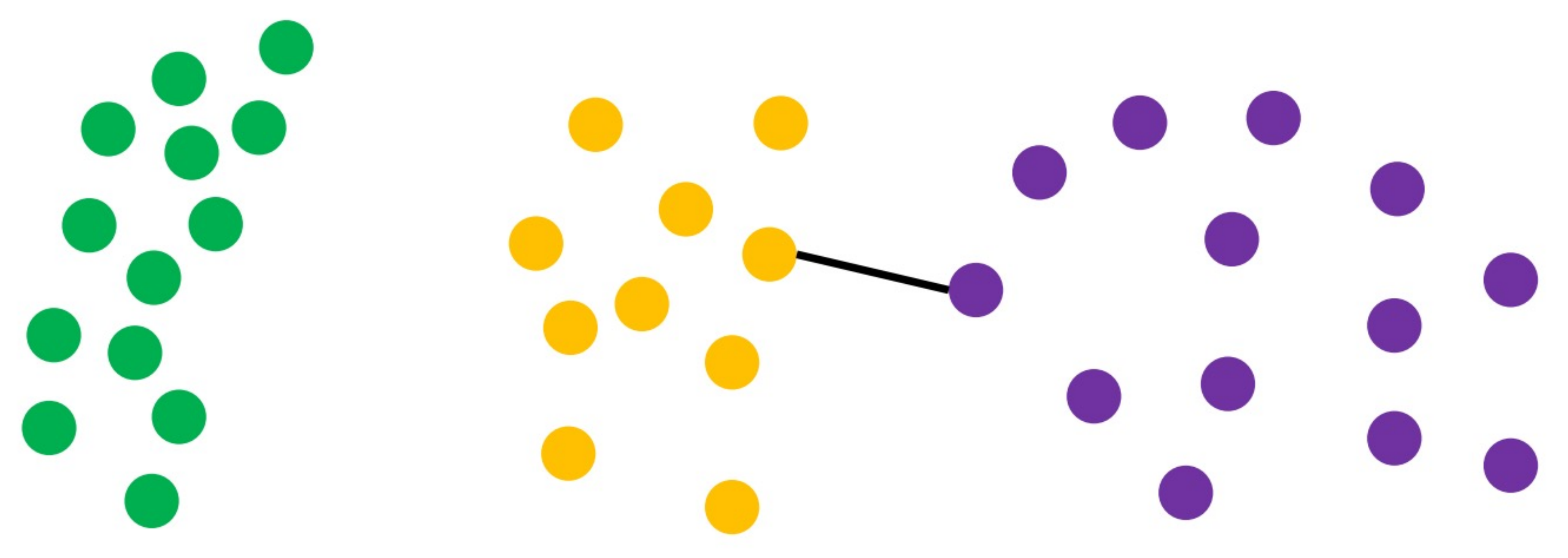
Hierarchical Clustering



Hierarchical Clustering

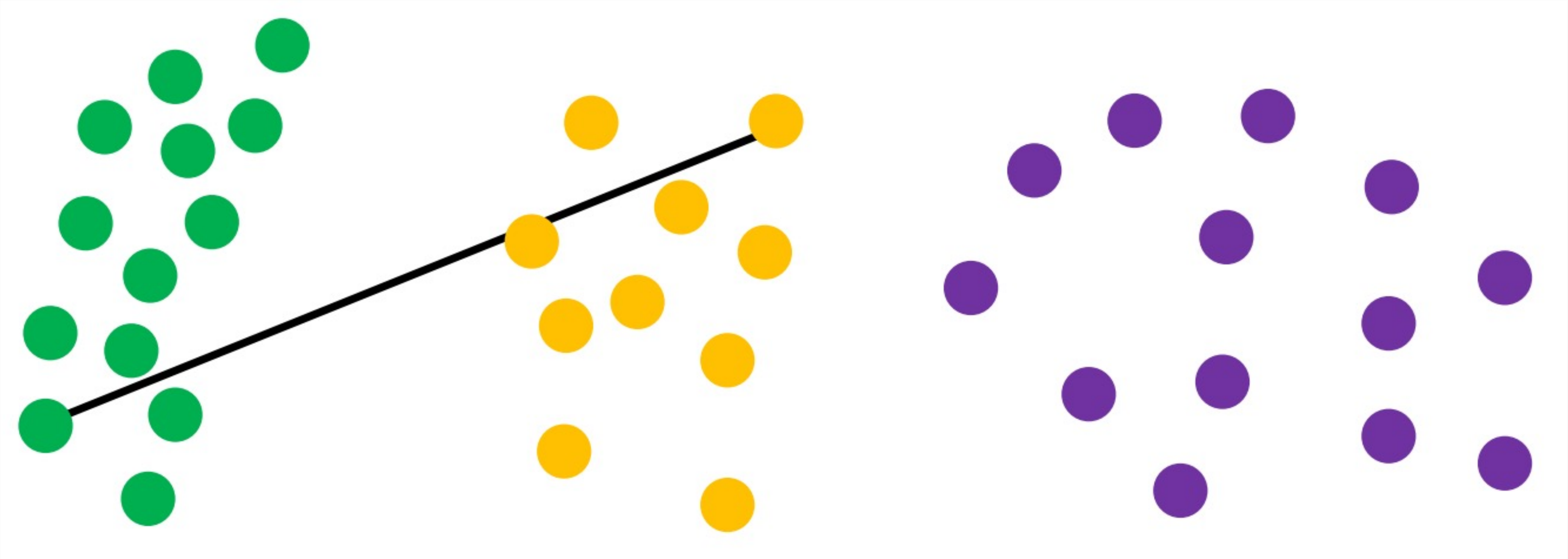
- Produces complete structure
- No predefined number of clusters
- Similarity between clusters:
  - single-linkage: $\min\{d(x,y) : x \in \mathcal{A}, y \in \mathcal{B}\}$
  - complete-linkage: $\max\{d(x,y) : x \in \mathcal{A}, y \in \mathcal{B}\}$
  - average linkage: $\frac{1}{|\mathcal{A}| \cdot |\mathcal{B}|} \sum_{x \in \mathcal{A}} \sum_{y \in \mathcal{B}} d(x,y)$

Single Linkage



$$\min\{d(x,y) : x \in \mathcal{A}, y \in \mathcal{B}\}$$

Complete Linkage

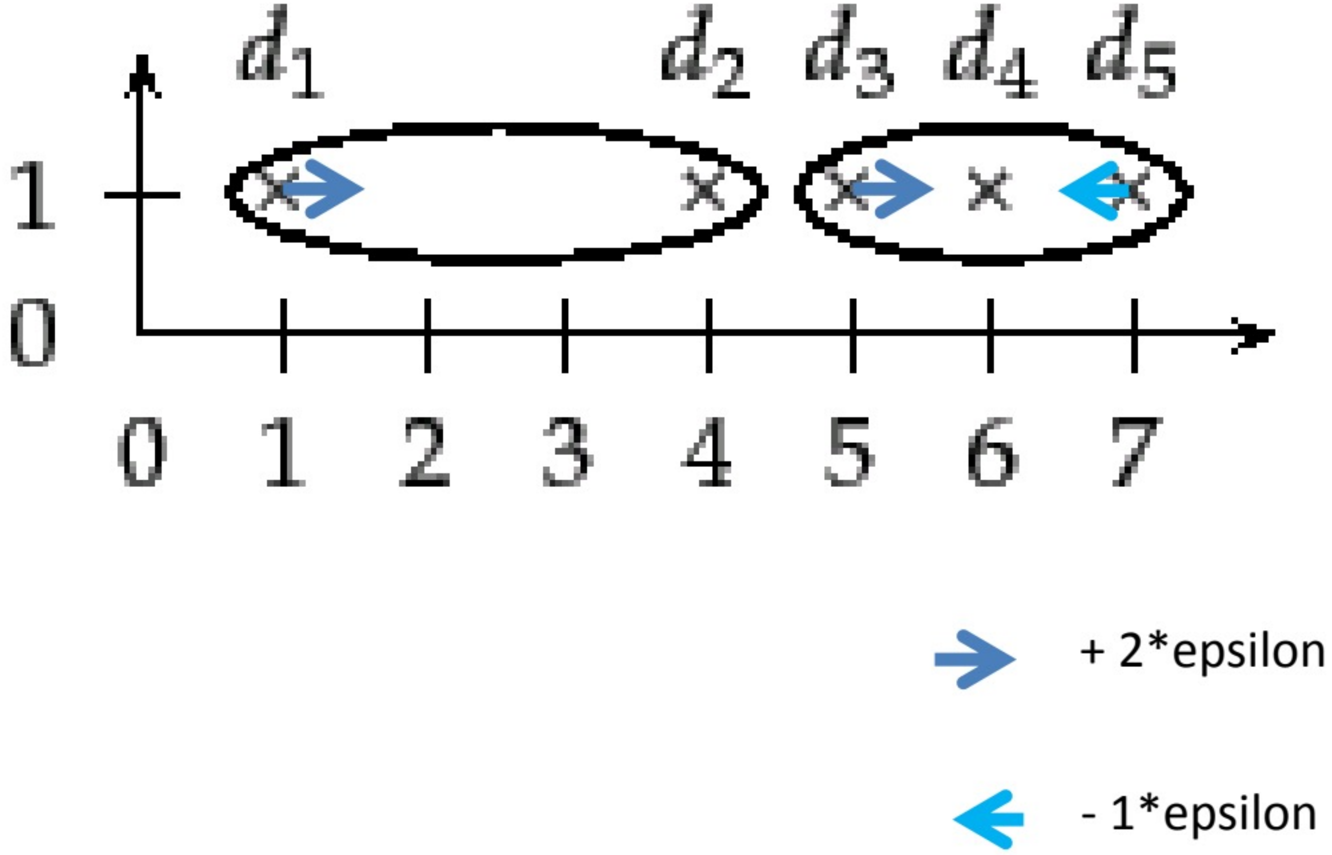


$$\max\{d(x,y) : x \in \mathcal{A}, y \in \mathcal{B}\}$$

Linkage Matters

- Single linkage: tendency to form long chains
- Complete linkage: Sensitive to outliers
- Average-link: Trying to compromise between the two

Outlier Sensitivity

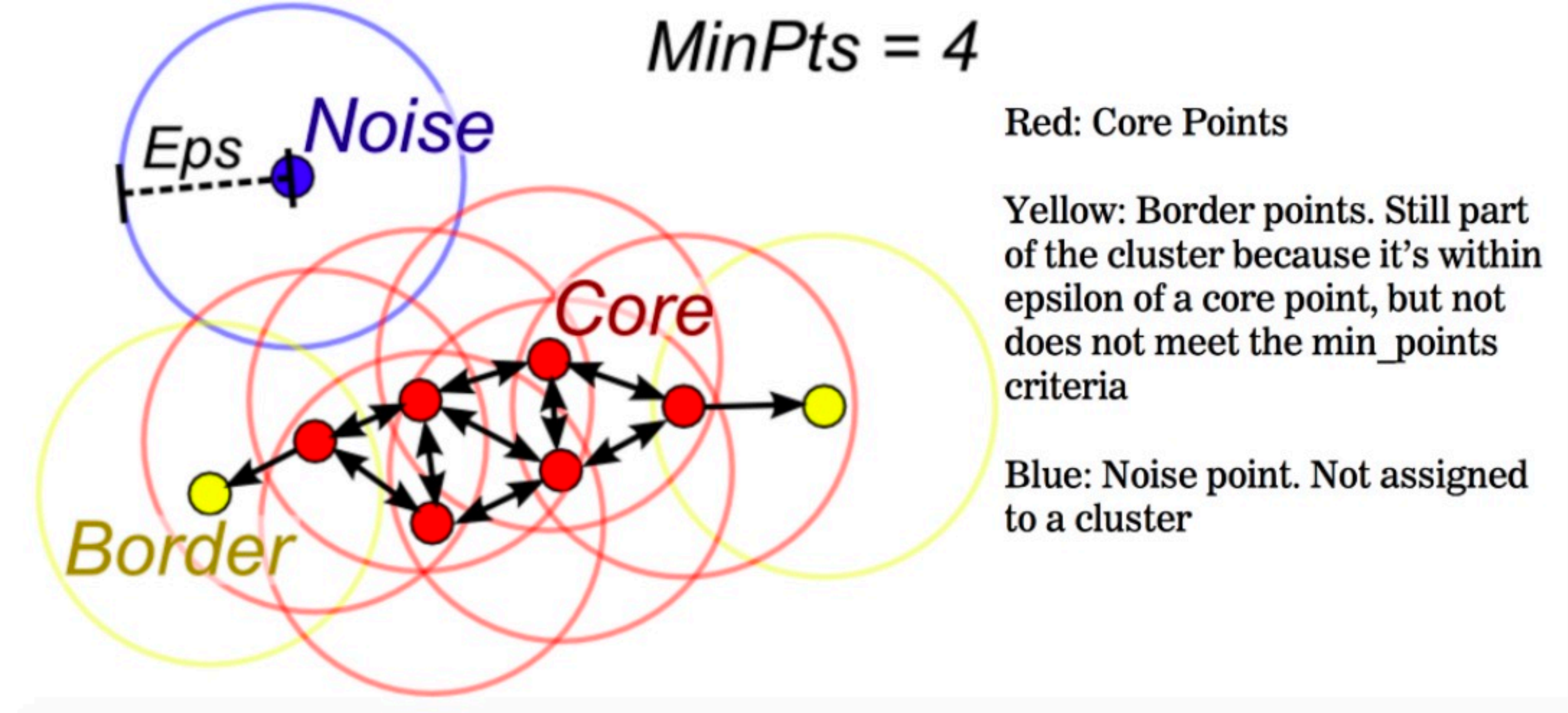


http://nlp.stanford.edu/IR-book/html/htmledition/img1569.png

DBSCAN

Density-based spatial clustering of applications with noise (DBSCAN)

- groups together points that are closely packed together
- marking as outliers points that lie alone in low-density regions

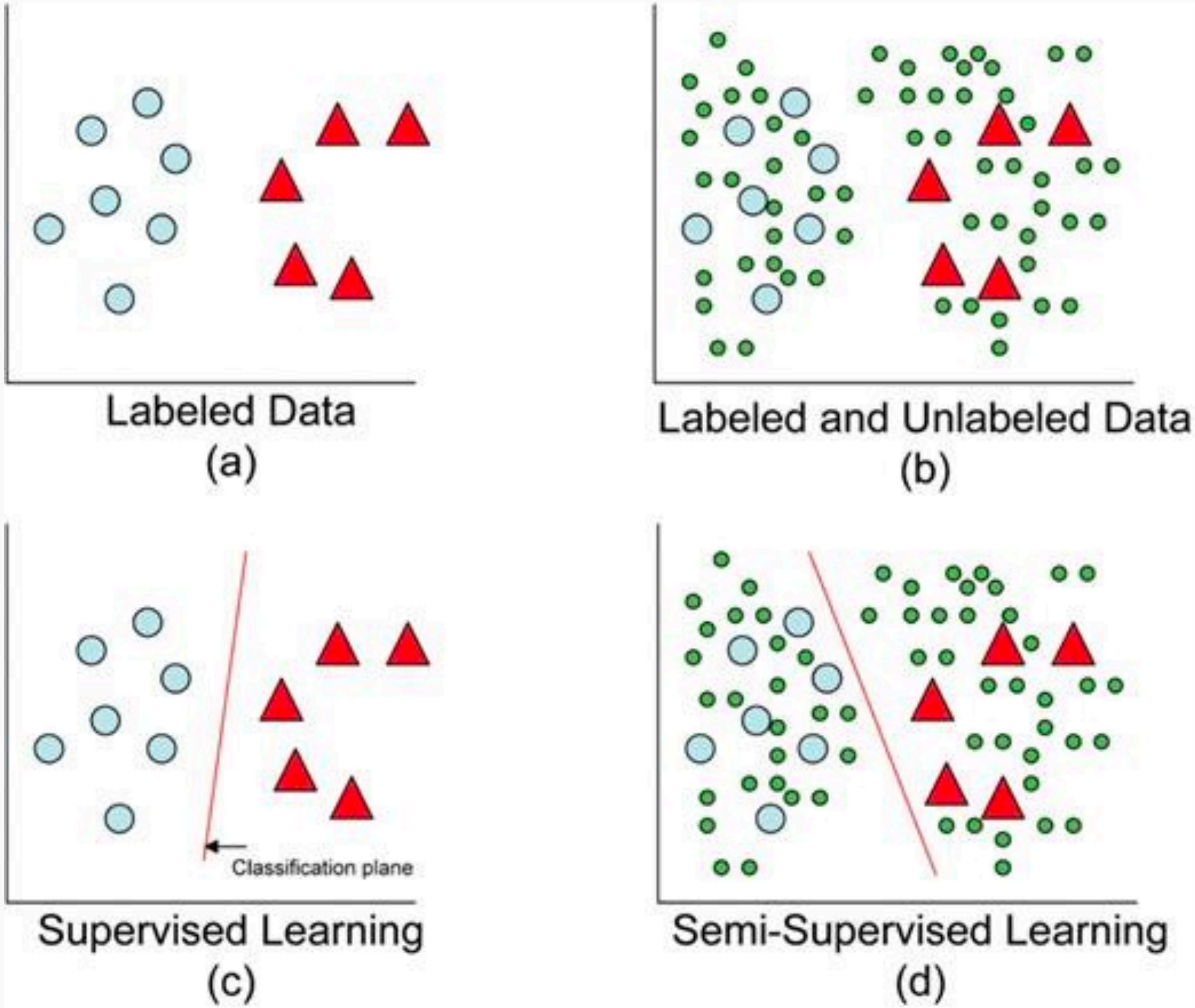


DBSCAN

- A point p is a *core point* if at least minPts points are within distance ε of it (including p).
- A point q is *directly reachable* from p if point q is within distance ε from core point p.
- A point q is *reachable* from p if there is a path $p_1,\ldots, p_n$ with $p_1 = p$ and $p_n = q$, where each p_{i+1} is directly reachable from p_i.
- All points not reachable from any other point are *outliers* or *noise points*.

Semi-supervised Learning (SSL)

Semi-supervised Learning



Some SSL Algorithms

- Self-Training
- Generative methods, mixture models
- Graph-based methods
- Co-Training
- Semi-supervised SVM
- Many others

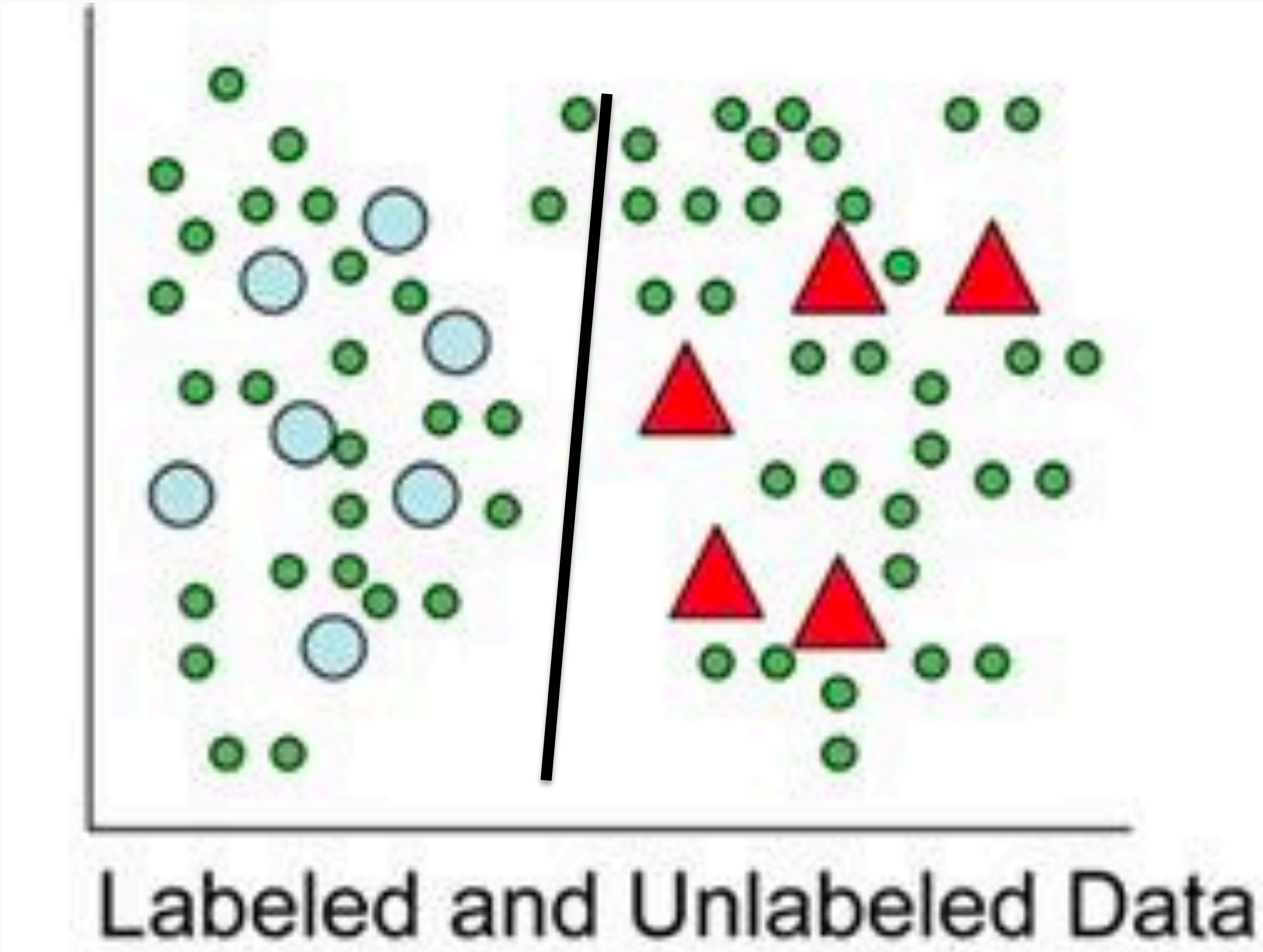
Self-training

Repeat:

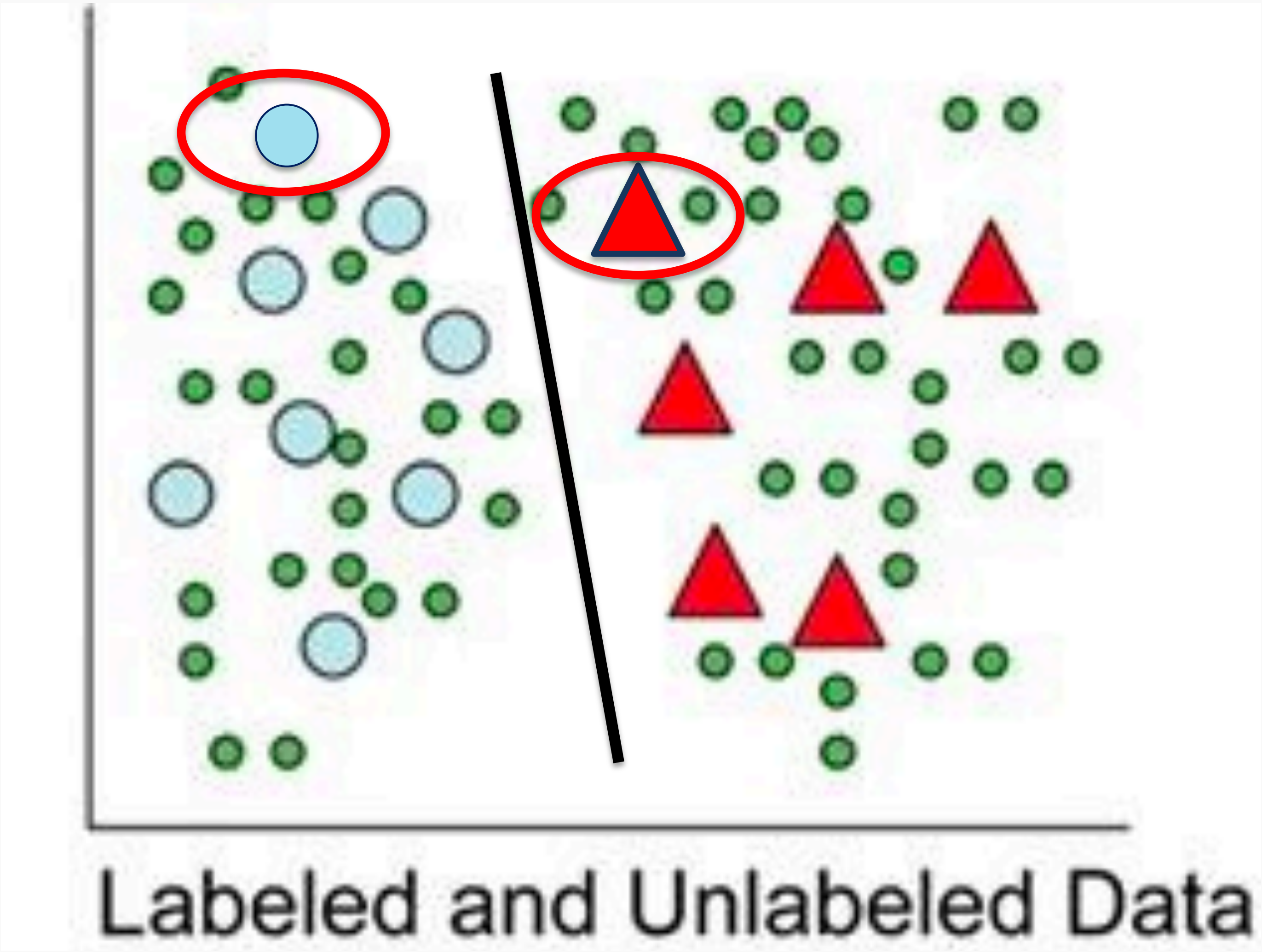
- Train the model f on the labeled data L (using supervised learning)
- Apply f to the unlabeled instances U
- Remove a subset S from U; add $\{(x, f(x)) \mid x \in S\}$ to L

- The choice of f is open
- Works well for many real world tasks
- But mistakes by f can reinforce itself

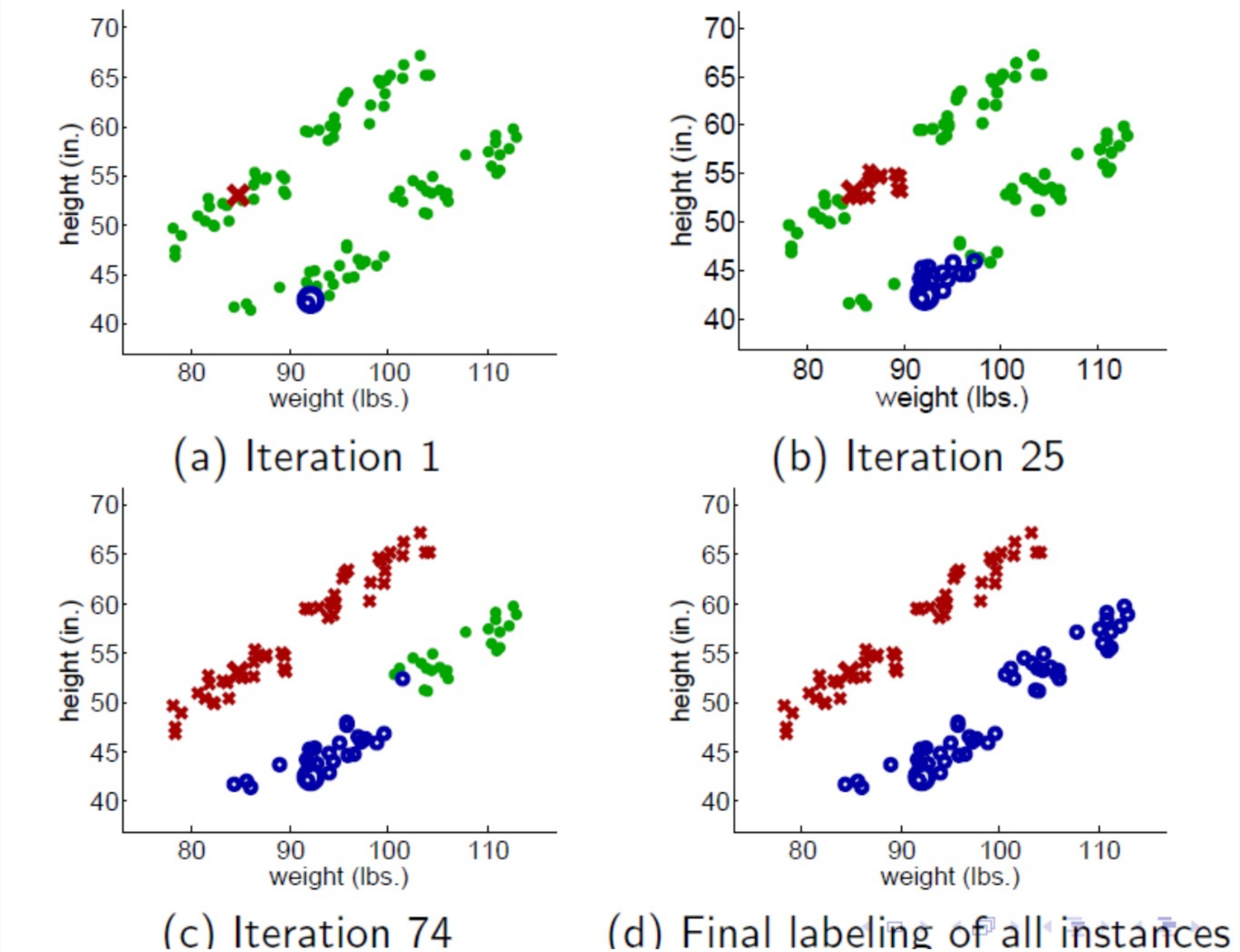
Self-training



Self-training



Propagating a 1-Nearest Neighbor: now it works

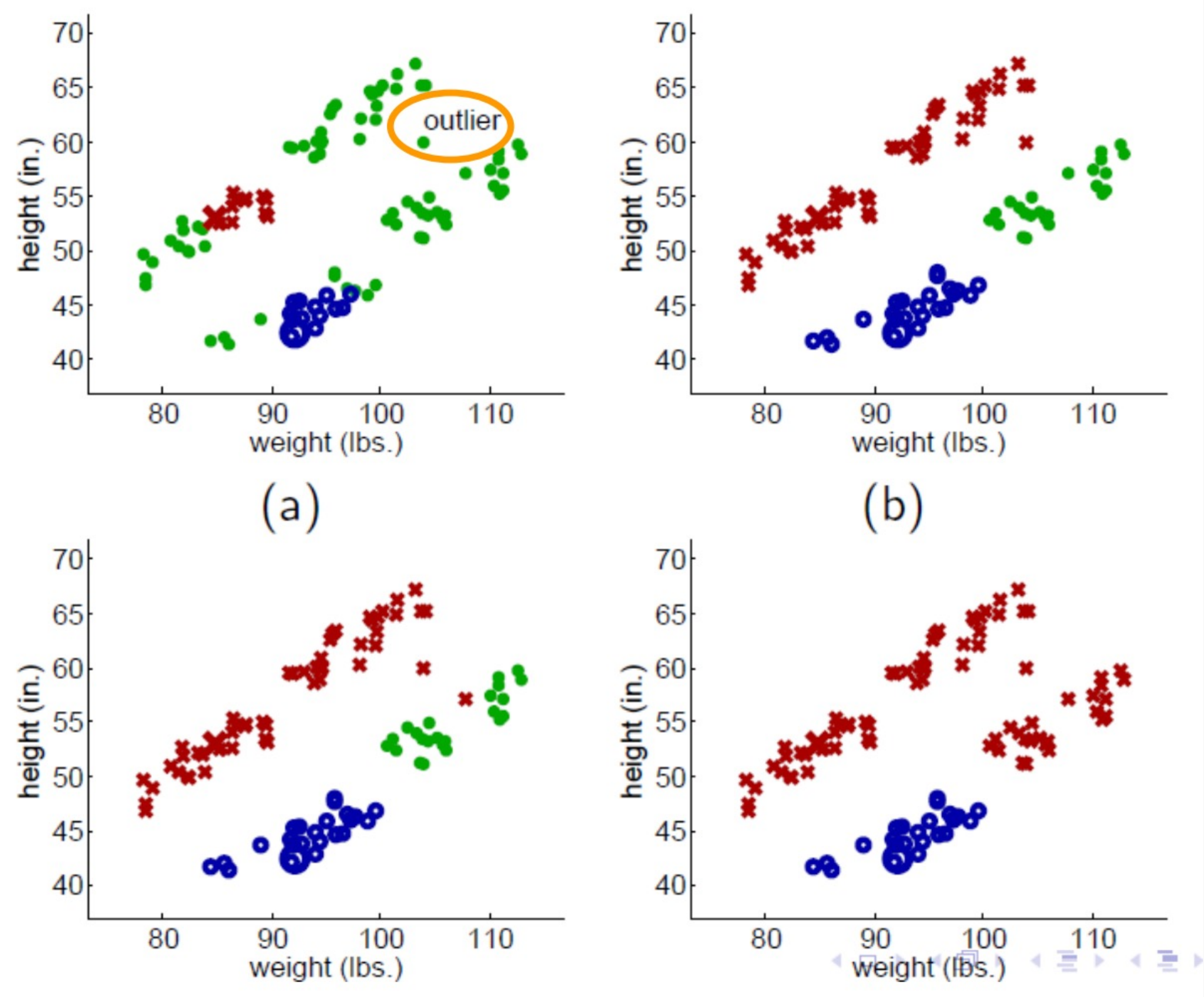


(a) Iteration 1 (b) Iteration 25

(c) Iteration 74 (d) Final labeling of all instances

Propagating a 1-Nearest Neighbor: now it does not work

But with a single outlier…



Summary

Regularization of NN

Optimization Challenges

Unsupervised Learning

- K-means
- Mean Shift
- Hierarchical Clustering
- DBSCAN

Semi-supervised Learning

- Self Training