Machine Learning for Large-Scale Data Analysis and Decision Making (MATH80629A)
Fall 2021

Week #10 - **Summary**

Announcement

- **Homework 1:** due **November 5, 2021**
- **Homework 2:** due **November 14, 2021**

Today

- **Fifth Quiz** on Gradescope!
- Summary of Unsupervised Learning
- Q&A
- Hands-on session

Quiz 4

Login to your Gradescope account

Recall

Experience (E)

- What data does f experience?
  - (Focus on algorithms that experience whole datasets)
  - Unsupervised. Examples alone $\{x_i\}_{i=0}^{n}$
  - Supervised. Examples come with labels $\{(x_i, y_i)\}_{i=0}^{n}$

Recall

Unsupervised

$$\{x_i\}_{i=0}^{n}$$

- Experience examples alone
- Learn “useful properties of the structure of the data”
  - E.g., clustering, density modeling (p(x)), PCA, FA.

Different tasks

- **Finding patterns**
  - **Clustering** $f : X \rightarrow \{1, 2, \ldots, K\}$ (K clusters)
  - **Dimensionality reduction** $f : X^p \rightarrow X^k, k << p$
  - **Density modelling** $f : X \rightarrow [0, 1]$
  - **…**

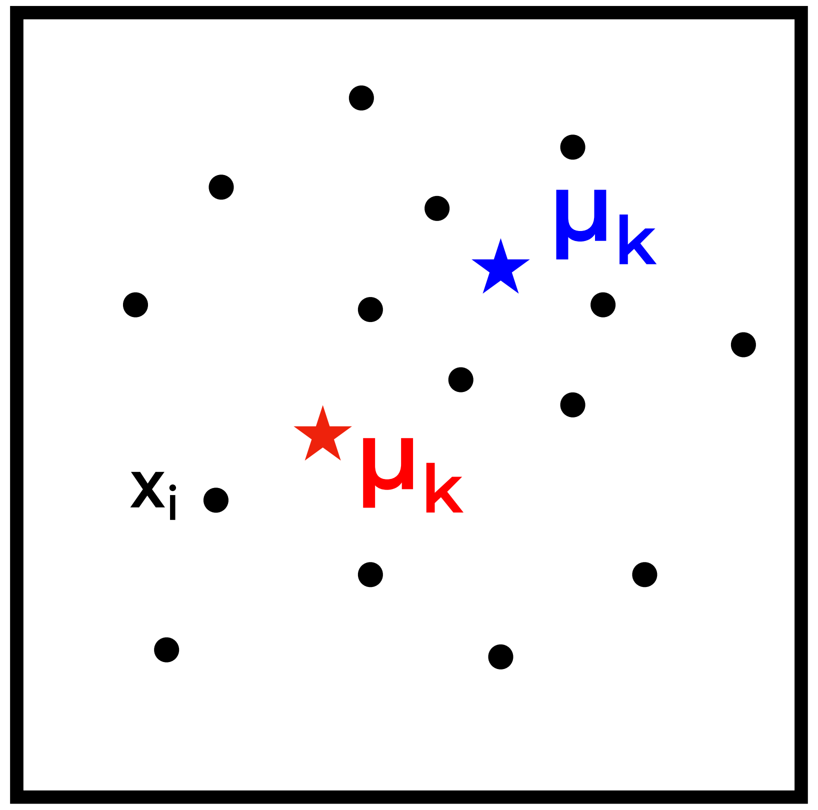
K-means clustering

- A particular clustering model (and accompanying algorithm)
  - There are K clusters. Each point belongs to a cluster. Clusters have centers: μ
- Objective: Find cluster centers μ_k that minimize the within cluster distance

$$\text{Objective} := \sum_{i=1}^{N} \sum_{k=1}^{K} \boxed{r_{ik}} ||x_i - \boxed{\mu_k}||^2$$

- Algorithm to minimize the objective:
  - Initialize the cluster centers
  - Until convergence:

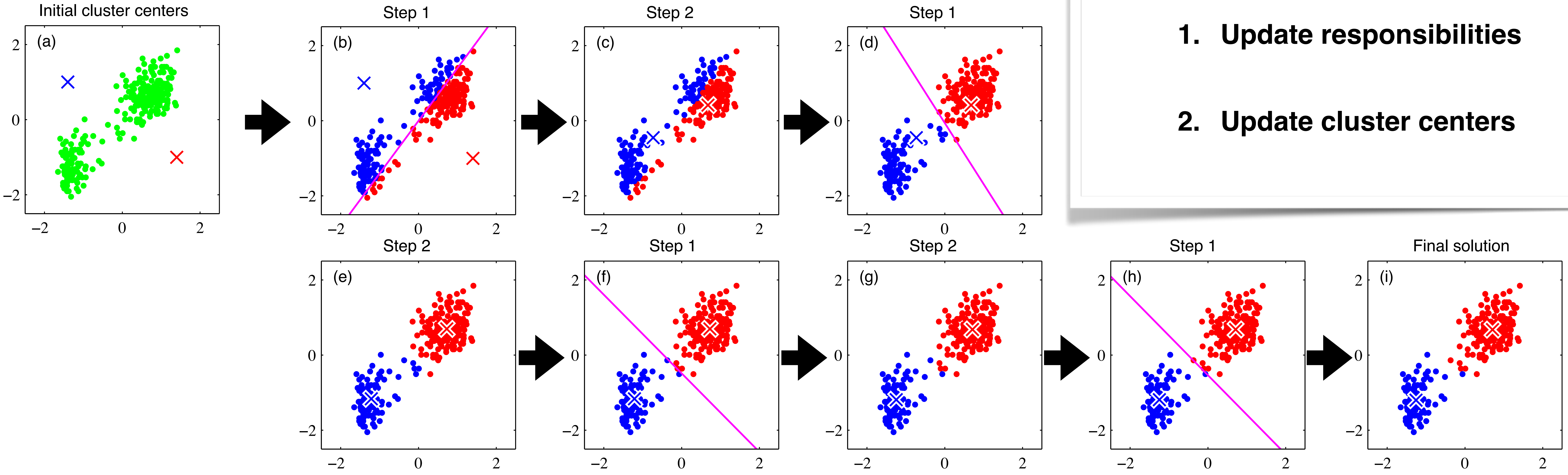
$$r = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}_{N \times 2}$$



1. Update responsibilities: r
2. Update cluster centers: $\mu_k \quad \forall k$

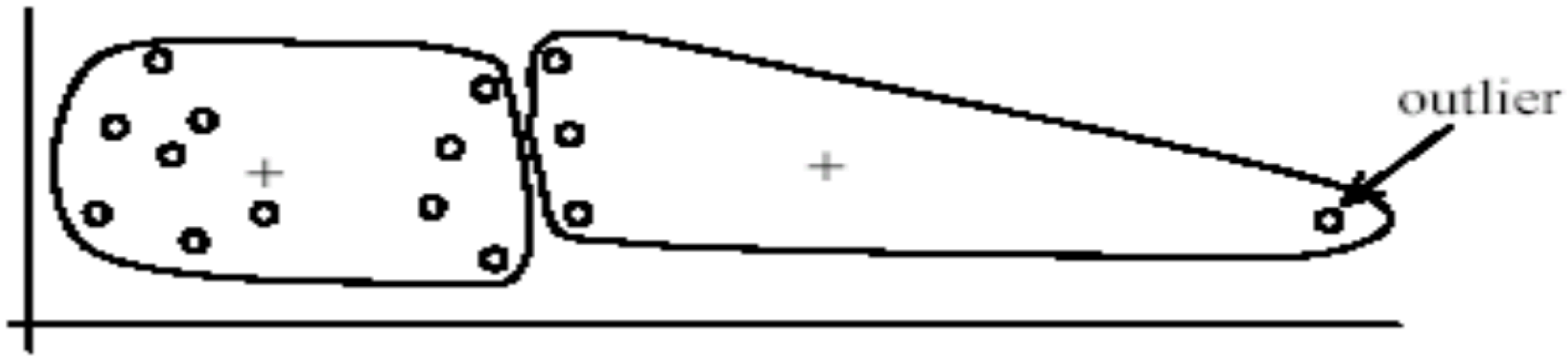
Algorithm

- **Initialize the cluster centers**
- **Until convergence:**
  1. **Update responsibilities**
  2. **Update cluster centers**

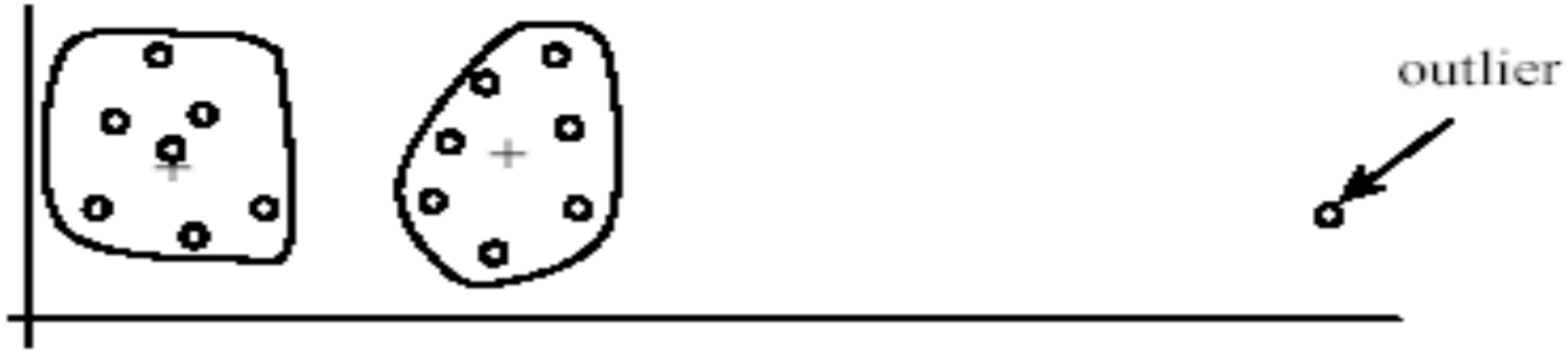


[Figure 9.1 from PRML]

K-means and sensitivity to outliers

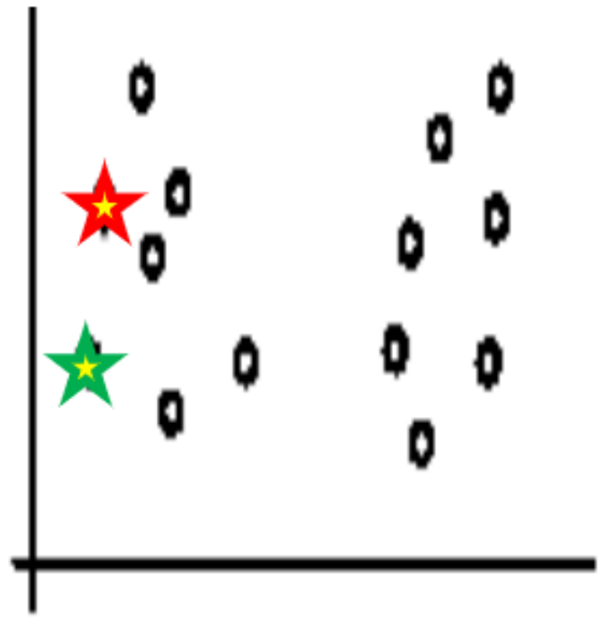


(A): Undesirable clusters

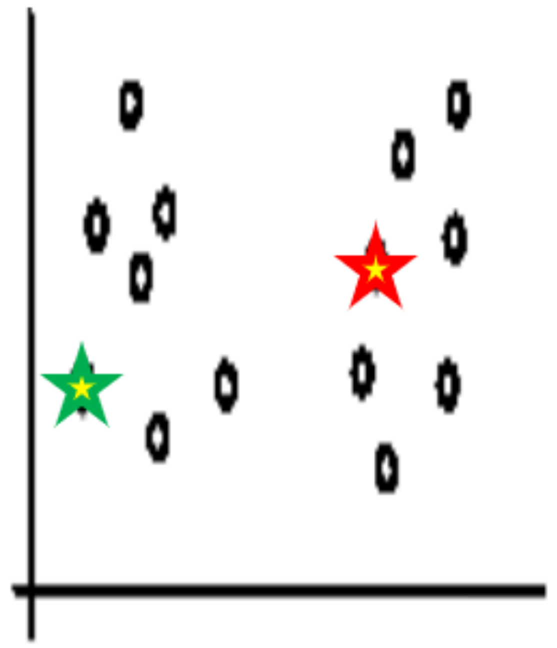


(B): Ideal clusters

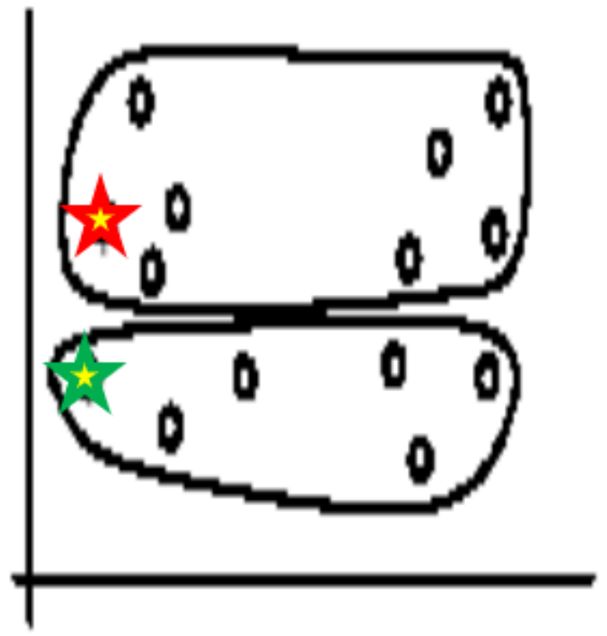
K-means and sensitivity to initial seeds



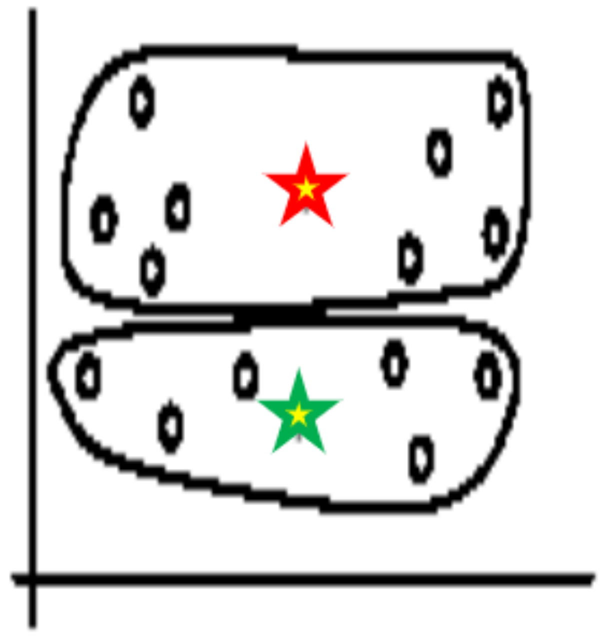
Random selection of seeds (centroids)



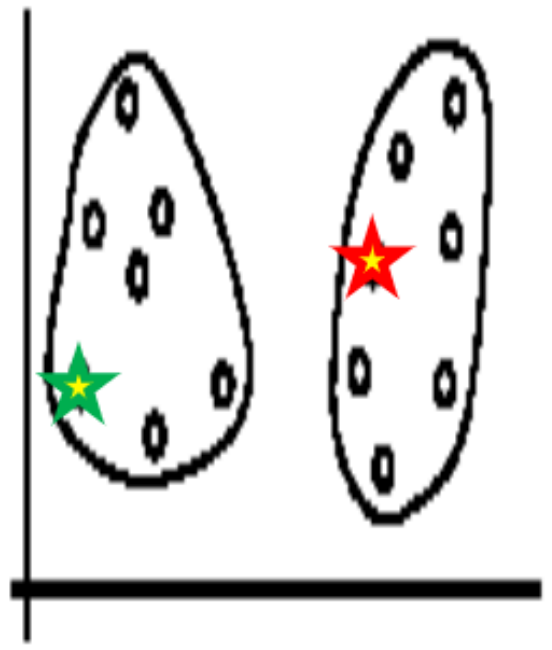
Random selection of seeds (centroids)



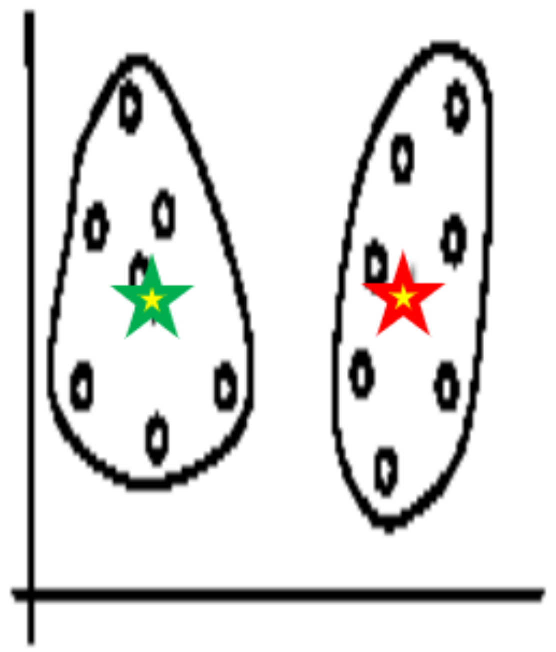
Iteration 1



Iteration 2



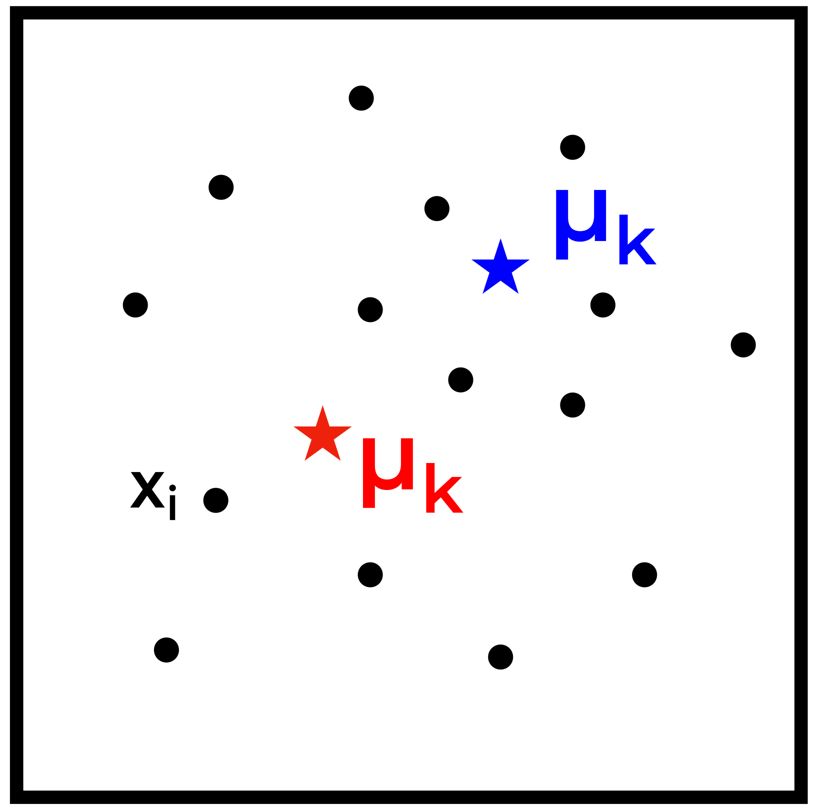
Iteration 1



Iteration 2

A probabilistic approach to k-means clustering

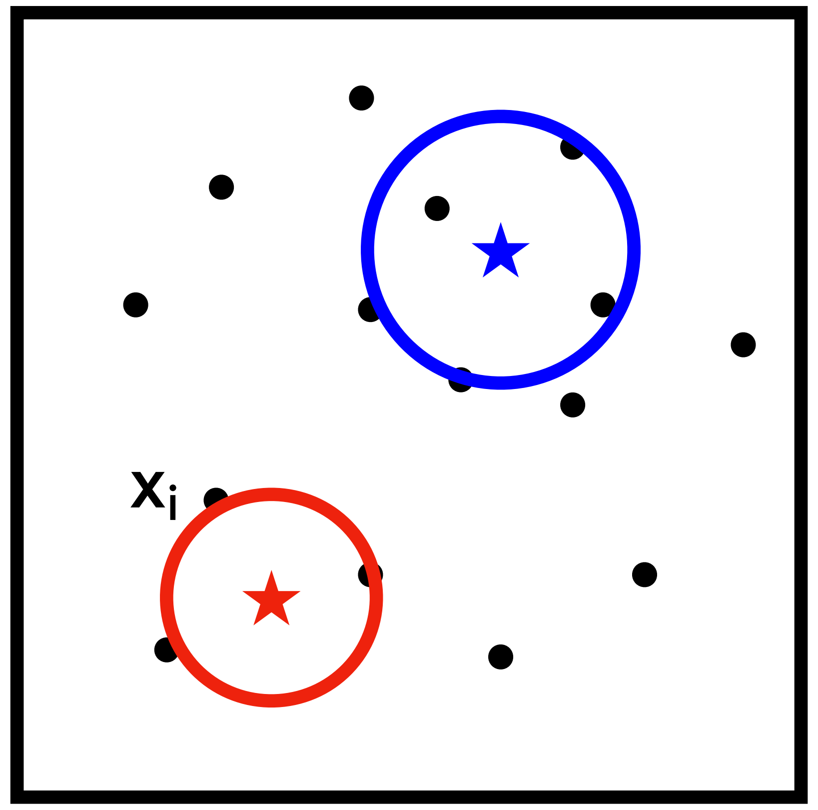
K-means Clustering



Responsibility   center

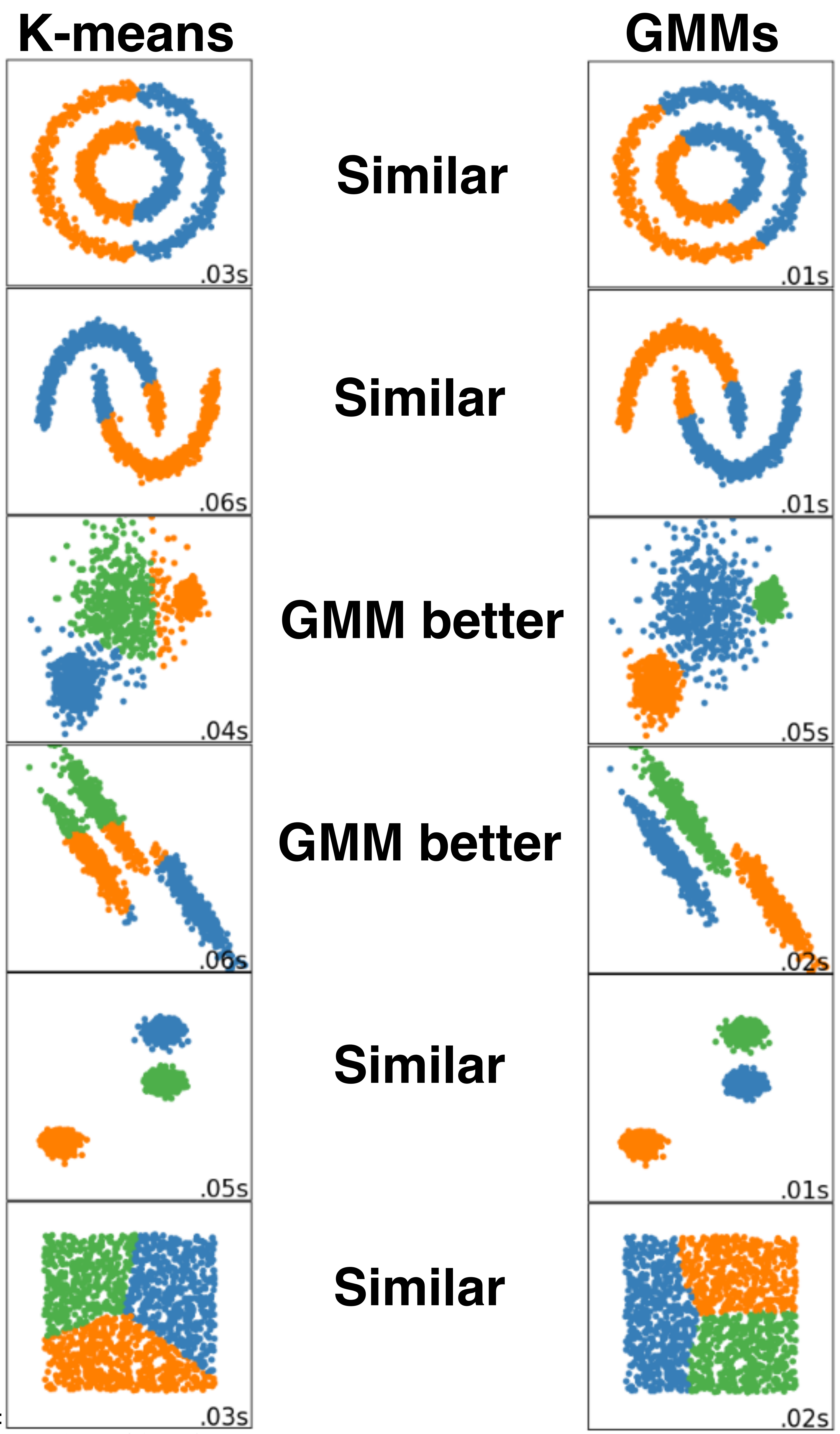
$$\text{Objective} := \sum_{i=1}^{N} \sum_{k=1}^{K} \boxed{r_{ik}} \, \| x_i - \boxed{\mu_k} \|^2$$

Soft K-means Clustering



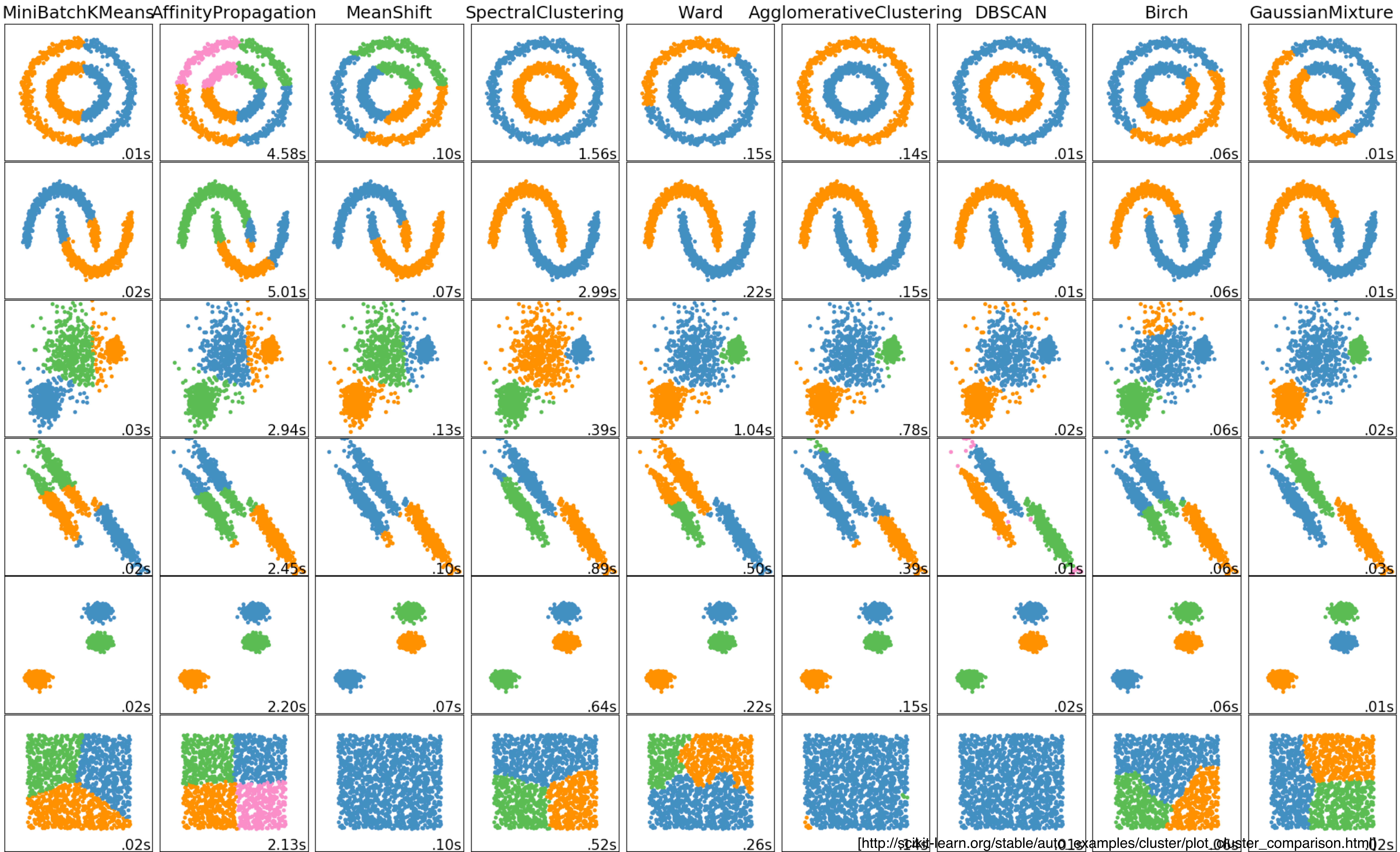
- Responsibilities are continuous [0, 1]
  - Each cluster has a responsibility: π_k
- Each cluster models data using a Gaussian:

$$\mathcal{N}(x_i \mid \mu_k, \Sigma_k)$$



Comparing K-means to GMMs

- GMMs learns covariance matrix
  - Per cluster variance
  - Covariance terms
- GMMs has many more parameters
  - Covariance matrix (MxM)

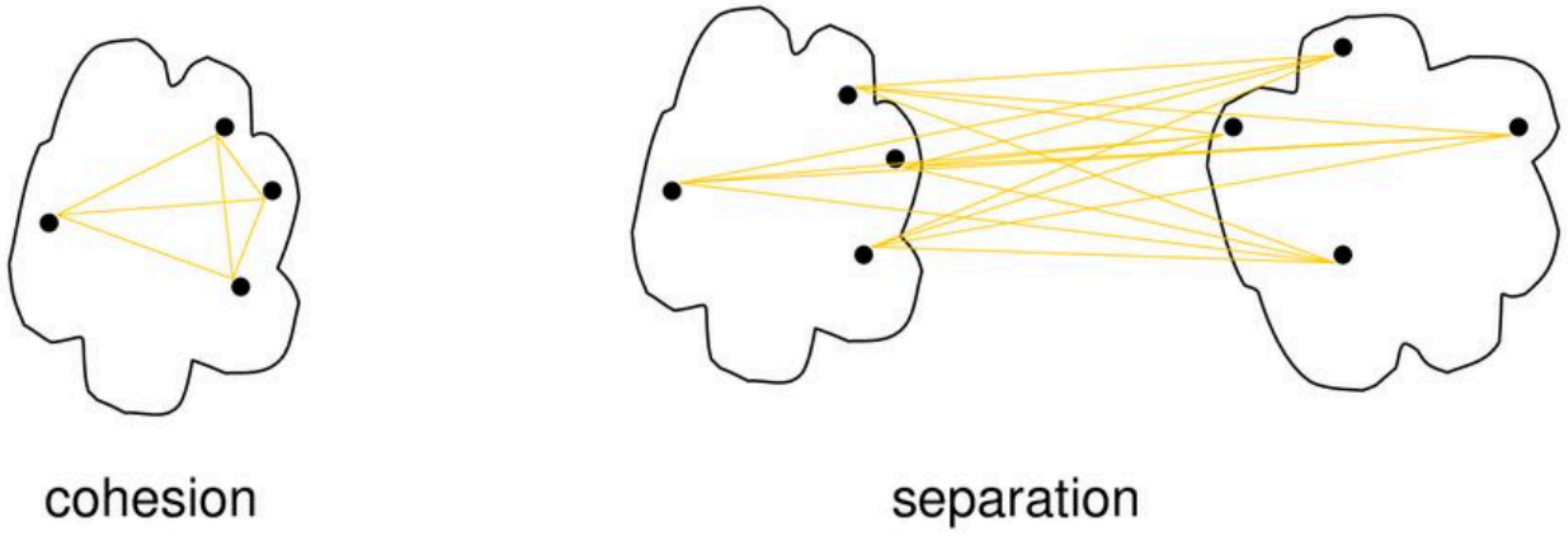


La

[http://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html]

Additional info

- **Evaluation**
  - Comparing different clustering algorithms is a difficult task. No one knows the correct clusters!
  - Internal evaluation and external evaluation
  - E.g., internal evaluation: Cohesion and Separation



Additional info

Cluster Cohesion: Measures how closely related are objects in a cluster

Cluster Separation: Measure how distinct or well-separated a cluster is from other clusters

Example: Squared Error

- Cohesion is measured by the within cluster sum of squares (SSE)

$$WSS = \sum_i \sum_{x \in C_i} (x - c_i)^2$$

We want this to be small

- Separation is measured by the between cluster sum of squares

$$BSS = \sum_i m_i (c - c_i)^2$$

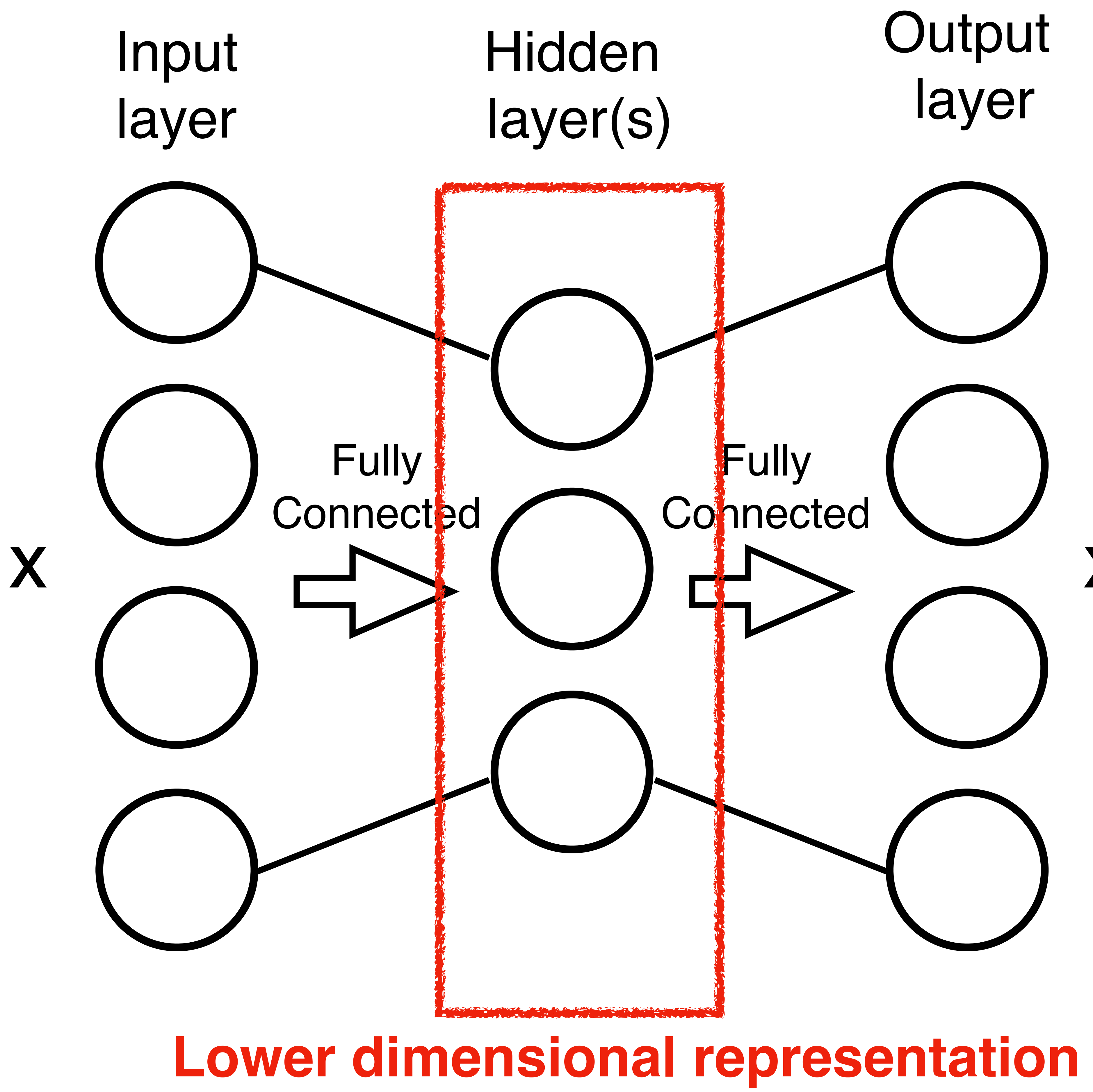
We want this to be large

Additional info

- **Evaluation**
  - Comparing different clustering algorithms is a difficult task. No one knows the correct clusters!
  - Internal evaluation and external evaluation
  - E.g., internal evaluation: Cohesion and Separation
- For some data, **hierarchical Clustering** is more appropriate, e.g., **biological taxonomy**

Autoencoders

- **A neural network architecture for unsupervised learning**



Objective:
How well the network predicts X ?

$$\text{Loss} := \sum_{i=1}^{N} (x_i - \hat{x}_i)^2$$
$$= \sum_{i=1}^{N} (x_i - f_2(f_1(x)))^2$$