Machine Learning I
60629A

Summary

Unsupervised Learning

— Week #7

Recall

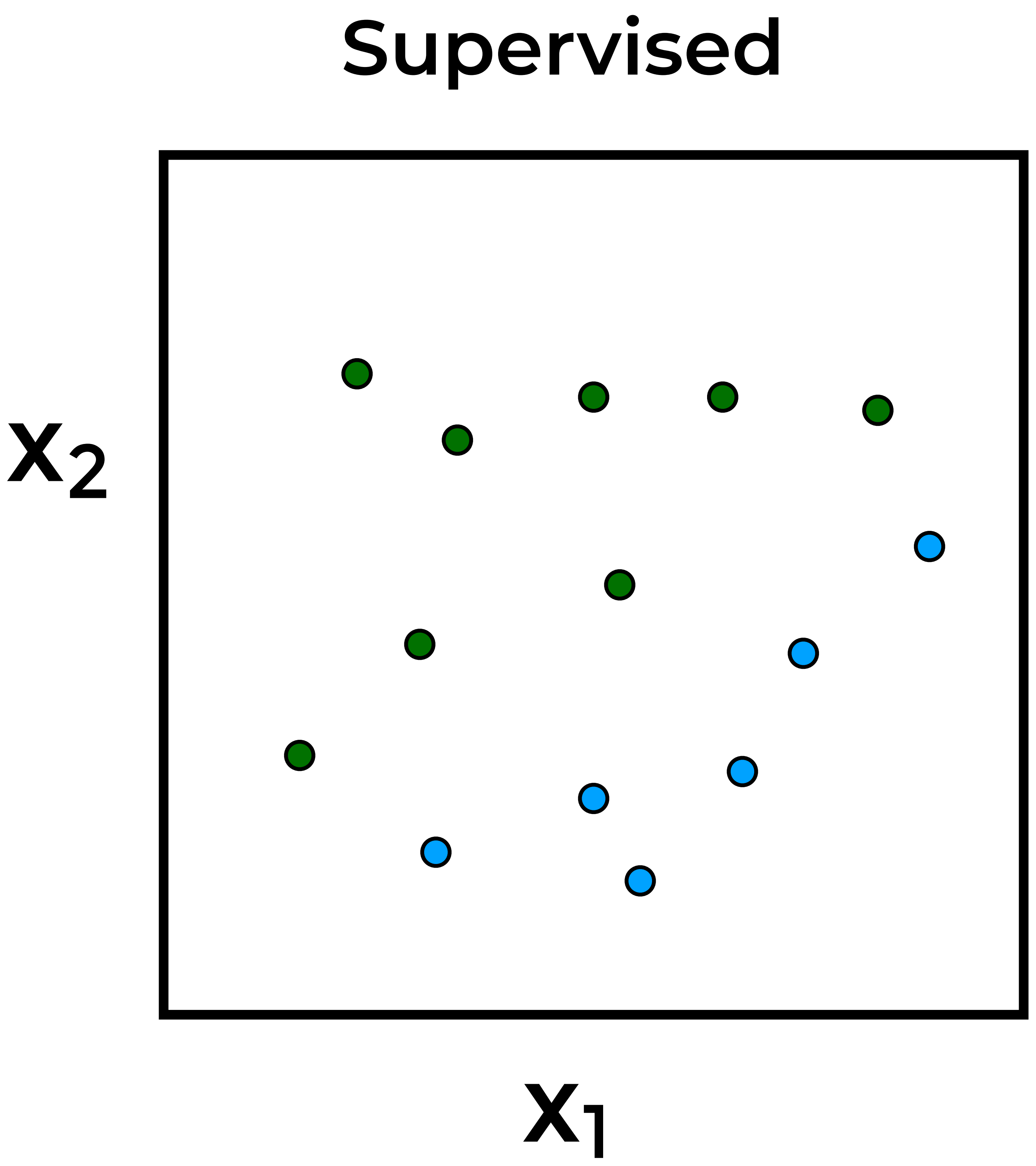
1. Unsupervised

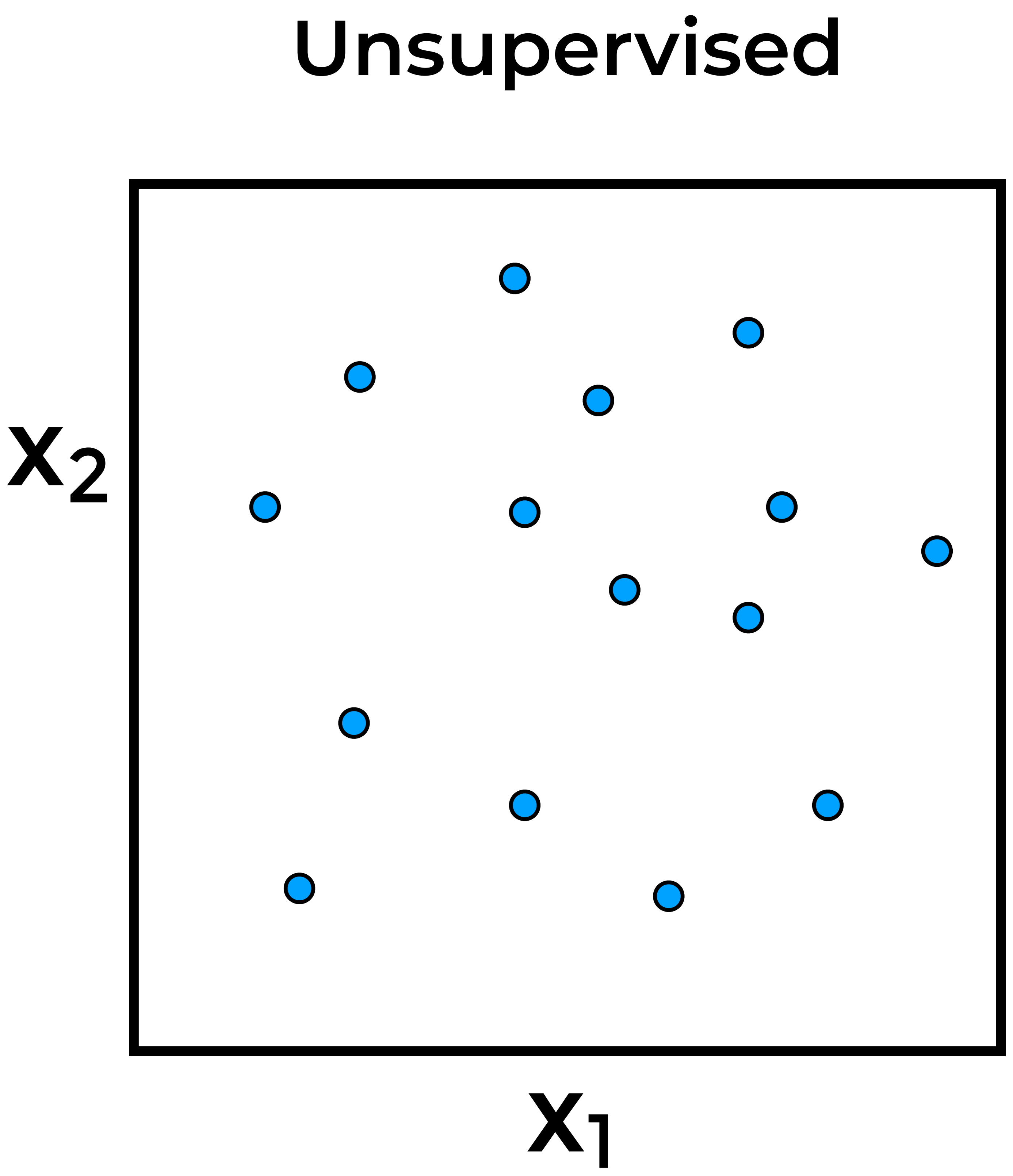
$$\{\mathbf{x}_i\}_{i=0}^{n}$$

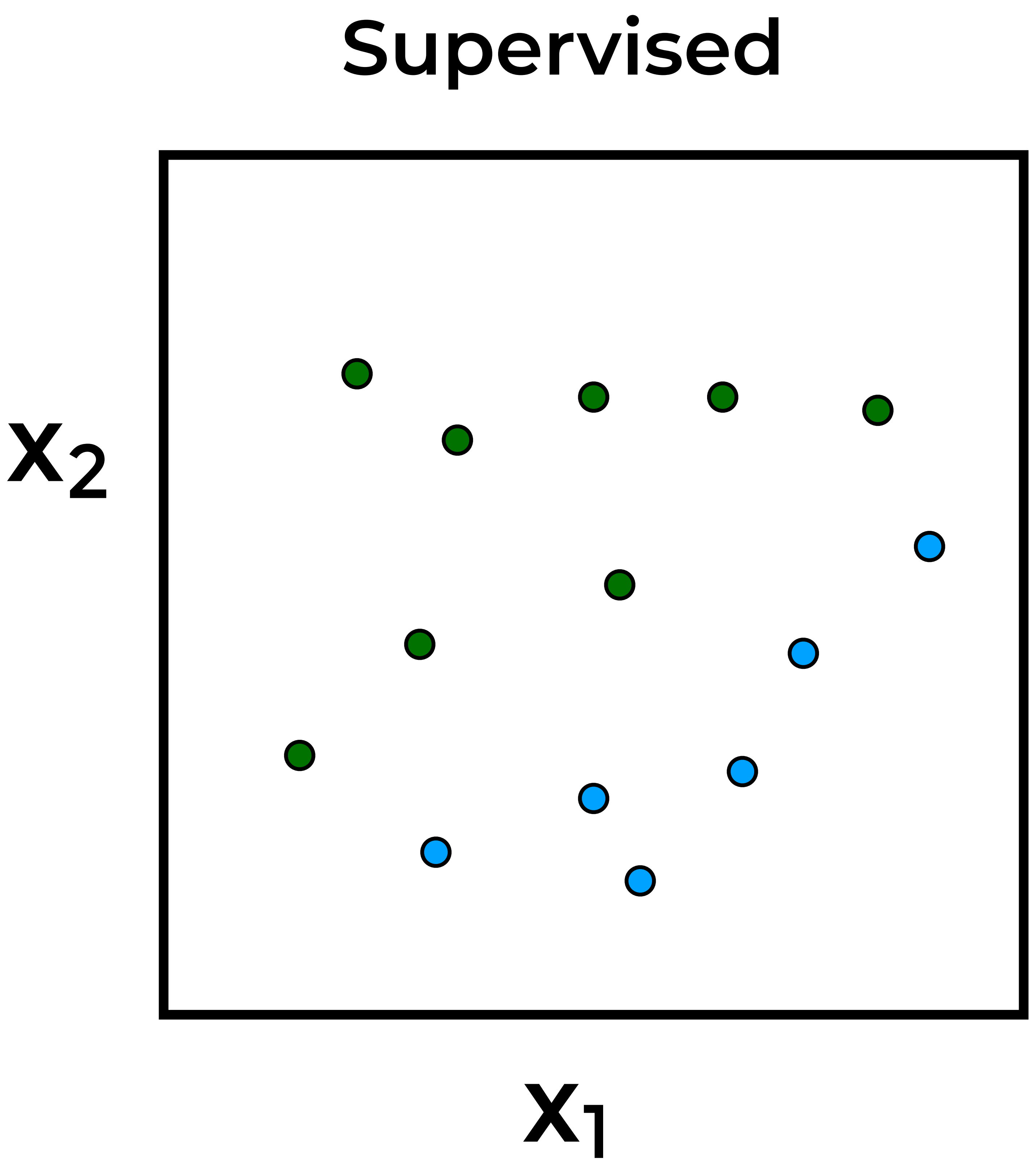
- Experience examples alone
- Learn "useful properties of the structure of the data"
  - E.g., clustering, density modeling (p(x)), PCA, FA.

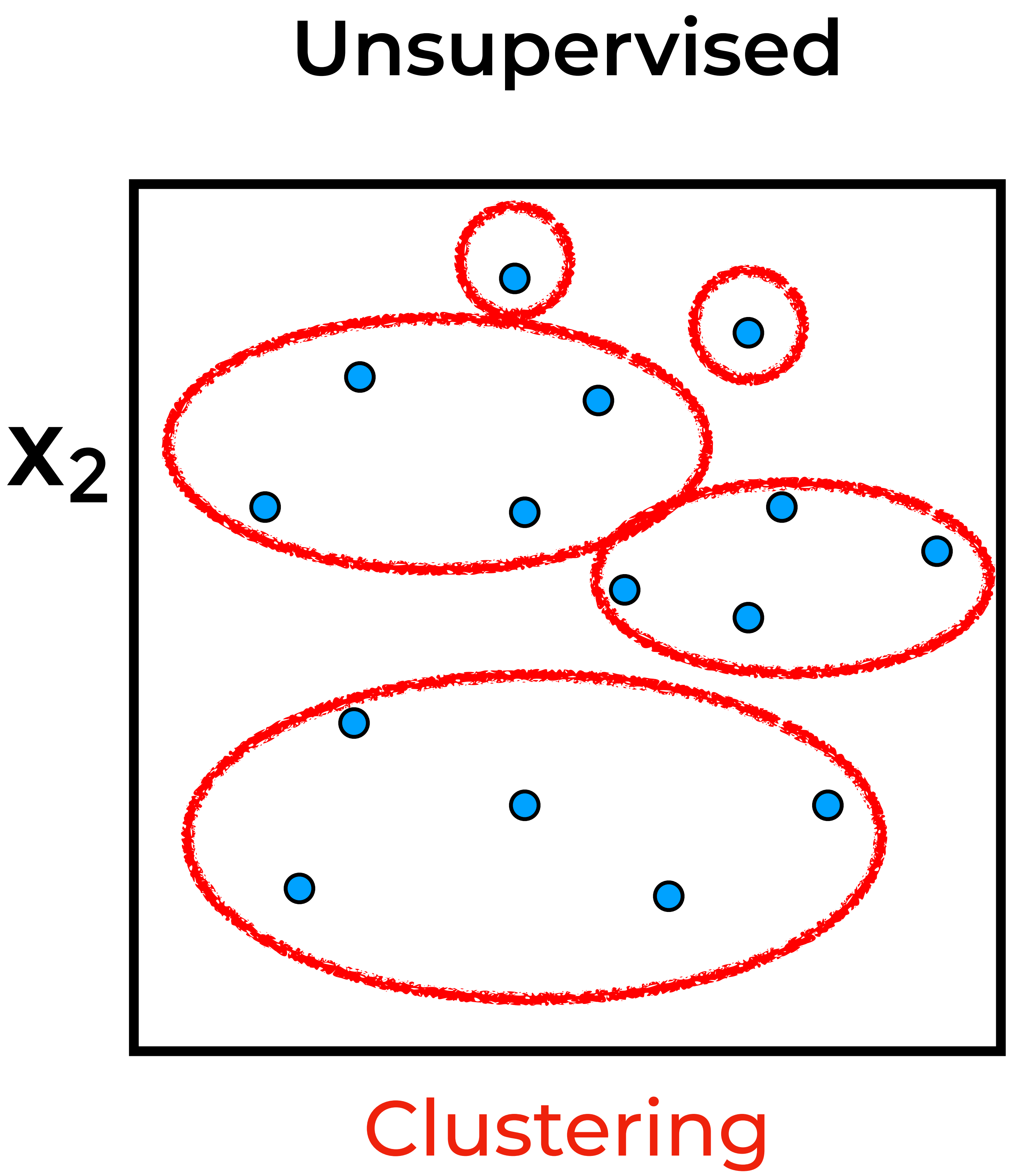
Different tasks

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


Supervised
x_2
x_1

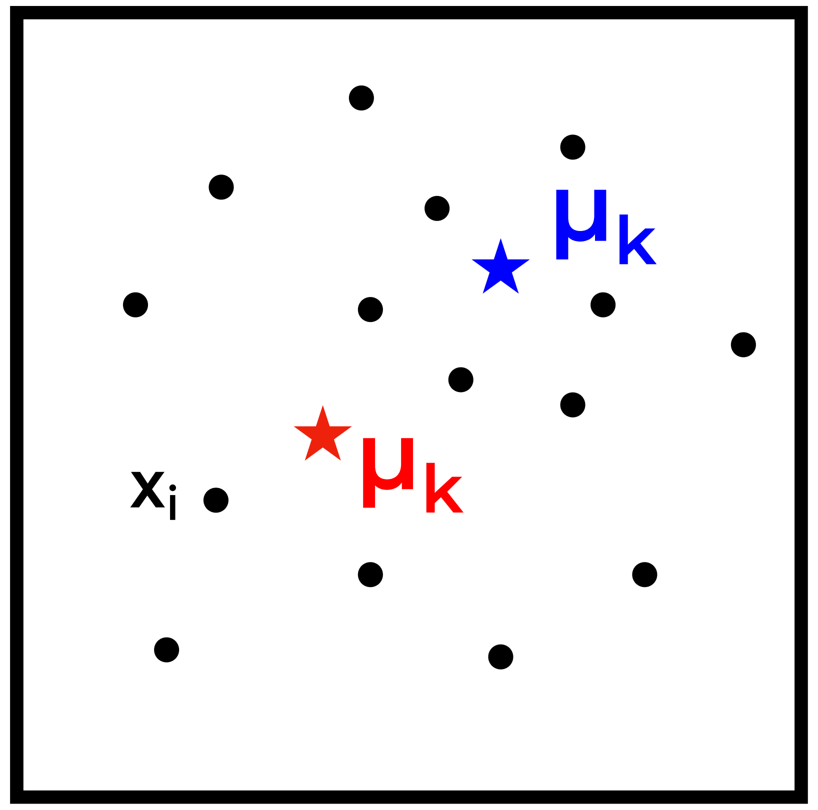

Unsupervised
x_2
x_1


Supervised
X2
X1


Unsupervised
X2
Clustering

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

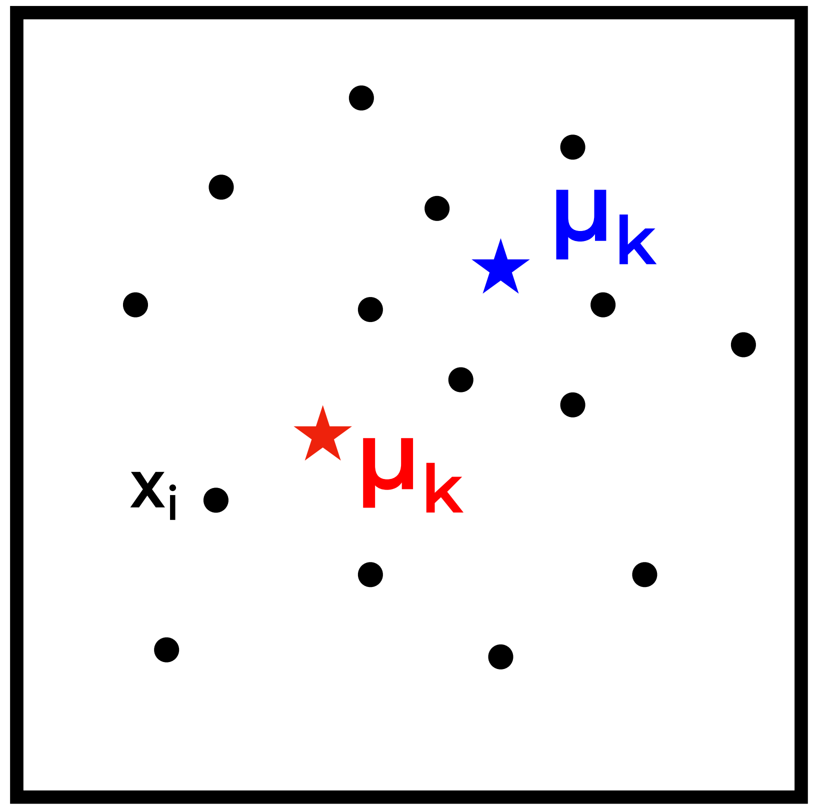


K-means clustering

- 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

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

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



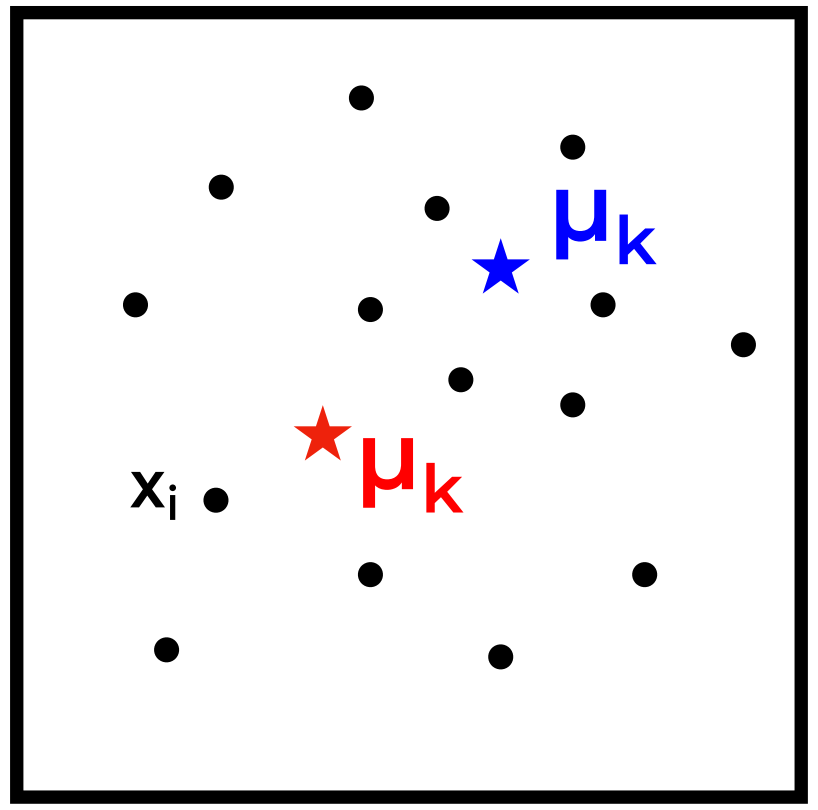
K-means clustering

- 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

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

- Algorithm to minimize the objective:
    - Initialize the cluster centers
    - Until convergence:
        1. Update responsibilities: r
        2. Update cluster centers: $\mu_k \quad \forall k$

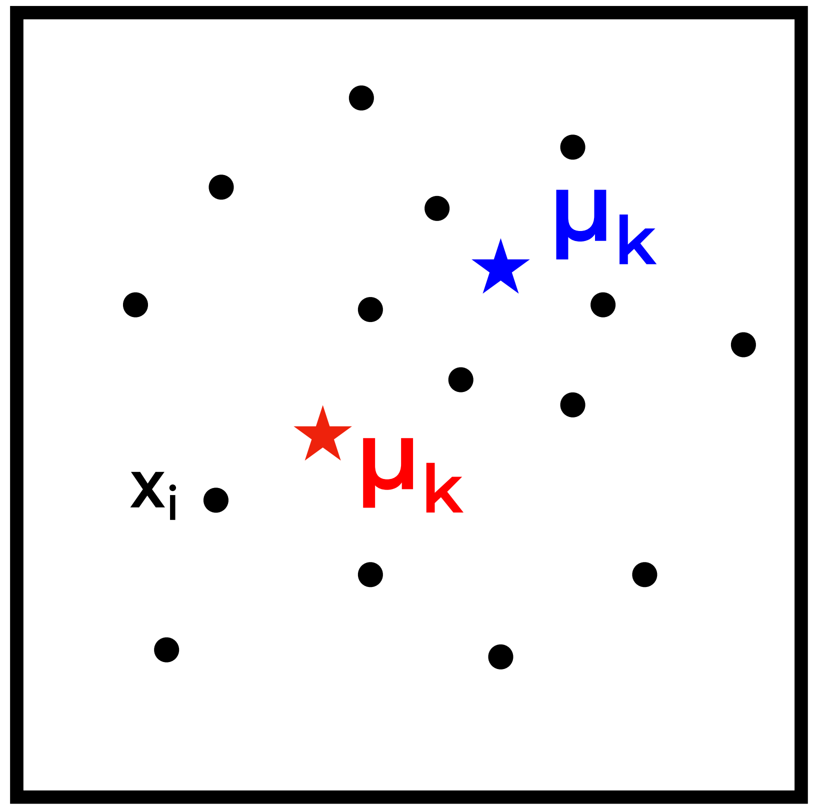
$$r = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}_{N \times 2}$$



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

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



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

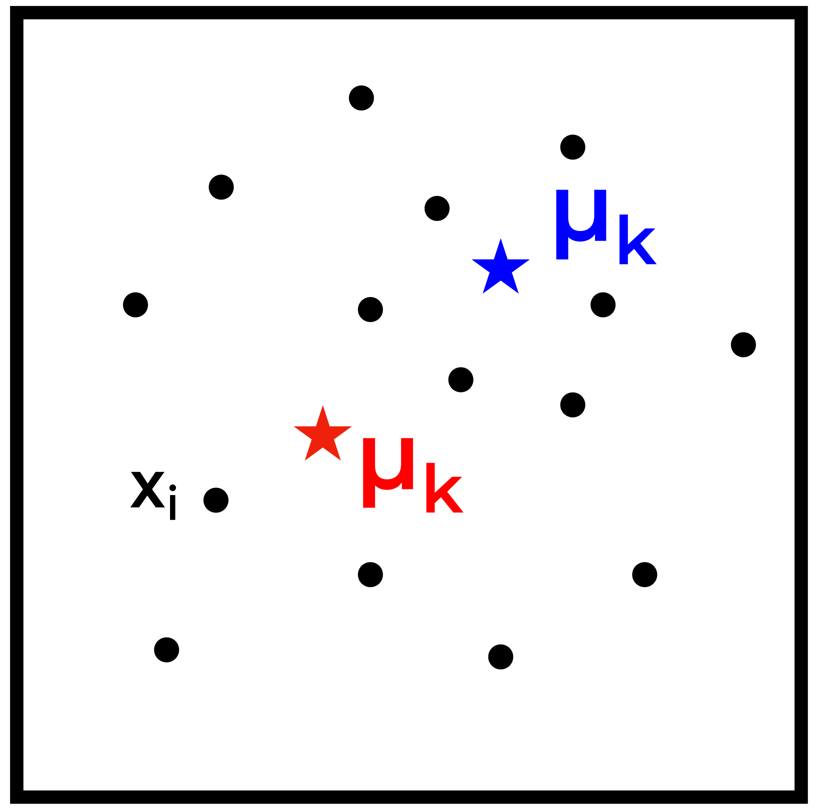
- Algorithm to minimize the objective:
  - Initialize the cluster centers
  - Until convergence:
    1. Update responsibilities: r
    2. Update cluster centers: $\mu_k \quad \forall k$

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

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

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



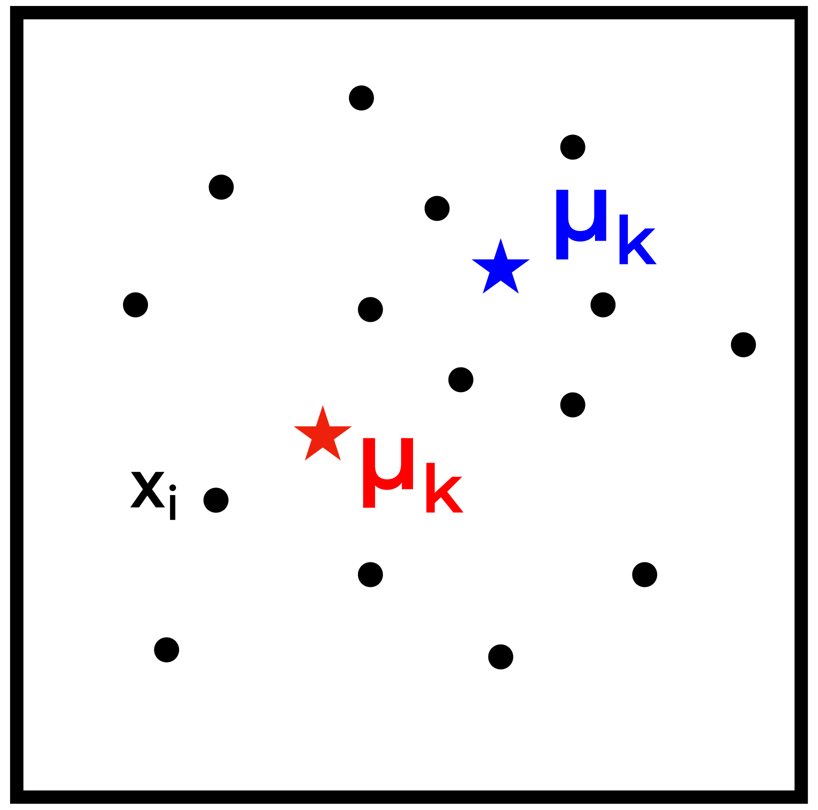
- Algorithm to minimize the objective:
  - Initialize the cluster centers
  - Until convergence:
    1. Update responsibilities: r
    2. Update cluster centers: $\mu_k \ \forall k$

K-means clustering

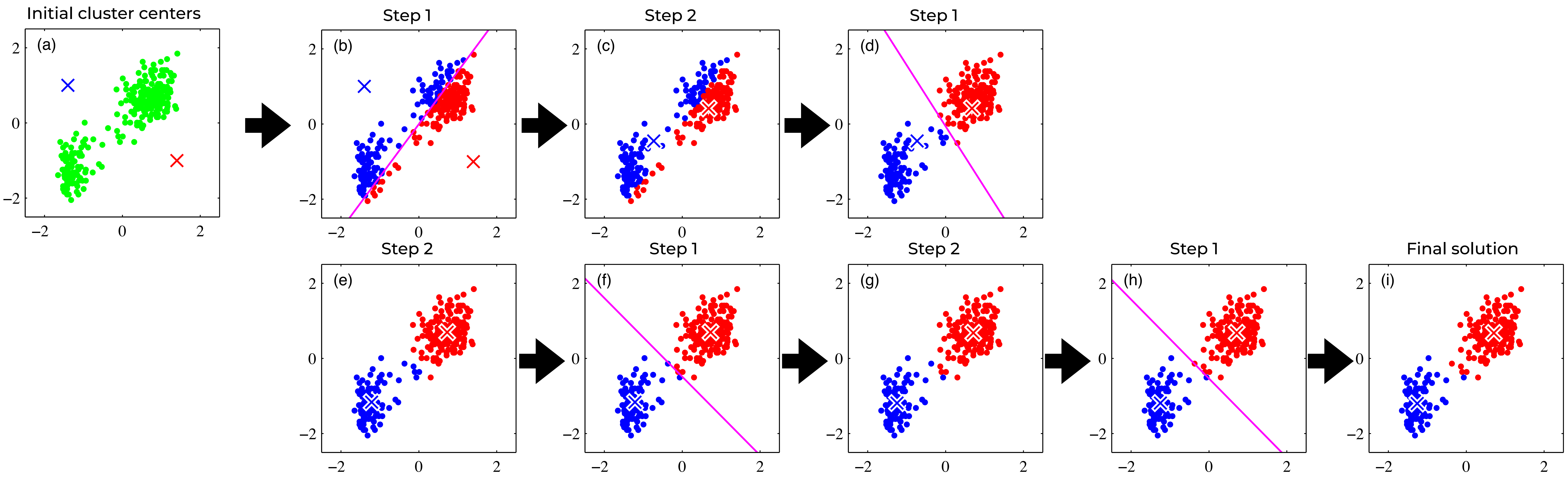
- A particular clustering model (and accompanying algorithm)
  - There are K clusters. Each point belongs to a cluster. Clusters have centers: μ
- Objective: Find cluster centers μ_k that minimize the within cluster distance

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

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



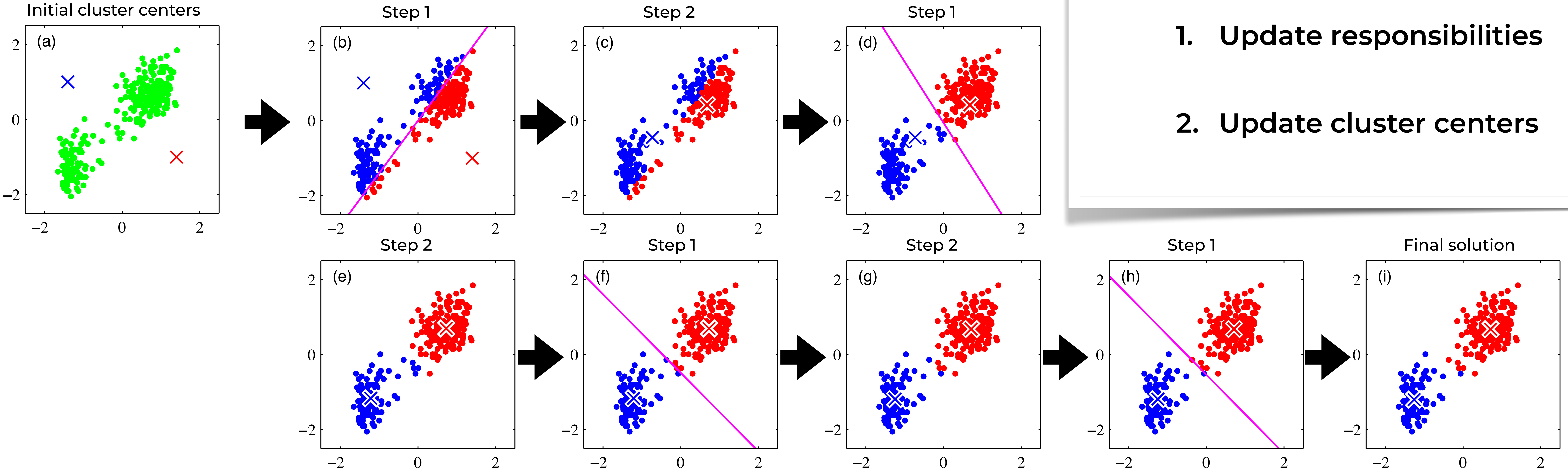
- Algorithm to minimize the objective:
  - Initialize the cluster centers
  - Until convergence:
    1. Update responsibilities: r
    2. Update cluster centers: $\mu_k \quad \forall k$



[Figure 9.1 from PRML]

Algorithm

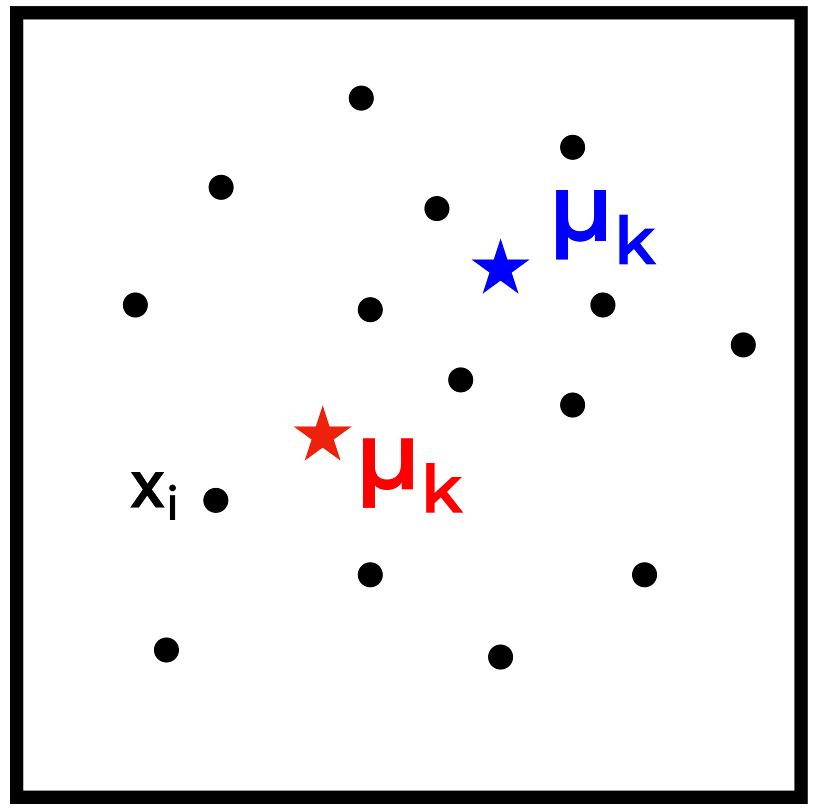
- Initialize the cluster centers
- Until convergence:
  1. Update responsibilities
  2. Update cluster centers



[Figure 9.1 from PRML]

A probabilistic approach to k-means clustering

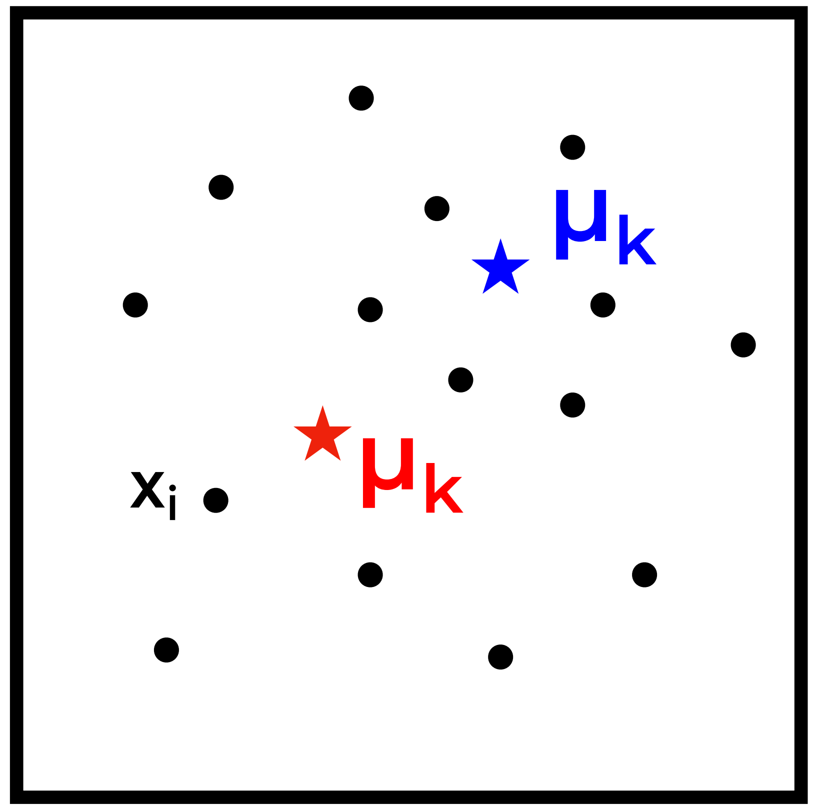
K-means Clustering



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

A probabilistic approach to k-means clustering

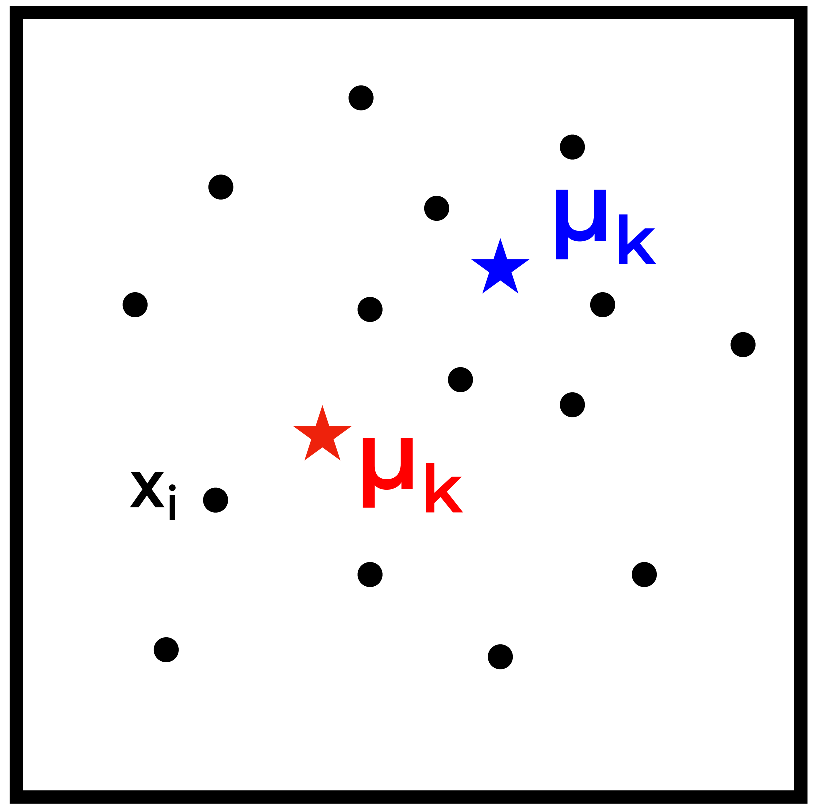
K-means Clustering



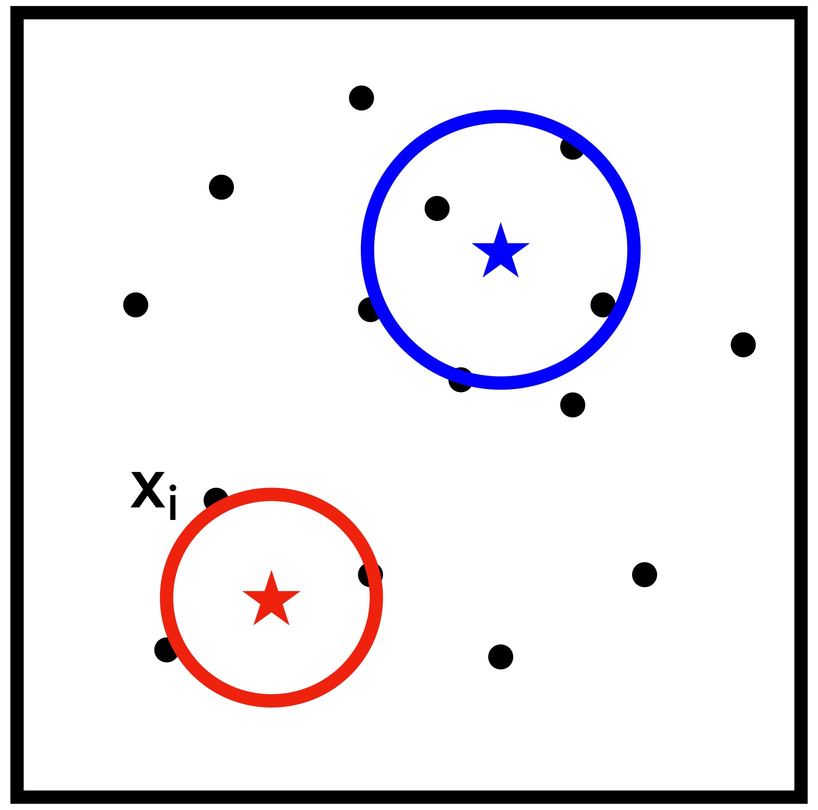
$$\text{Objective} := \sum_{i=1}^{N} \sum_{k=1}^{K} \overset{\text{Responsibility}}{\boxed{r_{ik}}} \| x_i - \overset{\text{center}}{\boxed{\mu_k}} \|^2$$

A probabilistic approach to k-means clustering

K-means Clustering



Soft K-means Clustering

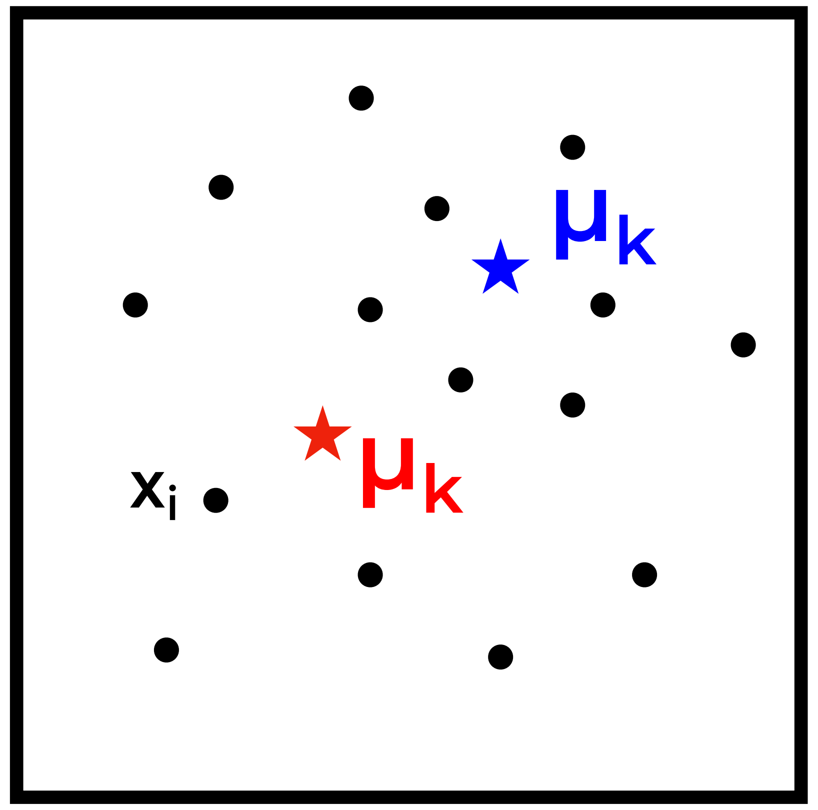


Responsibility center

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

A probabilistic approach to k-means clustering

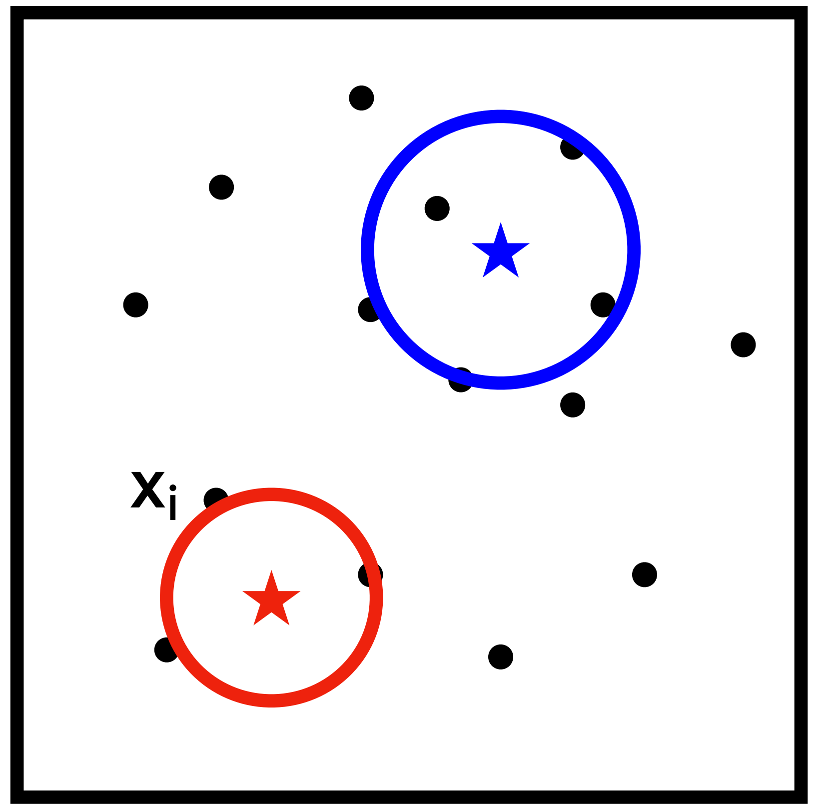
K-means Clustering



Responsibility center

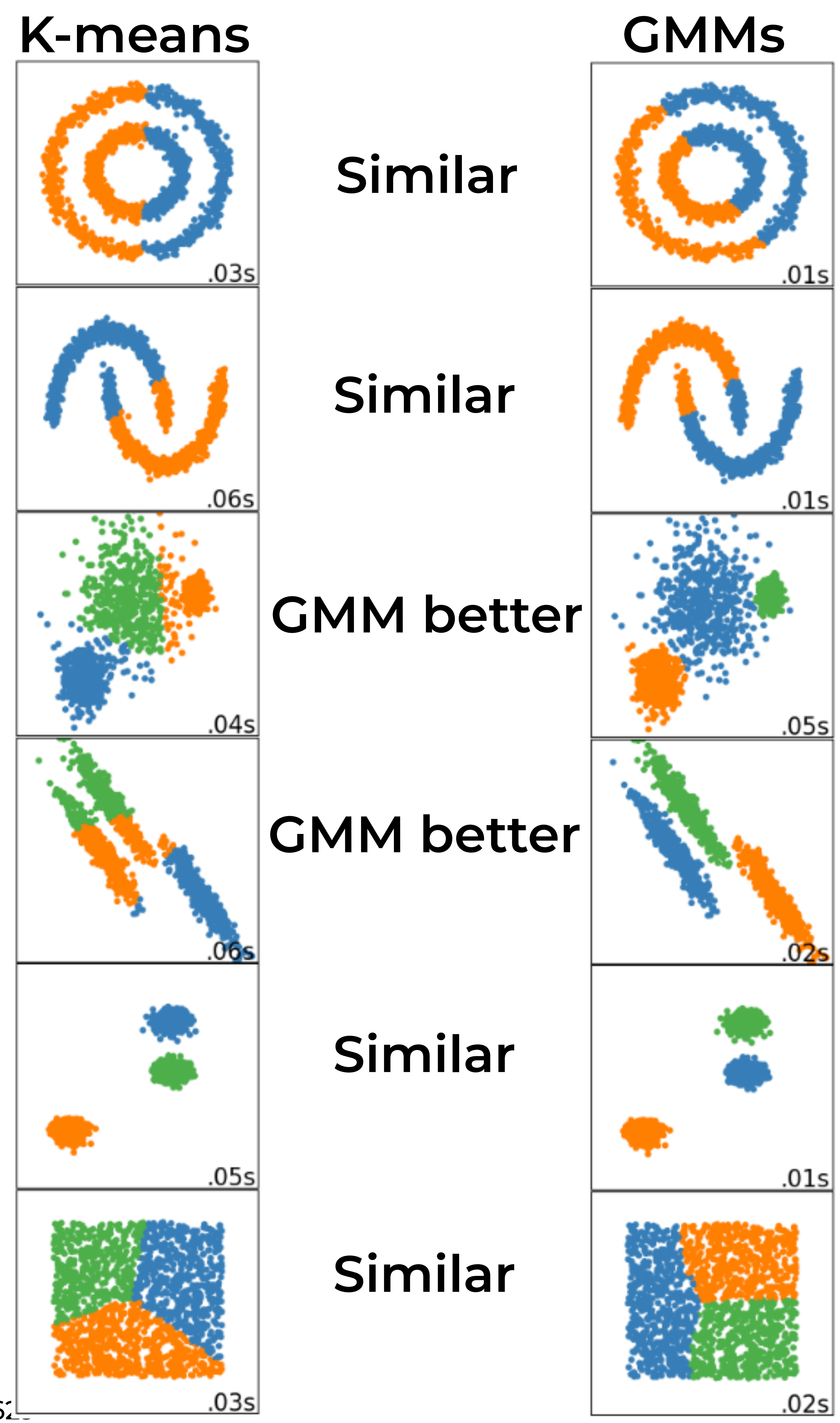
$$\text{Objective} := \sum_{i=1}^{N} \sum_{k=1}^{K} \boxed{r_{ik}} \| x_i - \boxed{\mu_k} \|^2$$

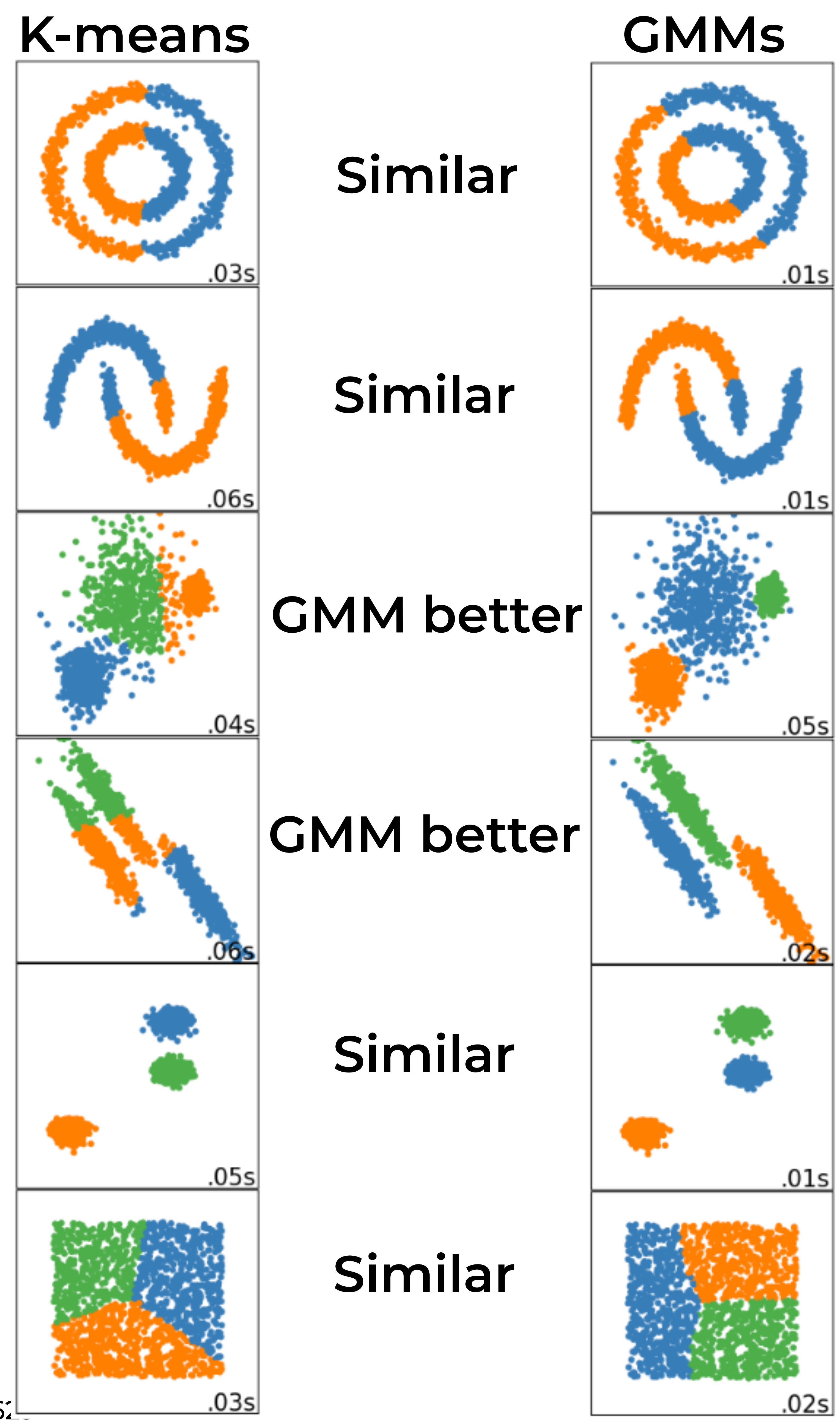
Soft K-means Clustering



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

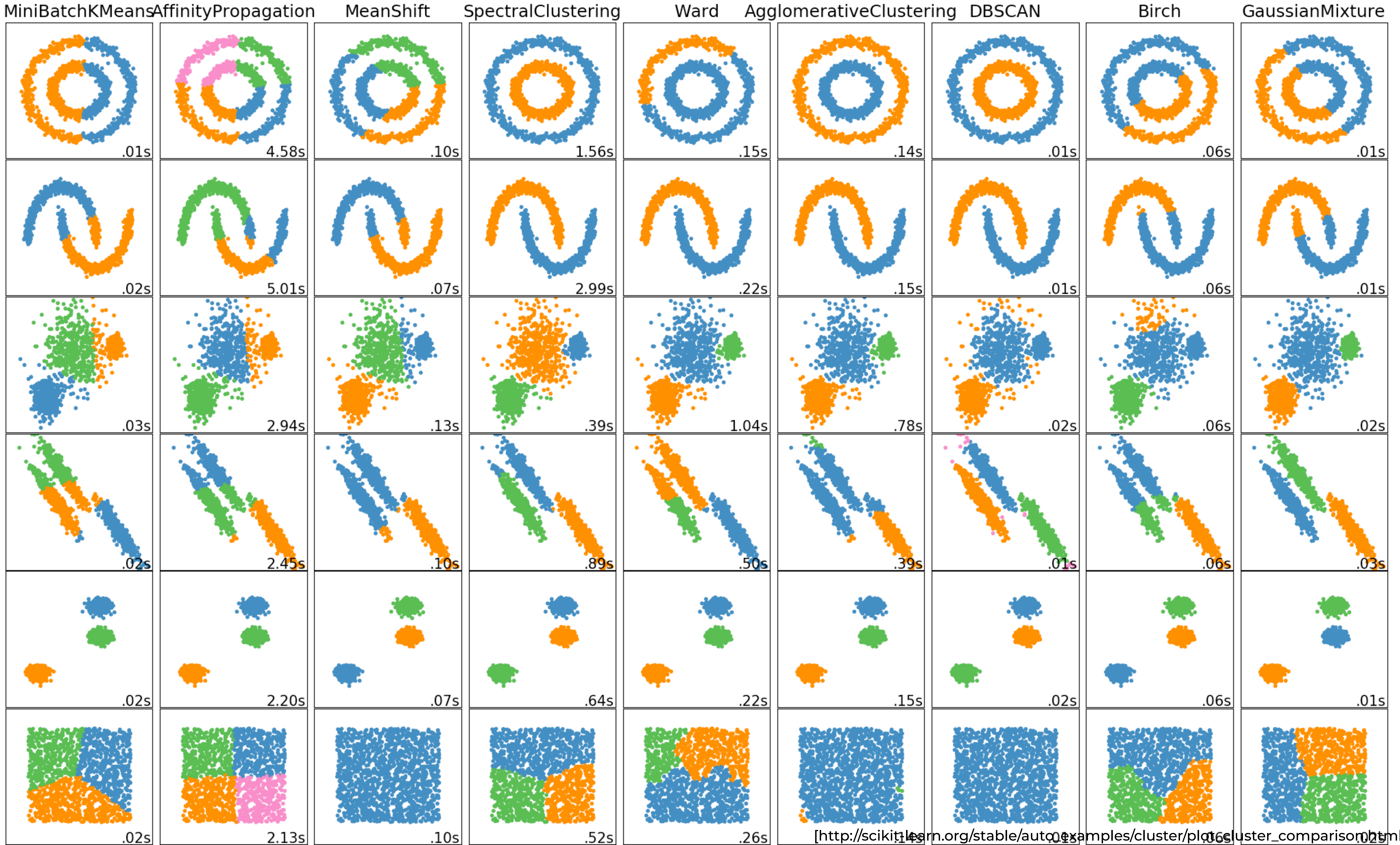
$$\mathcal{N}(x_i \mid \mu_k, \Sigma_k)$$


K-means
GMMs
Similar
Similar
GMM better
GMM better
Similar
Similar
.03s
.01s
.06s
.01s
.04s
.05s
.06s
.02s
.05s
.01s
.03s
.02s



Comparing K-means to GMMs

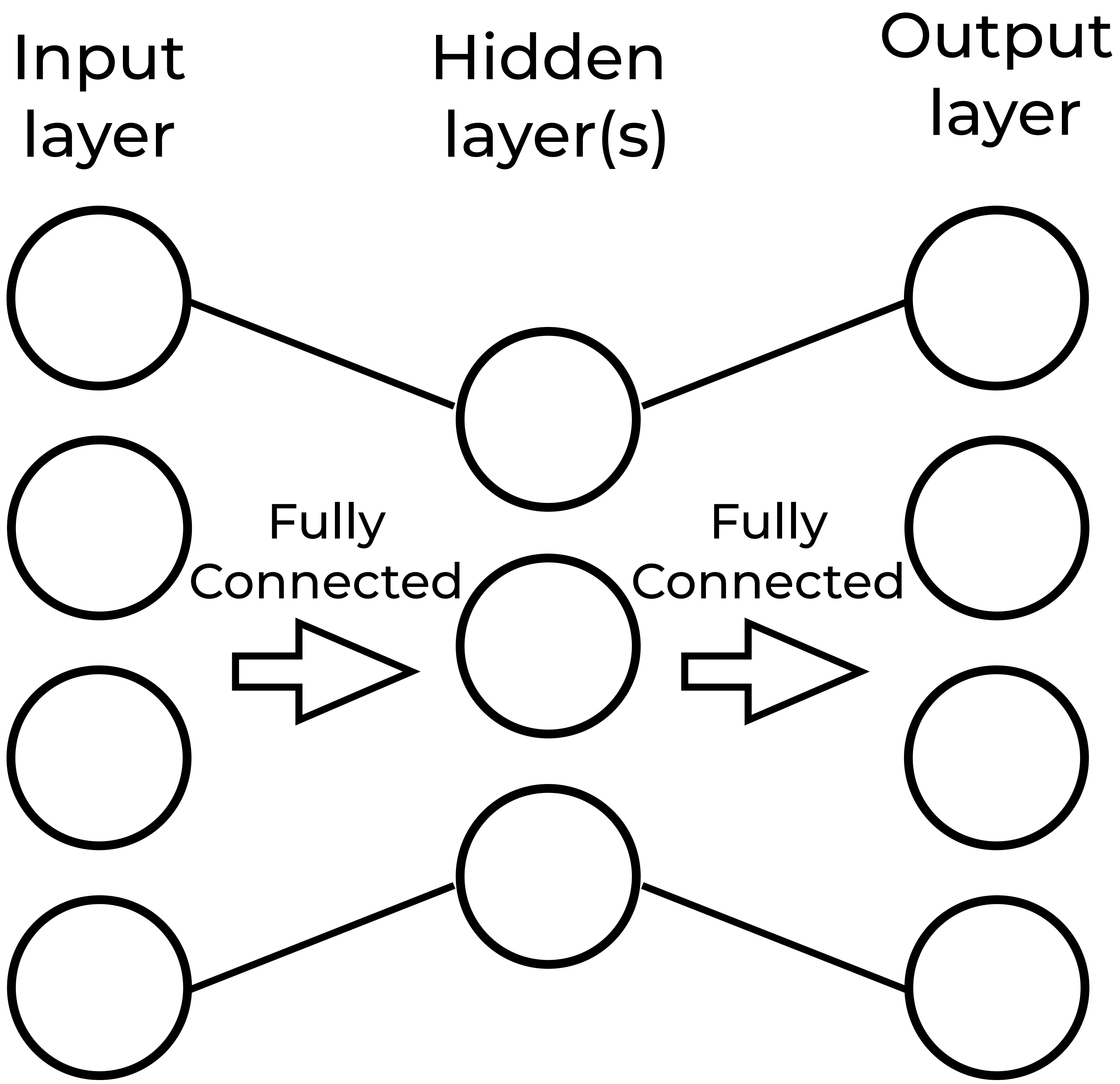
- GMMs learns covariance matrix
  - Per cluster variance
  - Covariance terms
- GMMs has many more parameters
  - Covariance matrix (MxM)



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

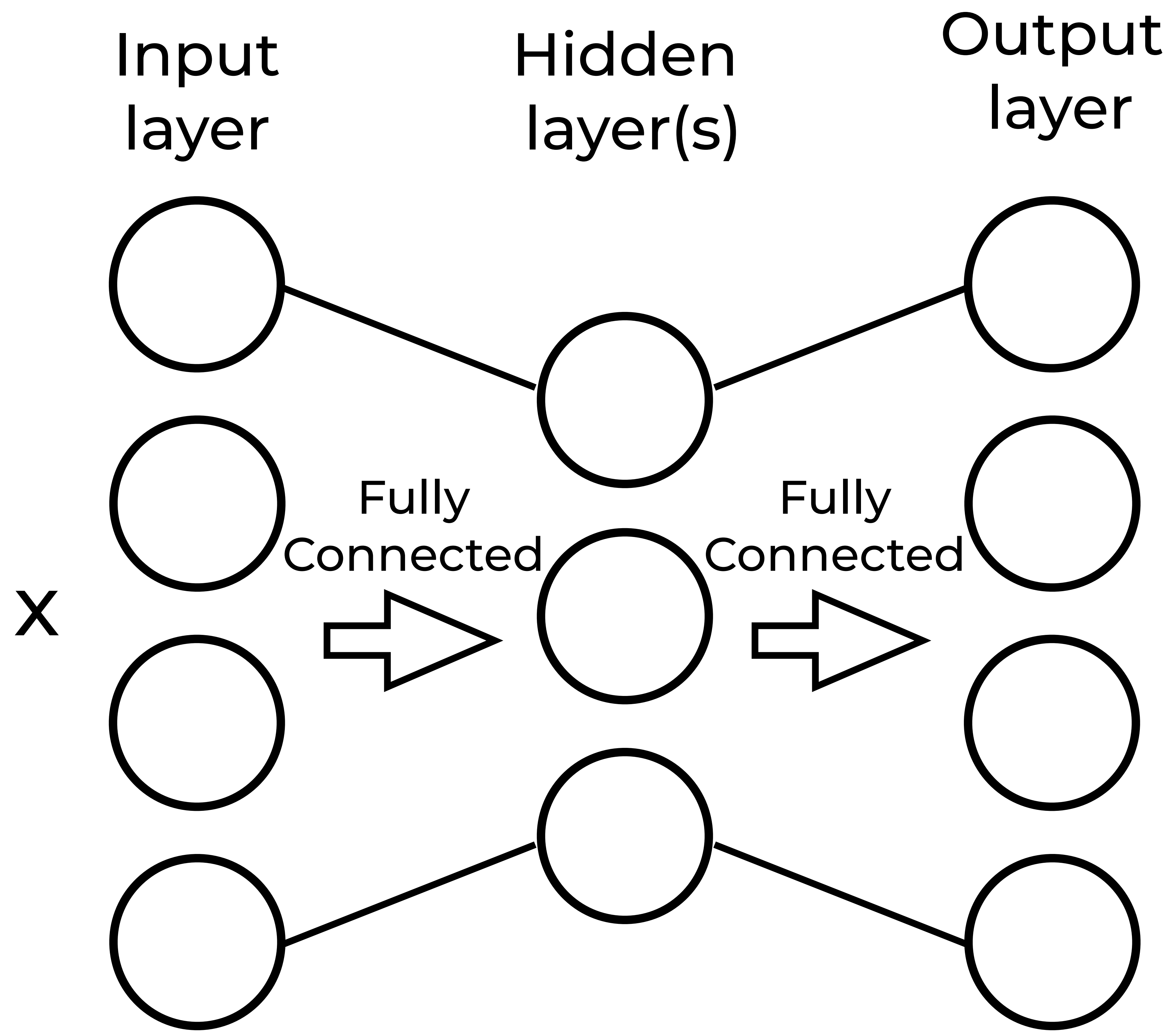
Autoencoders

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



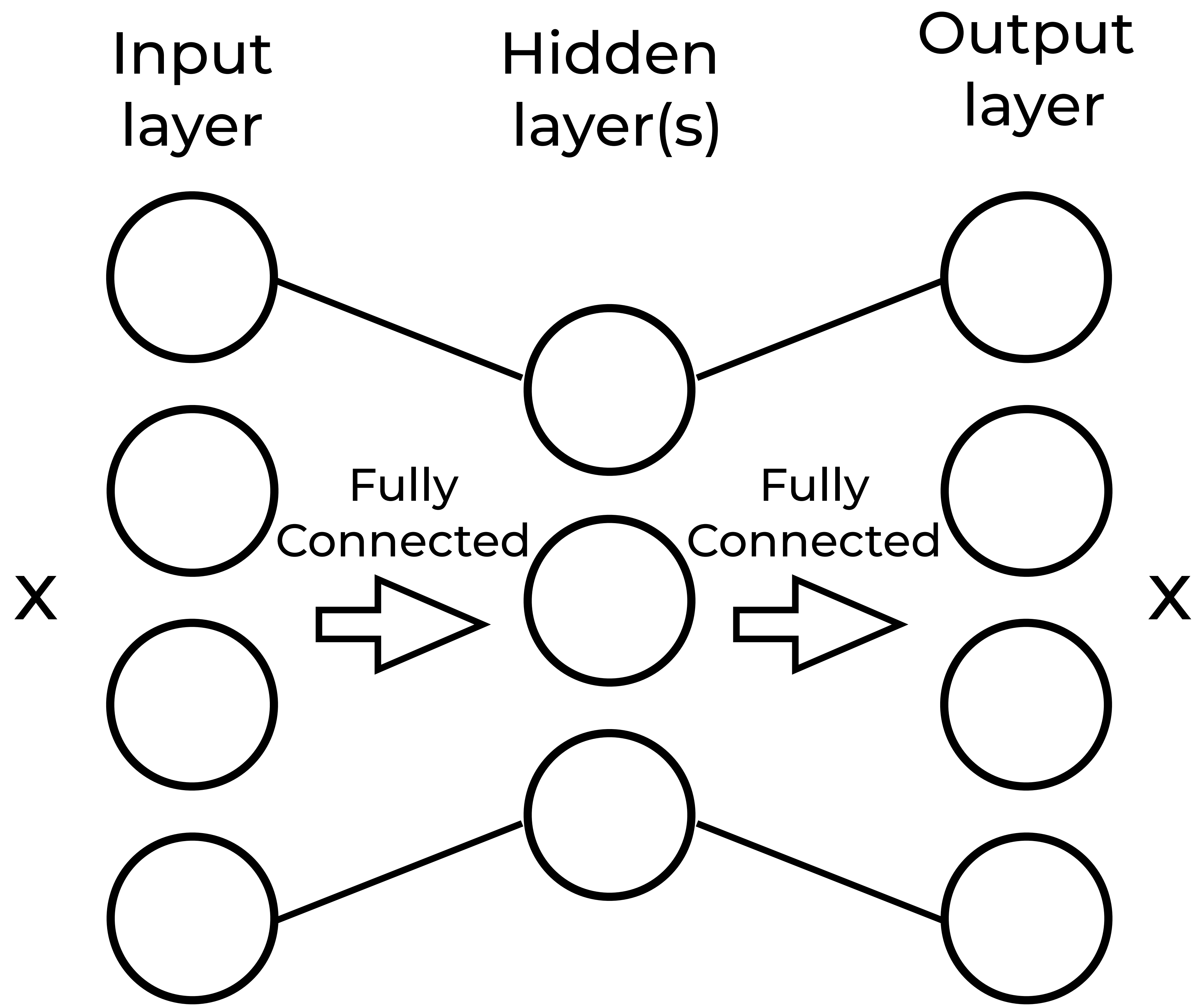
Autoencoders

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



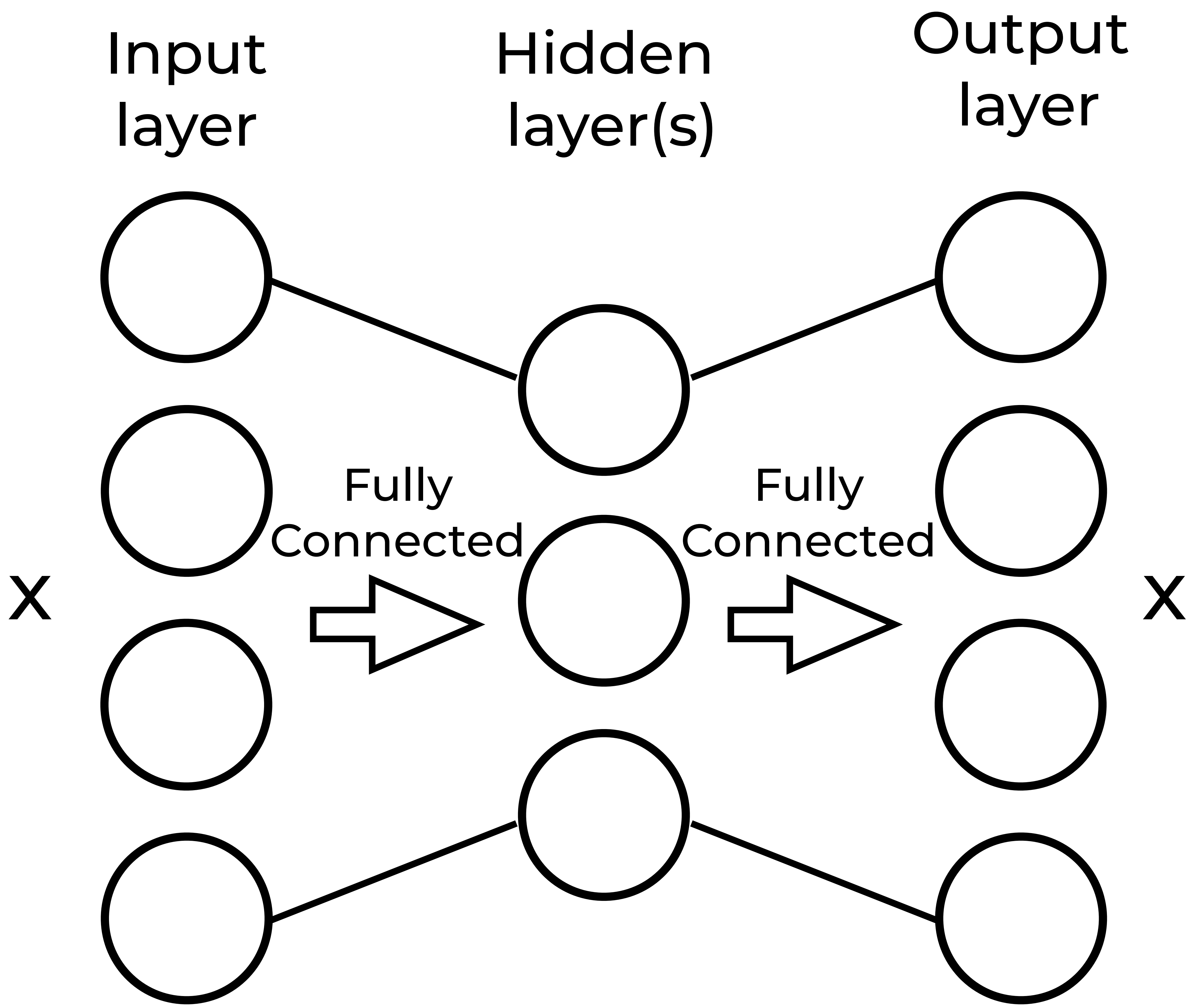
Autoencoders

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



Autoencoders

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

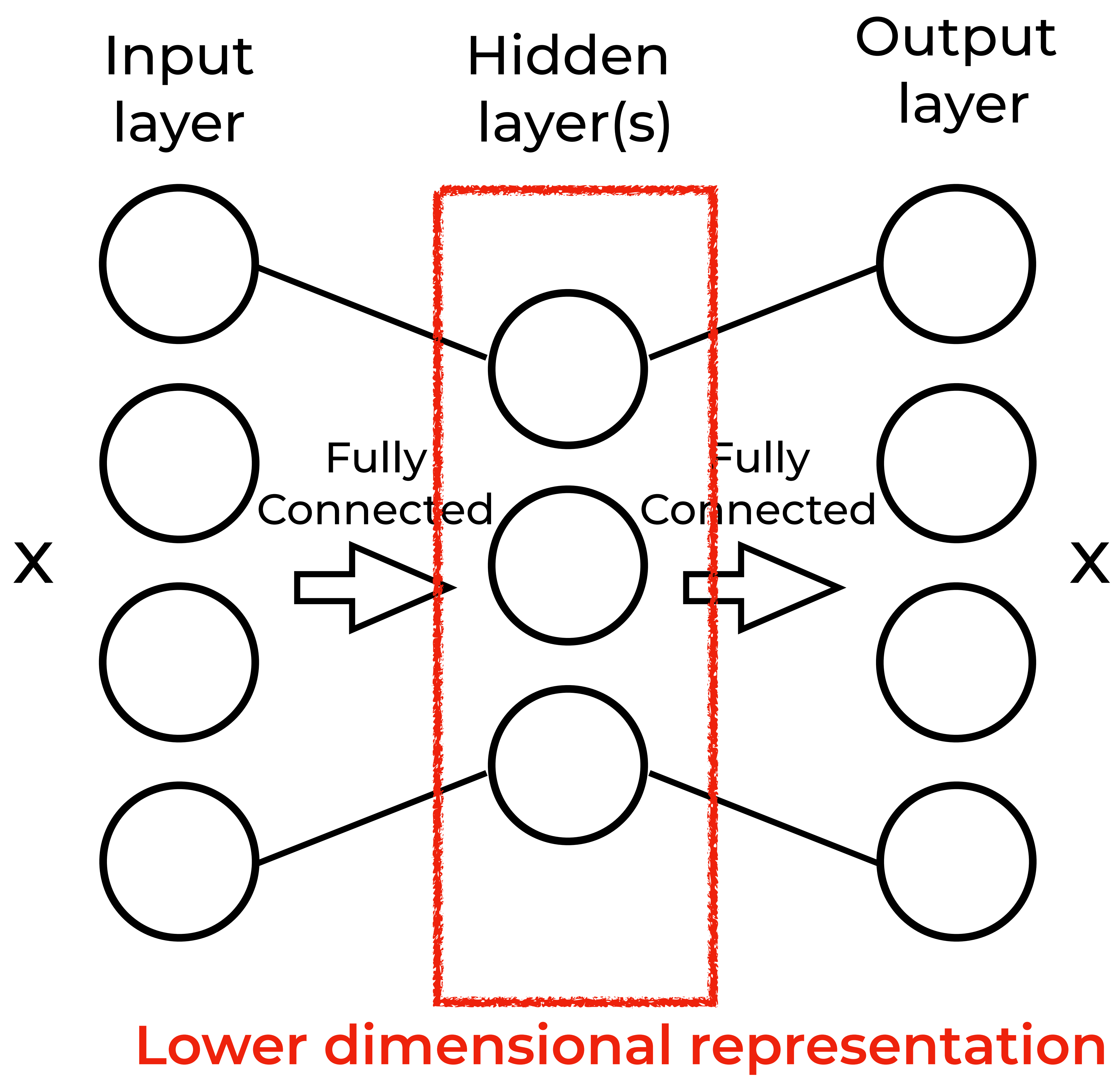


Objective:
How well the network predicts X ?

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

Autoencoders

- A neural network architecture for unsupervised learning



Objective:
How well the network predicts X ?

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