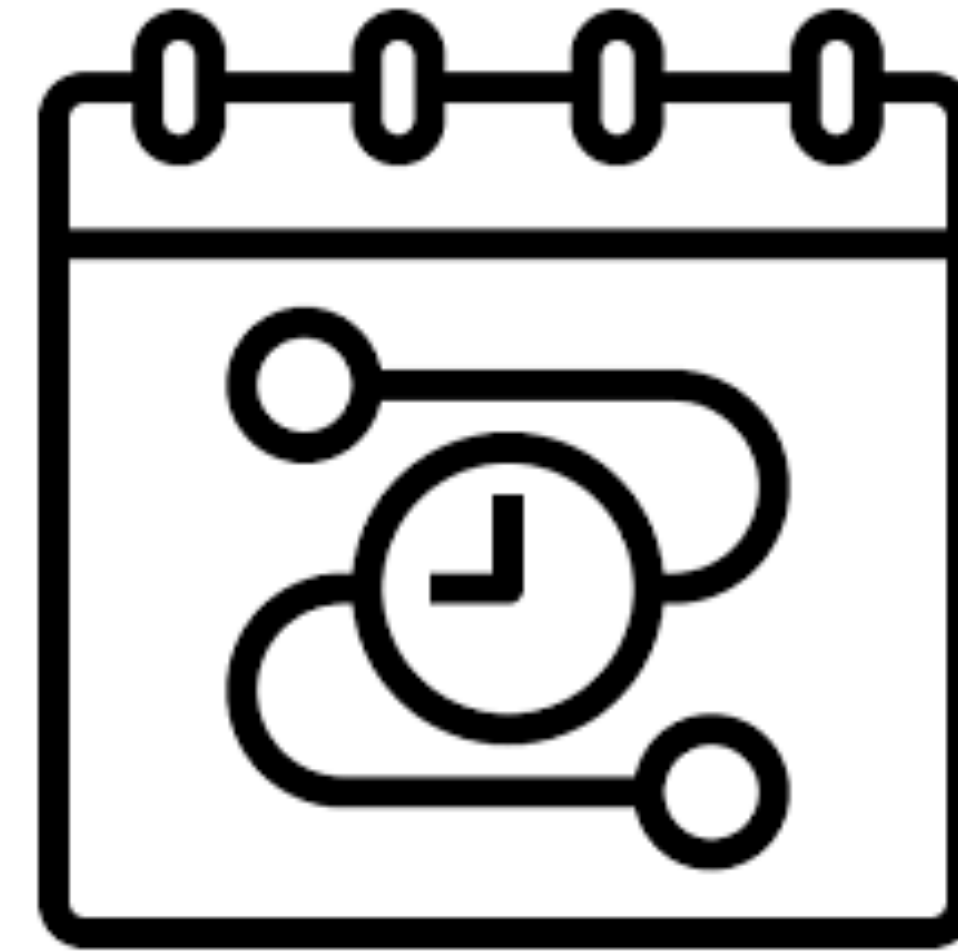


Machine Learning for Large-Scale Data Analysis and Decision Making (MATH80629A) Winter 2023

Week #7 - Summary

Announcement

- **Study Plan:** due **March 1, 2023**
- **Project Meeting:** **March 3, 2023**
- **Homework 1:** due **tonight**
- **Homework 2:** due **March 10, 2023**
(will be released tonight*)



Today

- Summary of Unsupervised Learning
- Q&A
- **Fifth Quiz** on Gradescope!
- Hands-on session (RNNs & CNNs)
- Hands-on session (Clustering)

Experience (E)

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

Unsupervised

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

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

Different tasks

- Finding patterns

- Clustering

$$f : X \rightarrow \{1, 2, \dots, K\} \text{ (K clusters)}$$

- Dimensionality reduction

$$f : X^p \rightarrow X^k, k \ll p$$

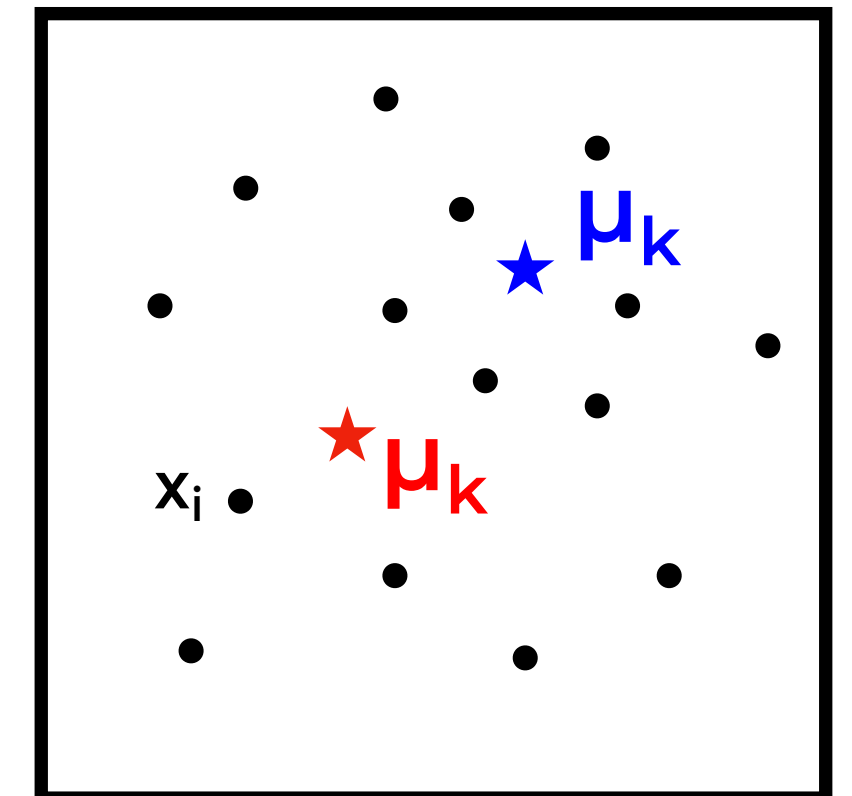
- Density modelling

$$f : X \rightarrow [0, 1]$$

- ...

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

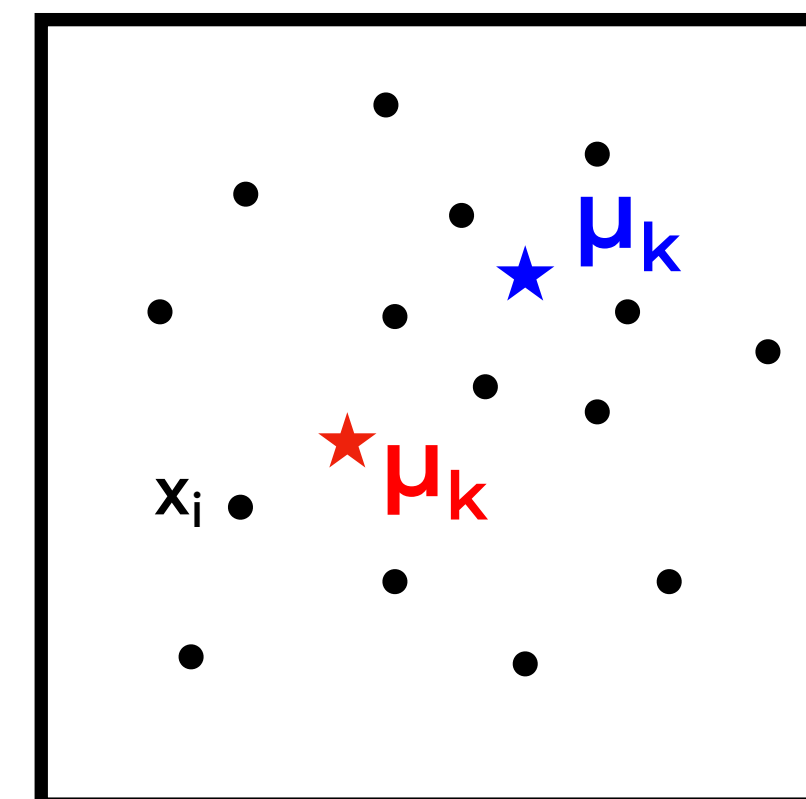


K-means clustering

- 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

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

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



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

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

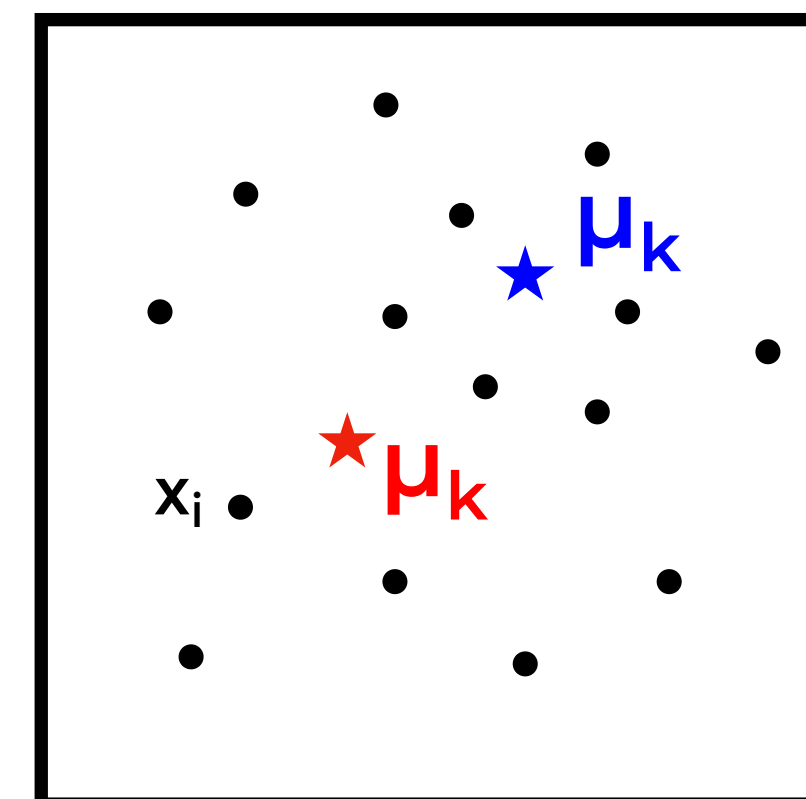
- Algorithm to minimize the objective:

- Initialize the cluster centers
- Until convergence:

1. Update responsibilities: r

2. Update cluster centers: $\mu_k \forall k$

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



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

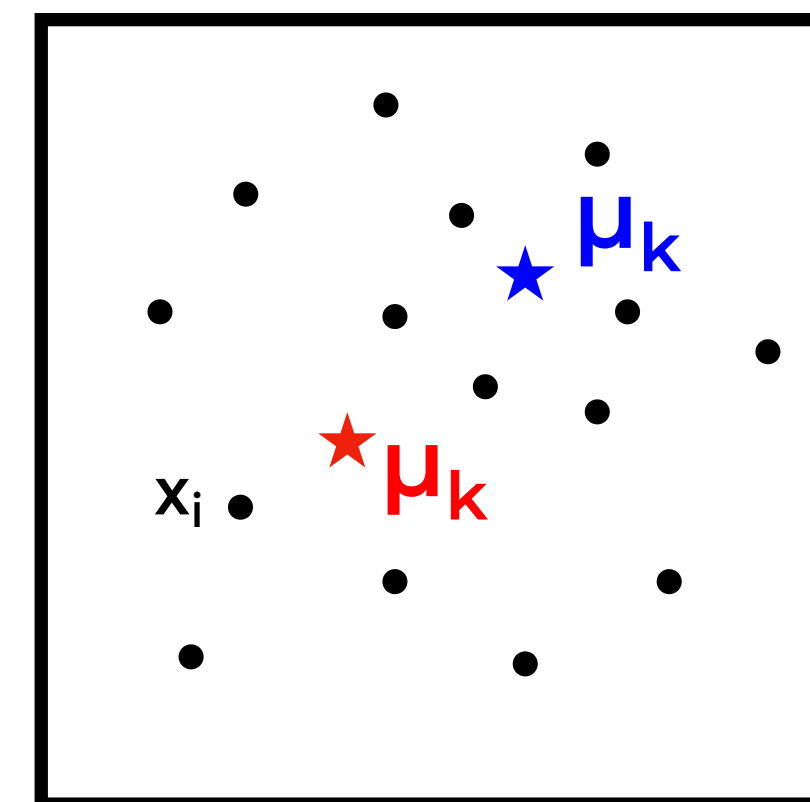
$$\text{Objective} := \sum_{i=1}^N \sum_{k=1}^K r_{ik} \|x_i - \mu_k\|^2$$

- 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

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

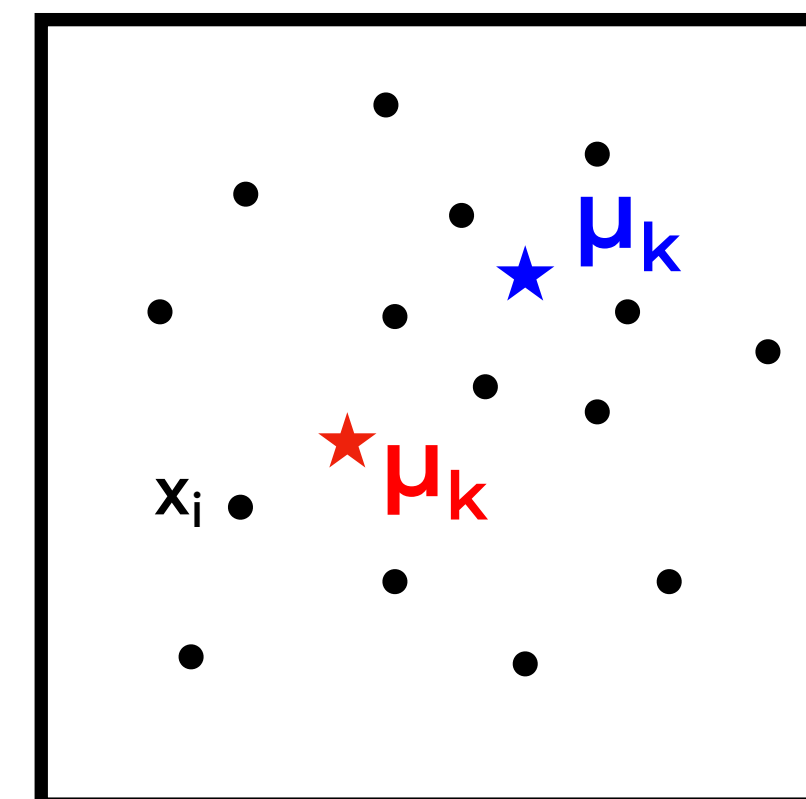
- Algorithm to minimize the objective:

- 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

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

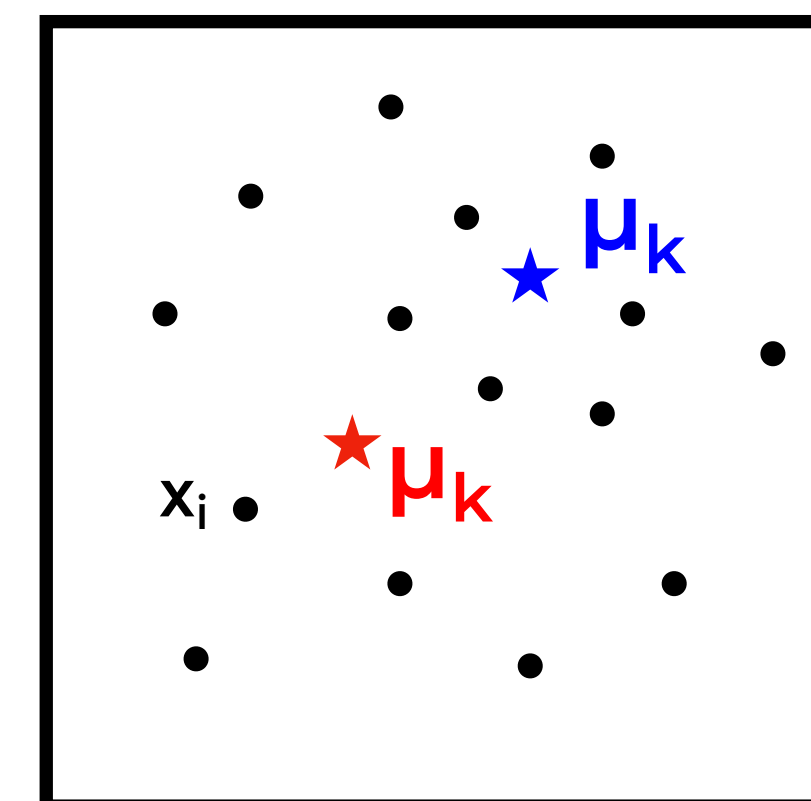
- Algorithm to minimize the objective:

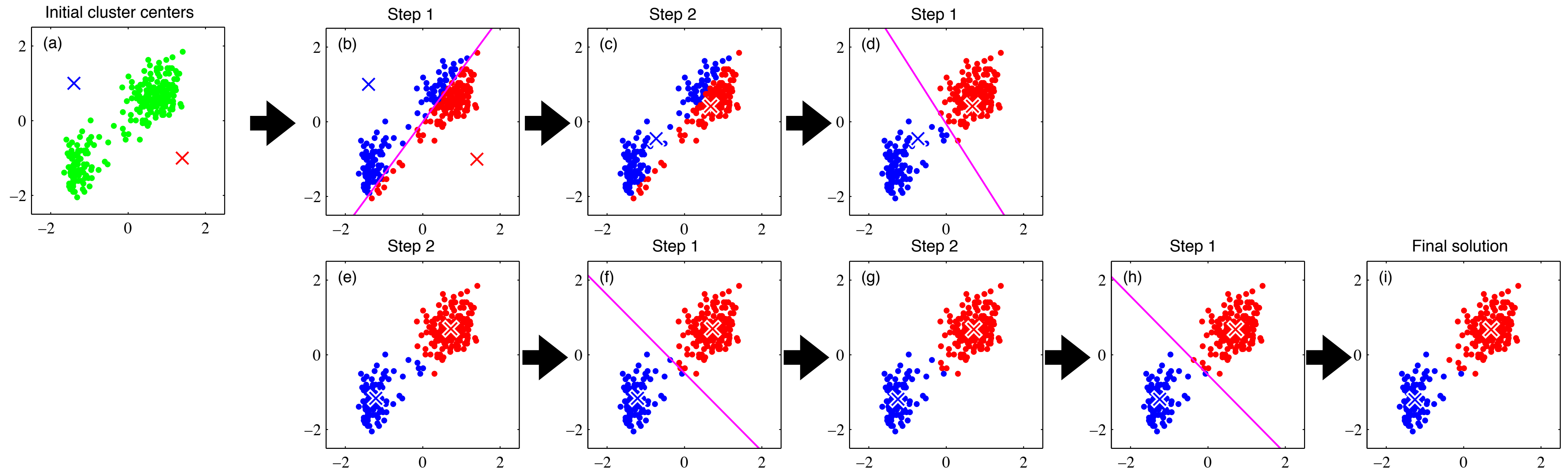
- Initialize the cluster centers
- Until convergence:

1. Update responsibilities: r

2. Update cluster centers: $\mu_k \forall k$

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





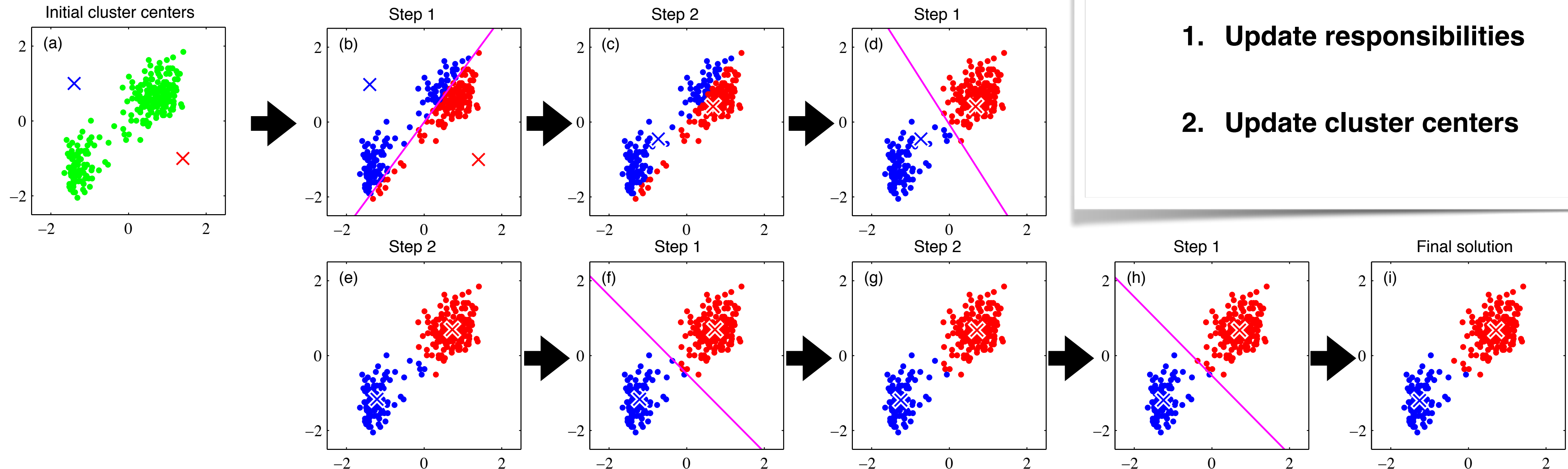
Algorithm

- Initialize the cluster centers

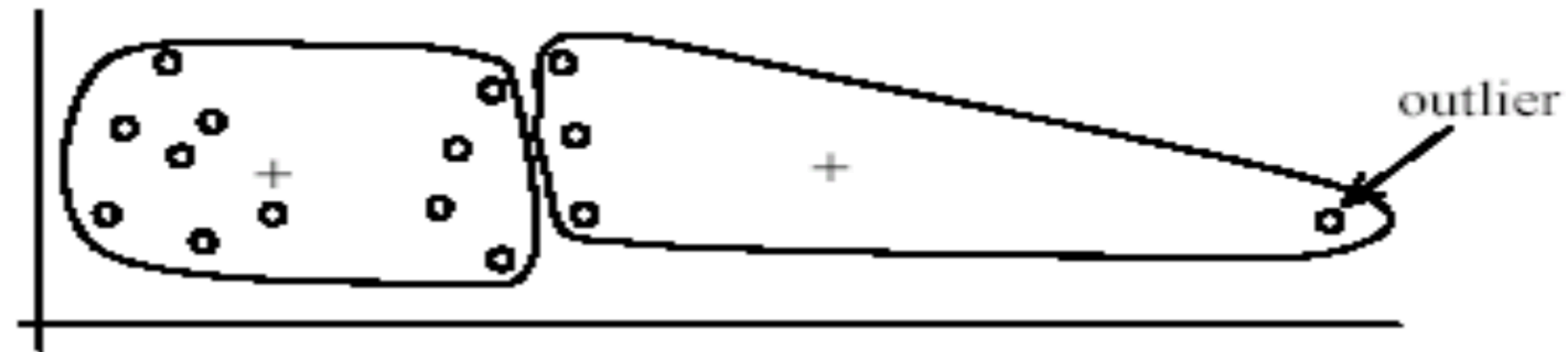
- Until convergence:

1. Update responsibilities

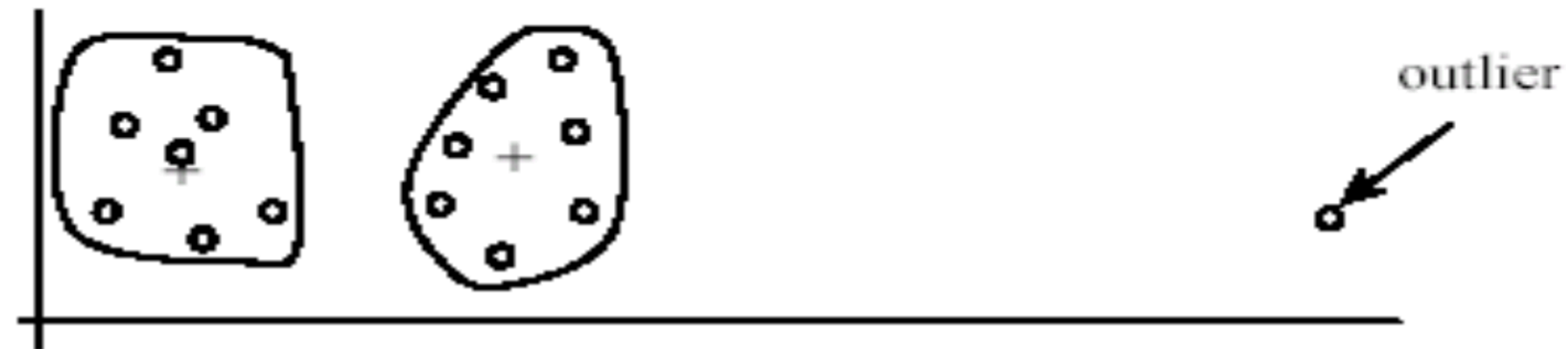
2. Update cluster centers



K-means and sensitivity to outliers

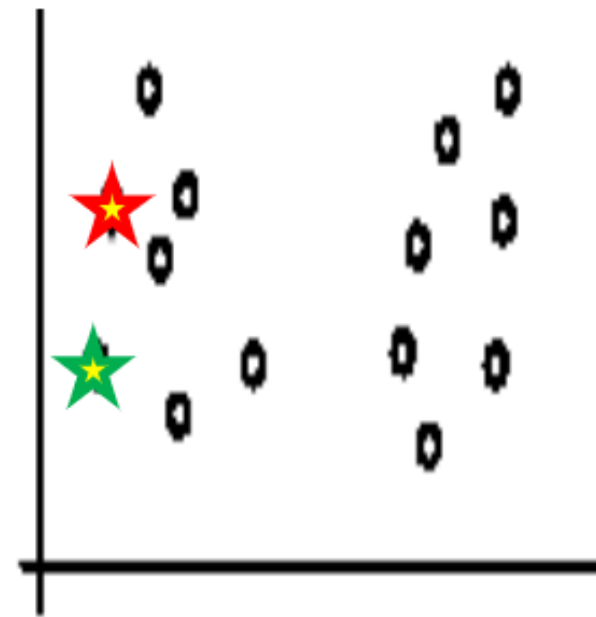


(A): Undesirable clusters

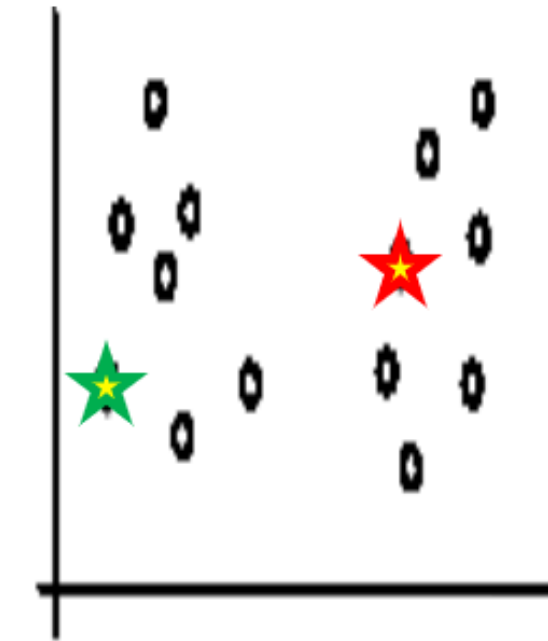


(B): Ideal clusters

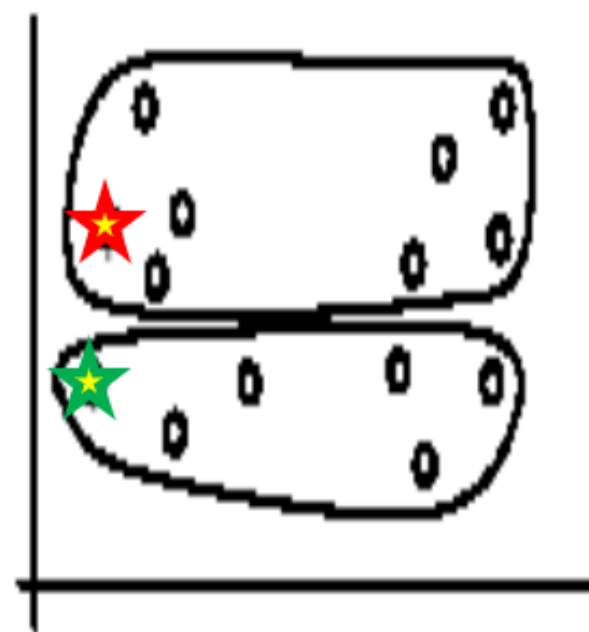
K-means and sensitivity to initial seeds



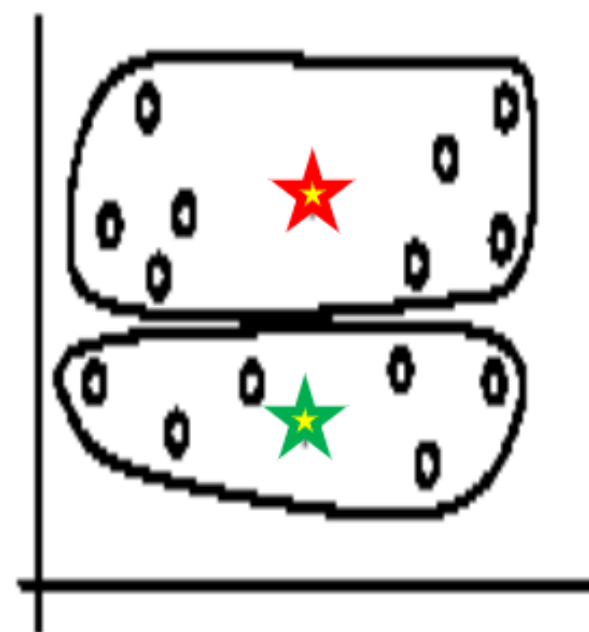
Random selection of seeds (centroids)



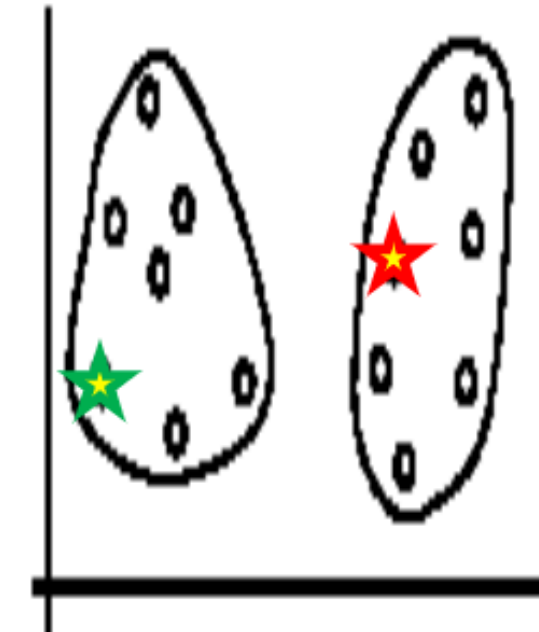
Random selection of seeds (centroids)



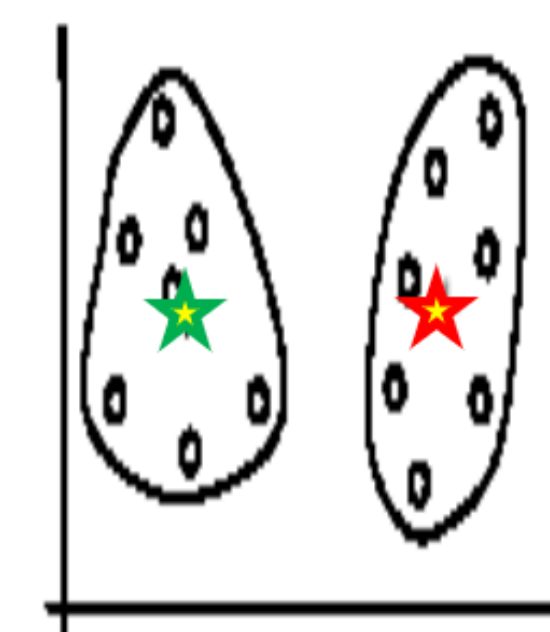
Iteration 1



Iteration 2



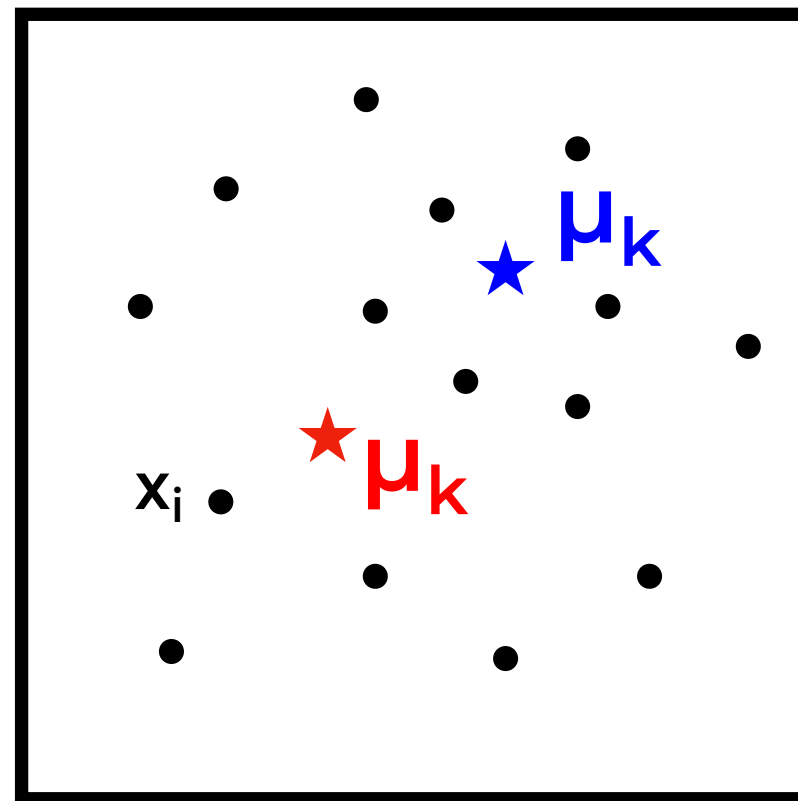
Iteration 1



Iteration 2

A probabilistic approach to k-means clustering

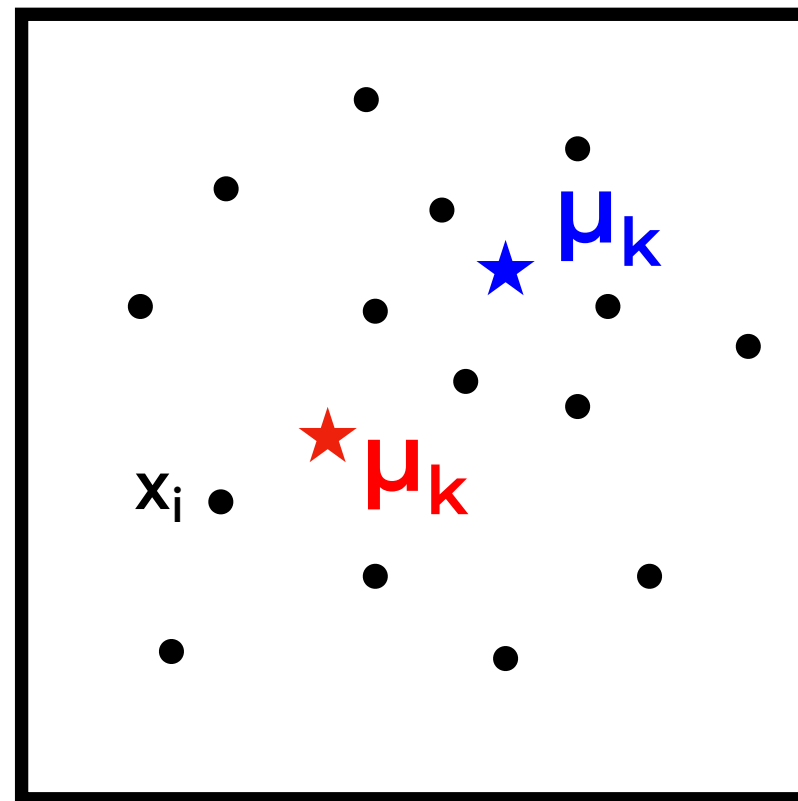
K-means Clustering



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

A probabilistic approach to k-means clustering

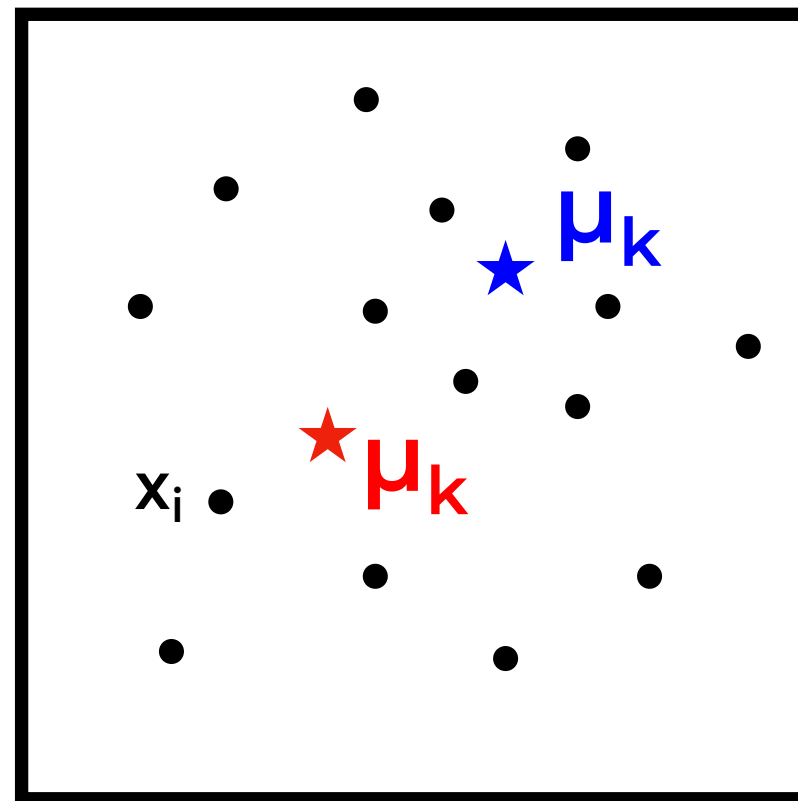
K-means Clustering



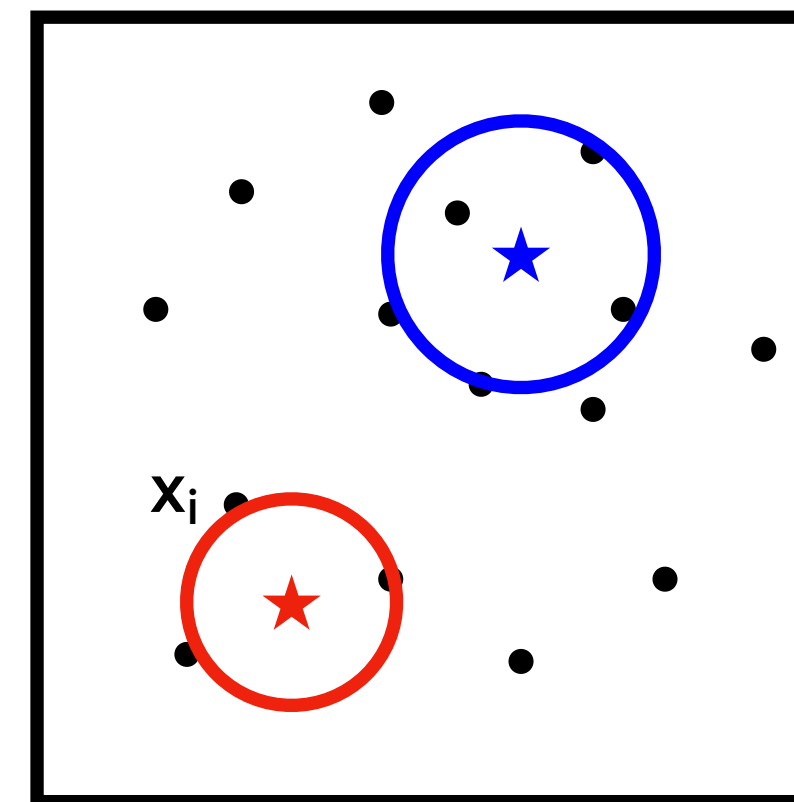
$$\text{Objective} := \sum_{i=1}^N \sum_{k=1}^K \overset{\text{Responsibility}}{\boxed{r_{ik}}} \left\| \mathbf{x}_i - \overset{\text{center}}{\boxed{\mu_k}} \right\|^2$$

A probabilistic approach to k-means clustering

K-means Clustering



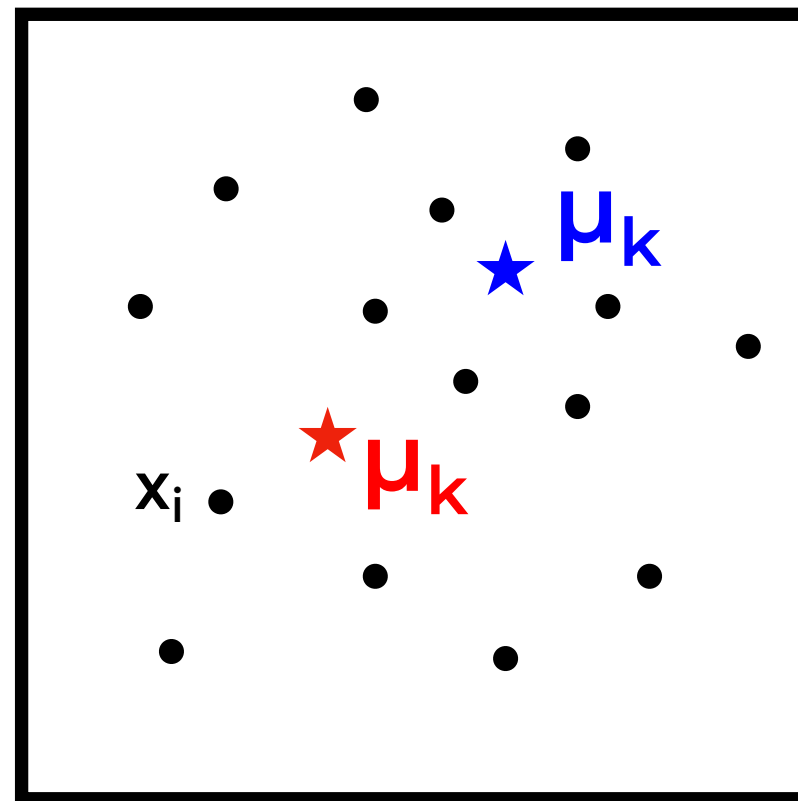
Soft K-means Clustering



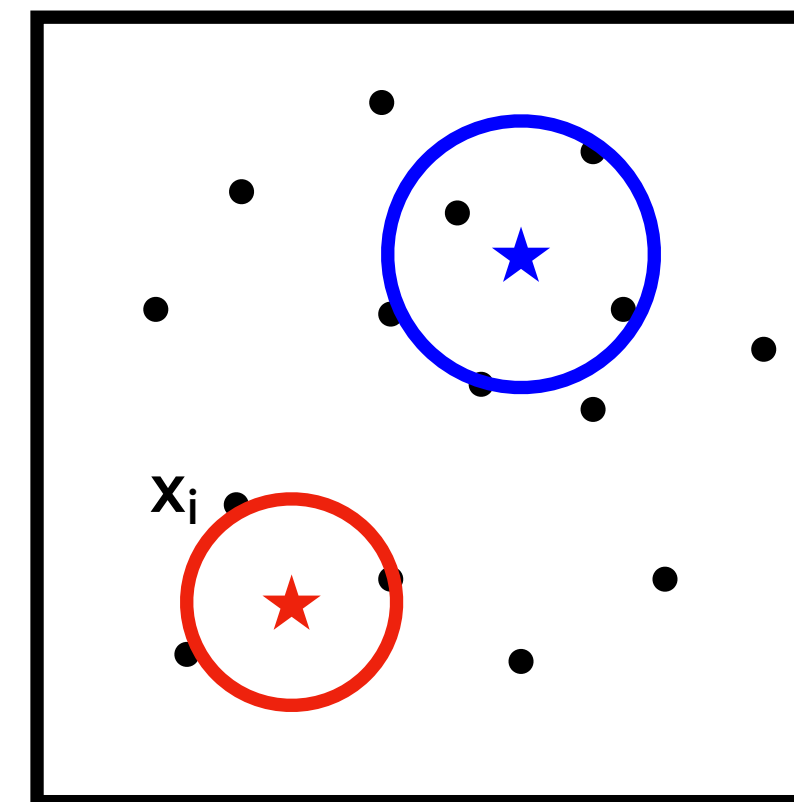
$$\text{Objective} := \sum_{i=1}^N \sum_{k=1}^K \overset{\text{Responsibility}}{\boxed{r_{ik}}} \left\| \mathbf{x}_i - \overset{\text{center}}{\boxed{\mu_k}} \right\|^2$$

A probabilistic approach to k-means clustering

K-means Clustering



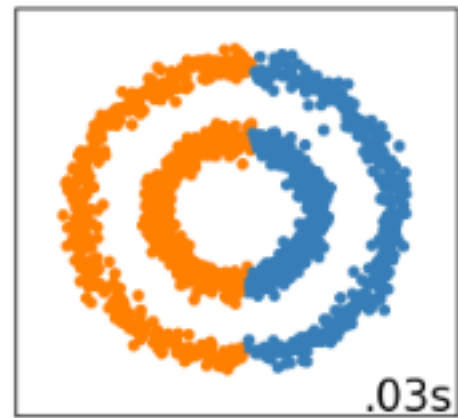
Soft K-means Clustering



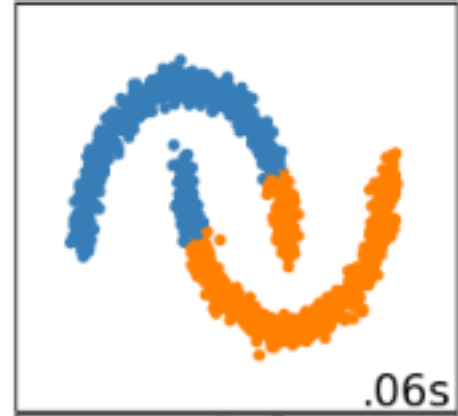
$$\text{Objective} := \sum_{i=1}^N \sum_{k=1}^K \overset{\text{Responsibility}}{\boxed{r_{ik}}} \| \mathbf{x}_i - \overset{\text{center}}{\boxed{\mu_k}} \|^2$$

- Responsibilities are continuous $[0, 1]$
 - Each cluster has a responsibility: π_k
- Each cluster models data using a Gaussian: $\mathcal{N}(\mathbf{x}_i | \mu_k, \Sigma_k)$

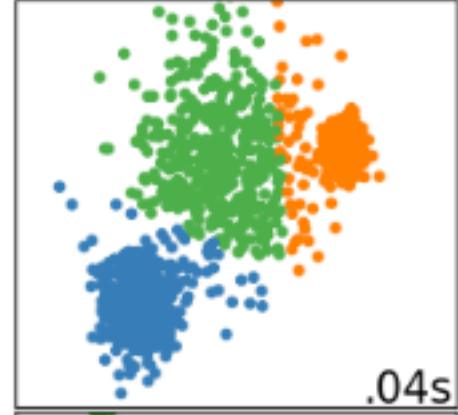
K-means



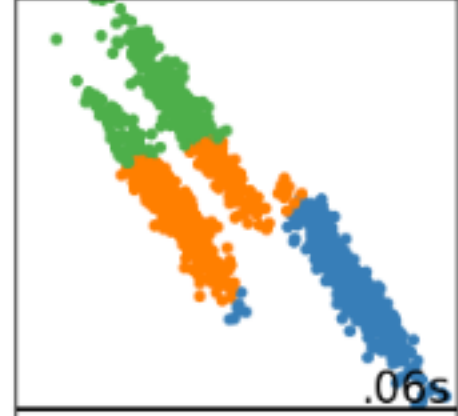
Similar



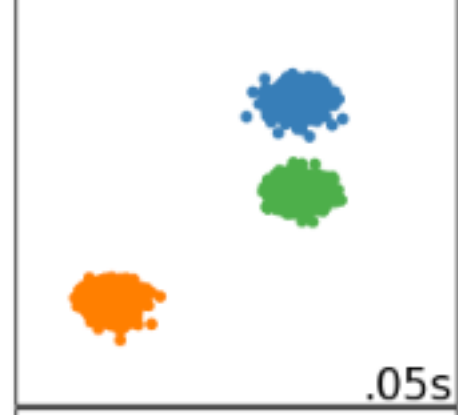
Similar



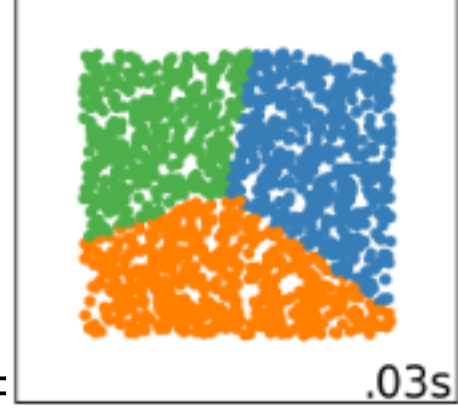
GMM better



GMM better

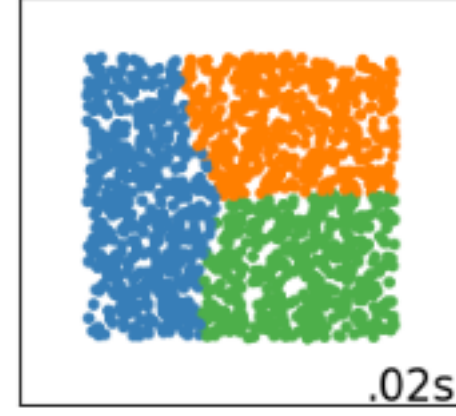
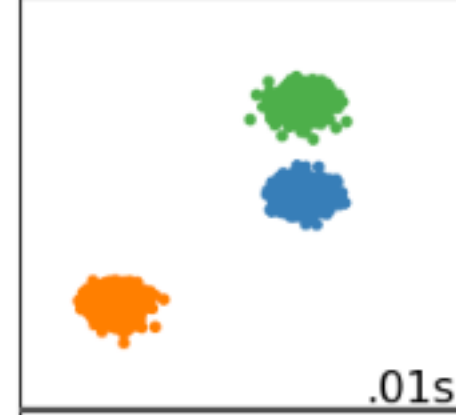
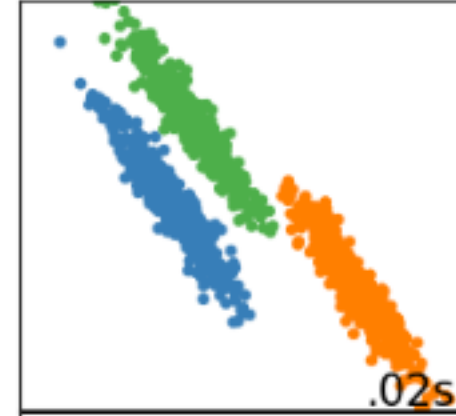
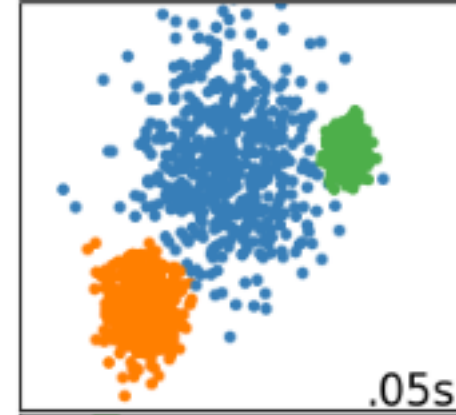
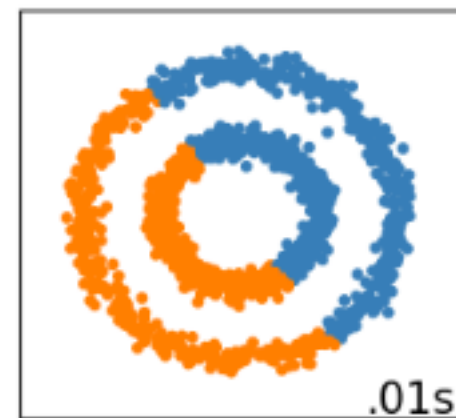


Similar

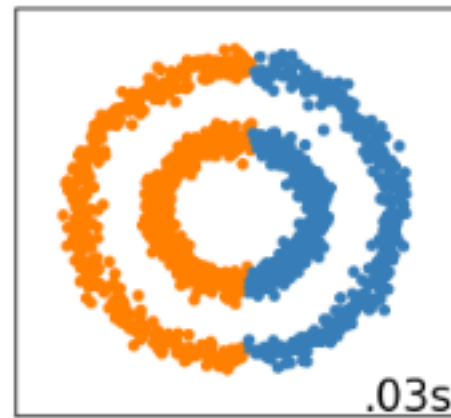


Similar

GMMs



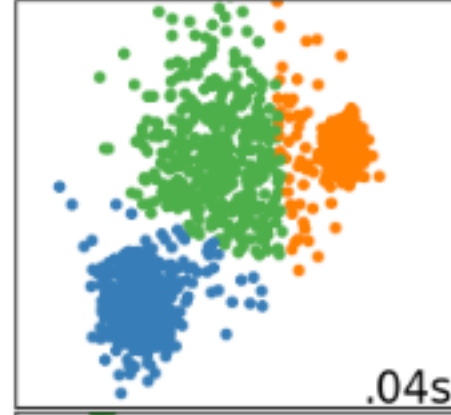
K-means



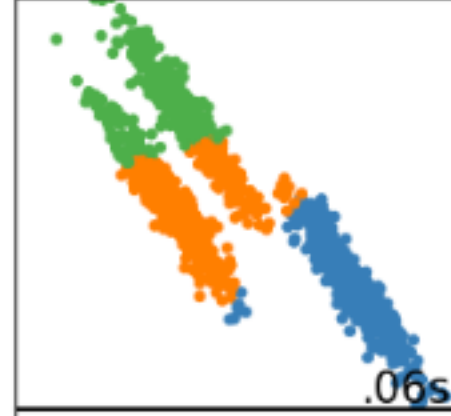
Similar



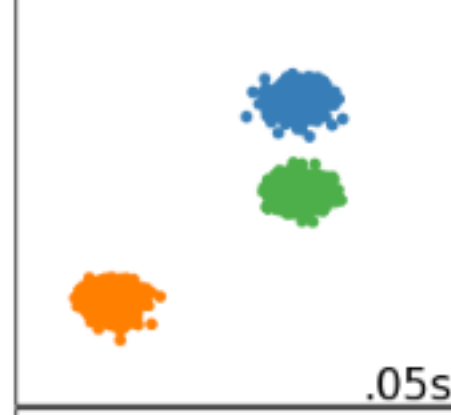
Similar



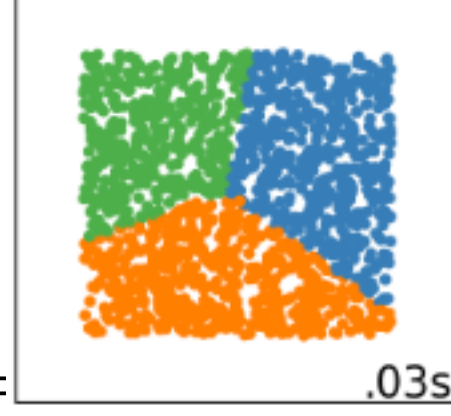
GMM better



GMM better

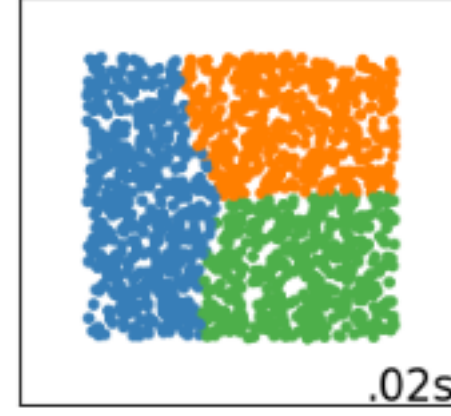
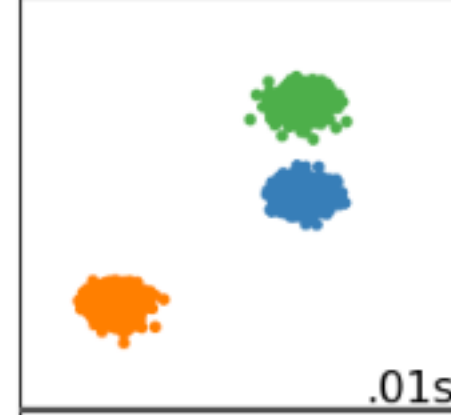
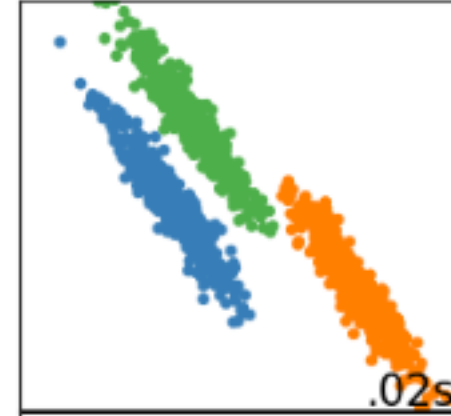
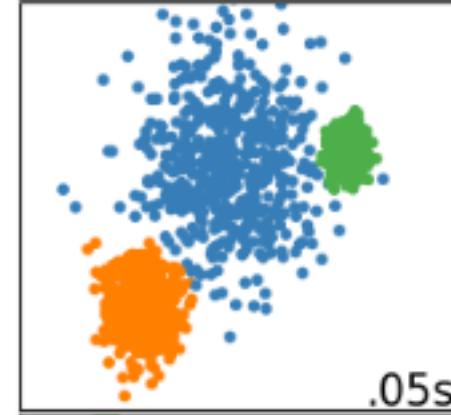
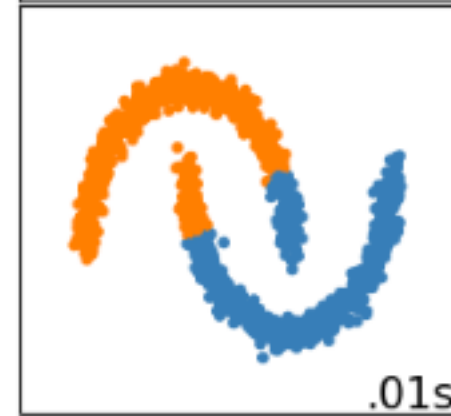
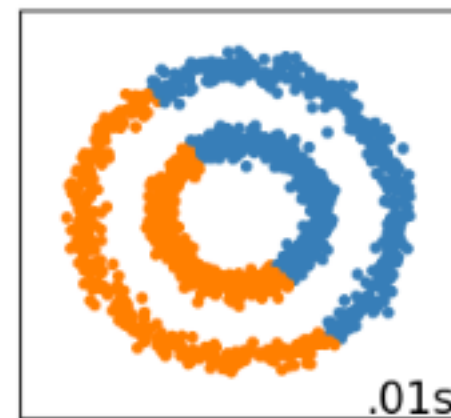


Similar



Similar

GMMs

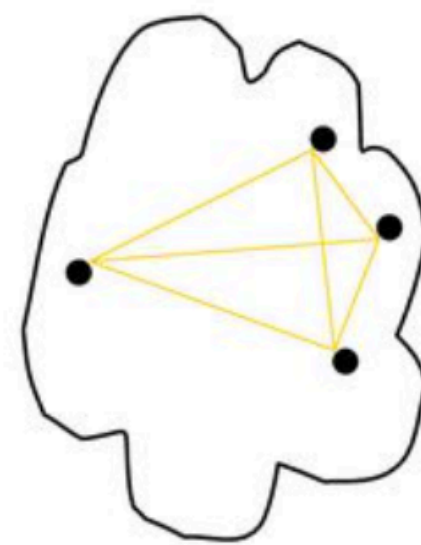


Comparing K-means to GMMs

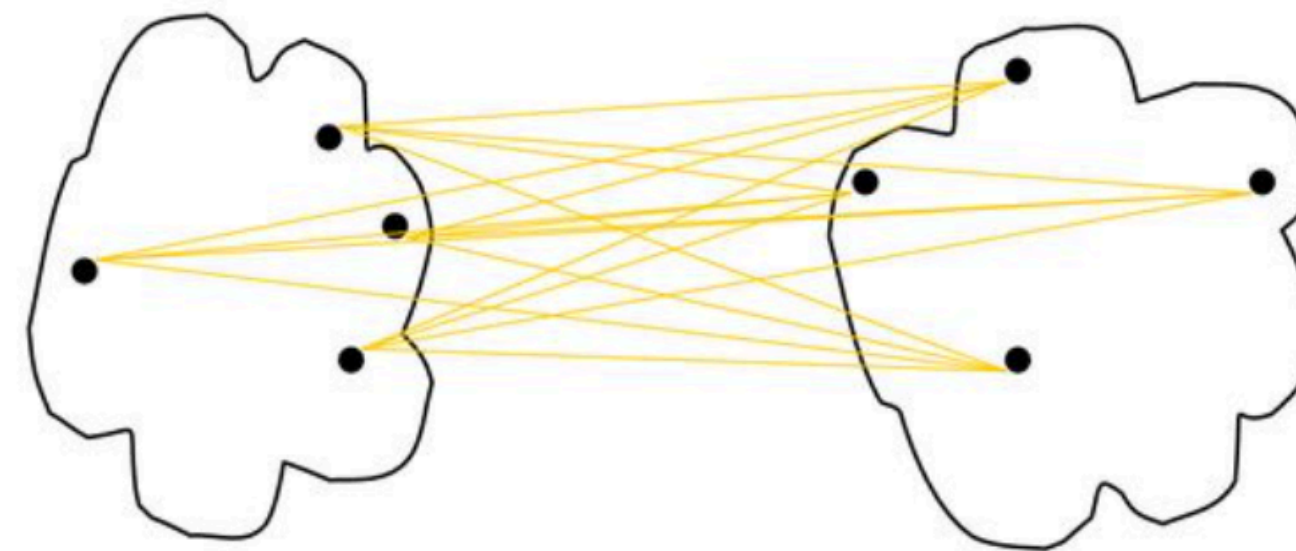
- GMMs learns covariance matrix
- Per cluster variance
- Covariance terms
- GMMs has many more parameters
- Covariance matrix ($M \times M$)

Additional info

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



cohesion



separation

Additional info

- **Cluster Cohesion:** how closely are the objects within the cluster

WSS (Within Clusters Sum of Squares)

We want this to be small

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

- **Cluster Separation:** measure how distinct clusters are wrt each other

BSS (Between Clusters Sum of Squares)

We want this to be large

$$BSS = \sum_i m_i (C_i - C)^2$$

the number of observations in the j-th cluster

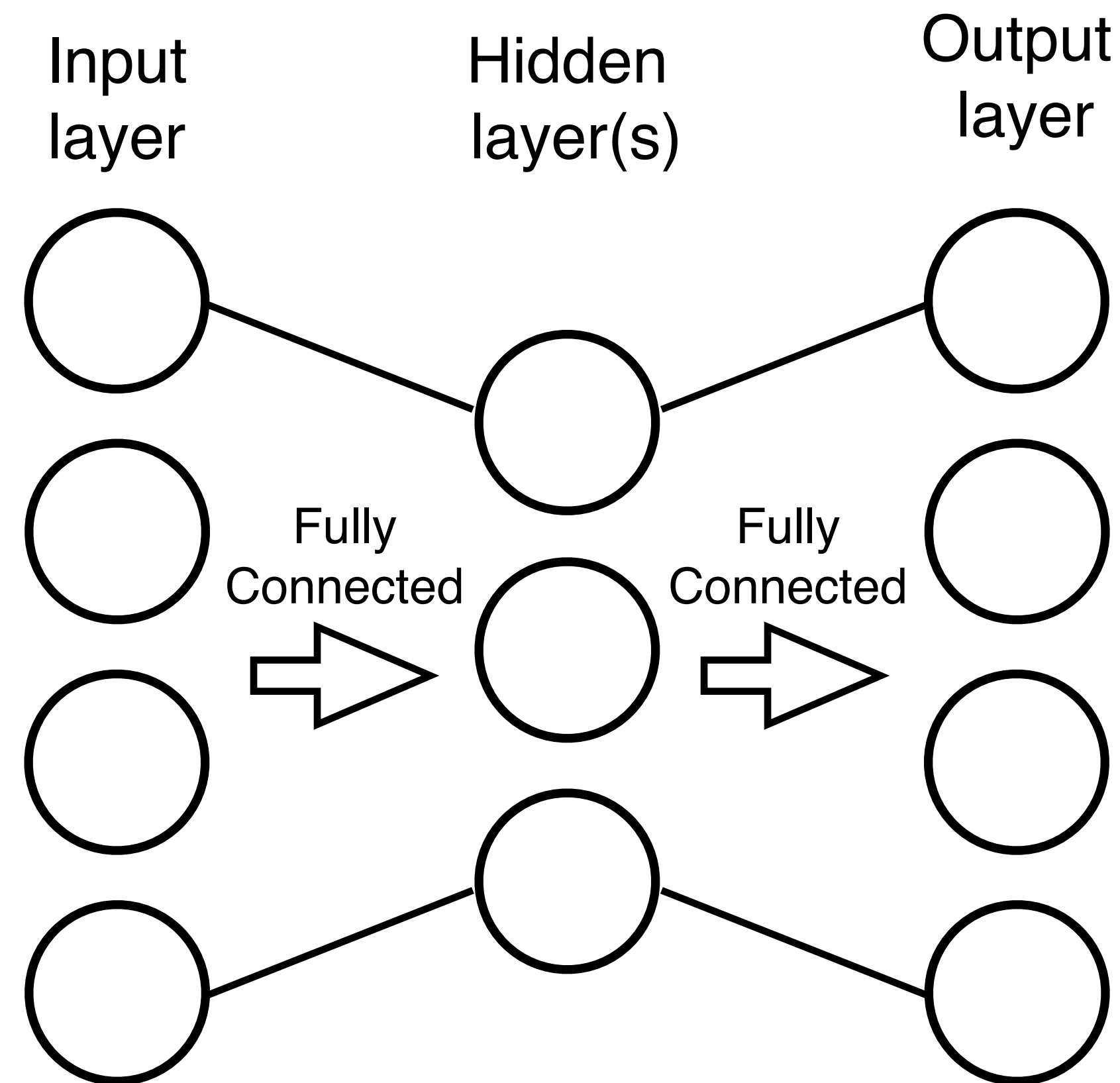
$$BSS + WSS = \text{CONSTANT}$$

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**

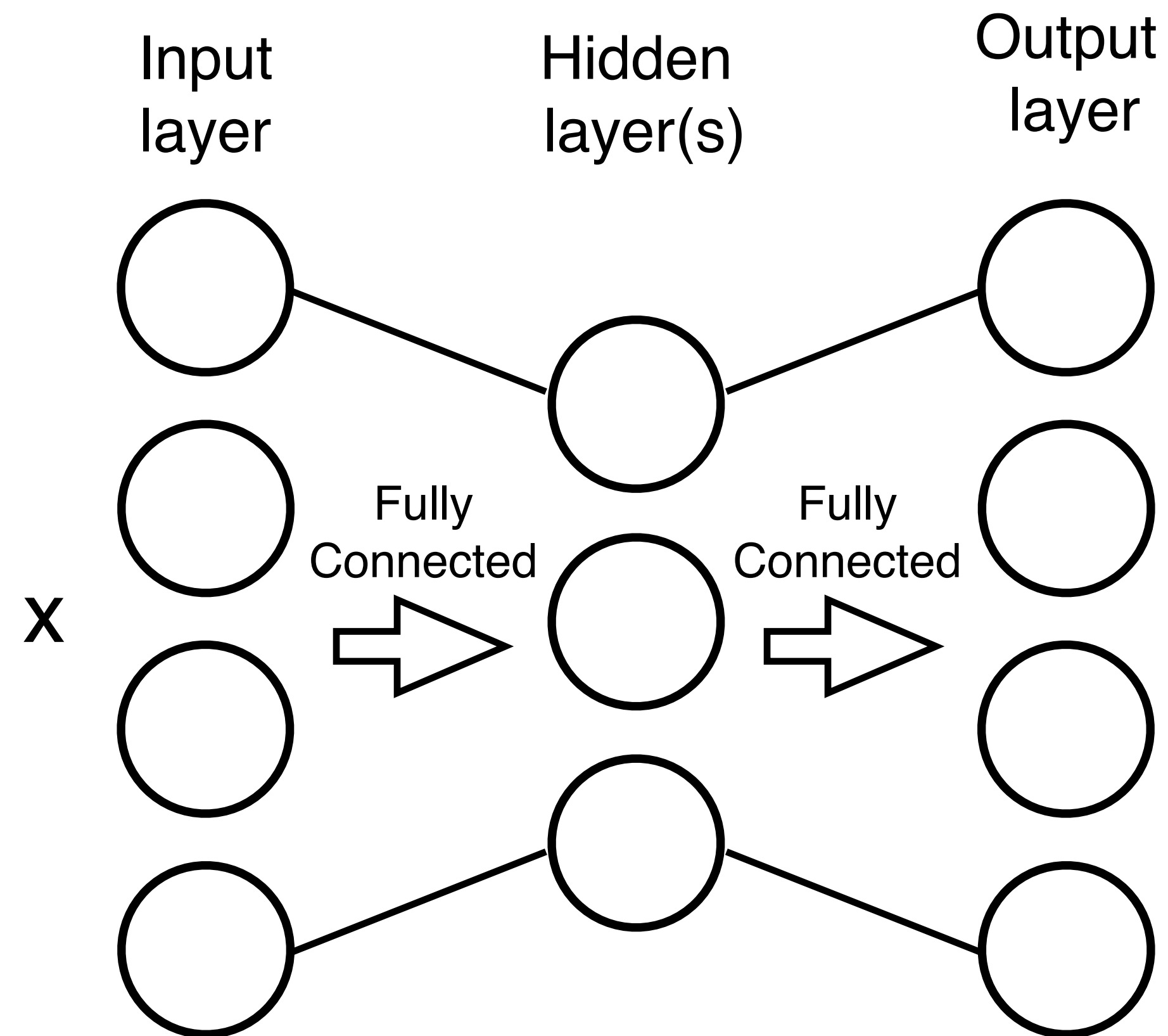
Autoencoders

- A neural network architecture for unsupervised learning



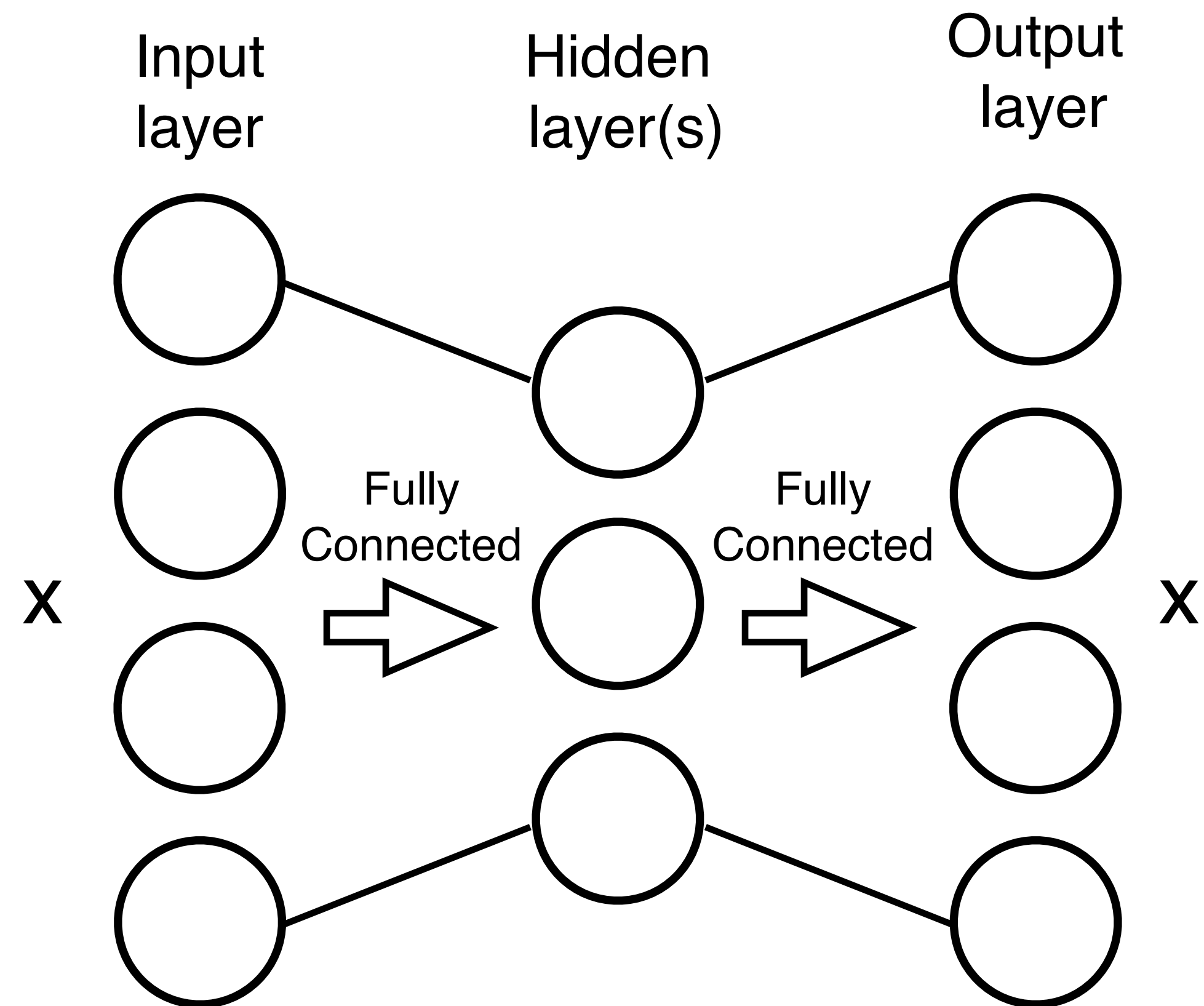
Autoencoders

- A neural network architecture for unsupervised learning



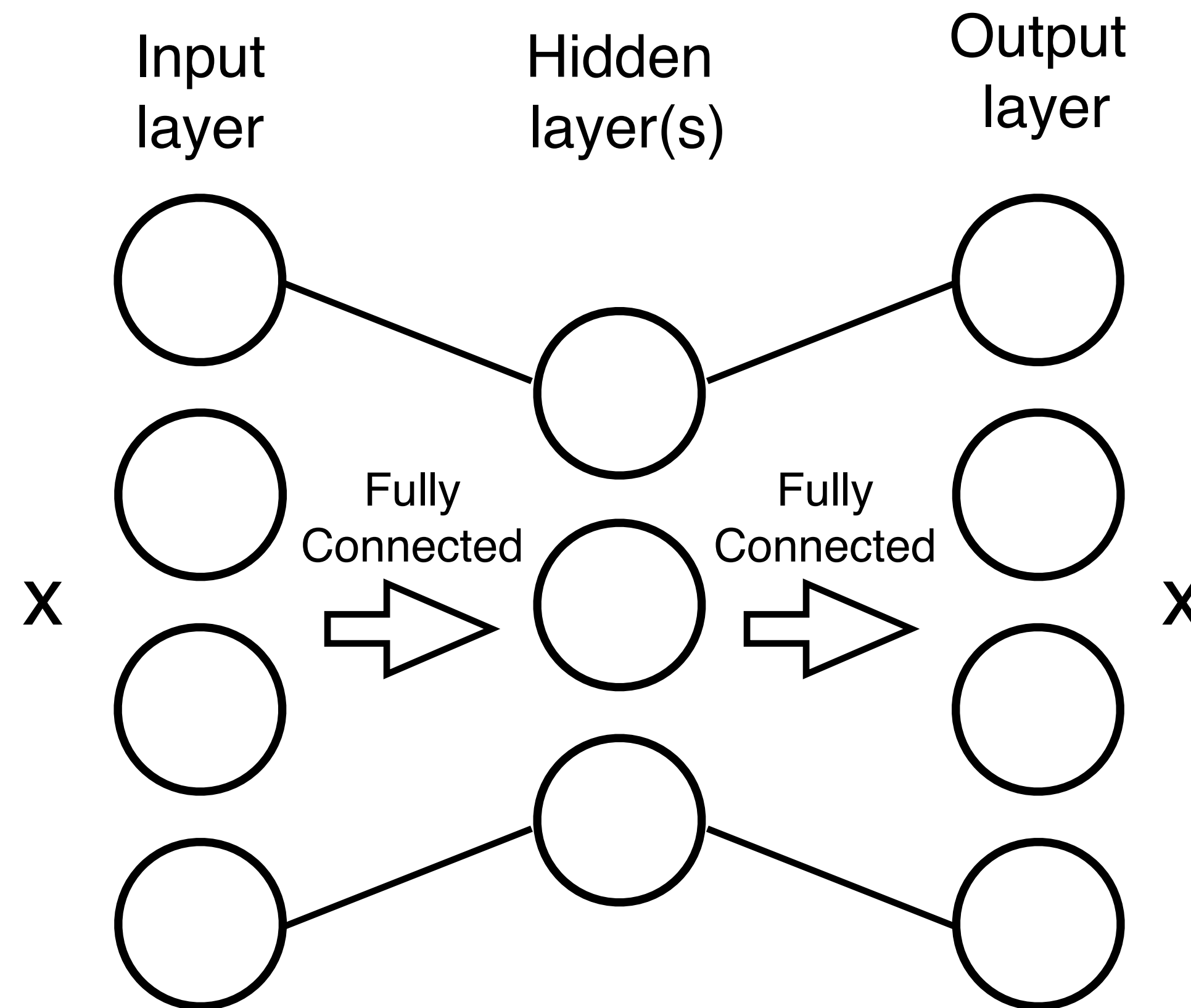
Autoencoders

- A neural network architecture for unsupervised learning



Autoencoders

- A neural network architecture for unsupervised learning



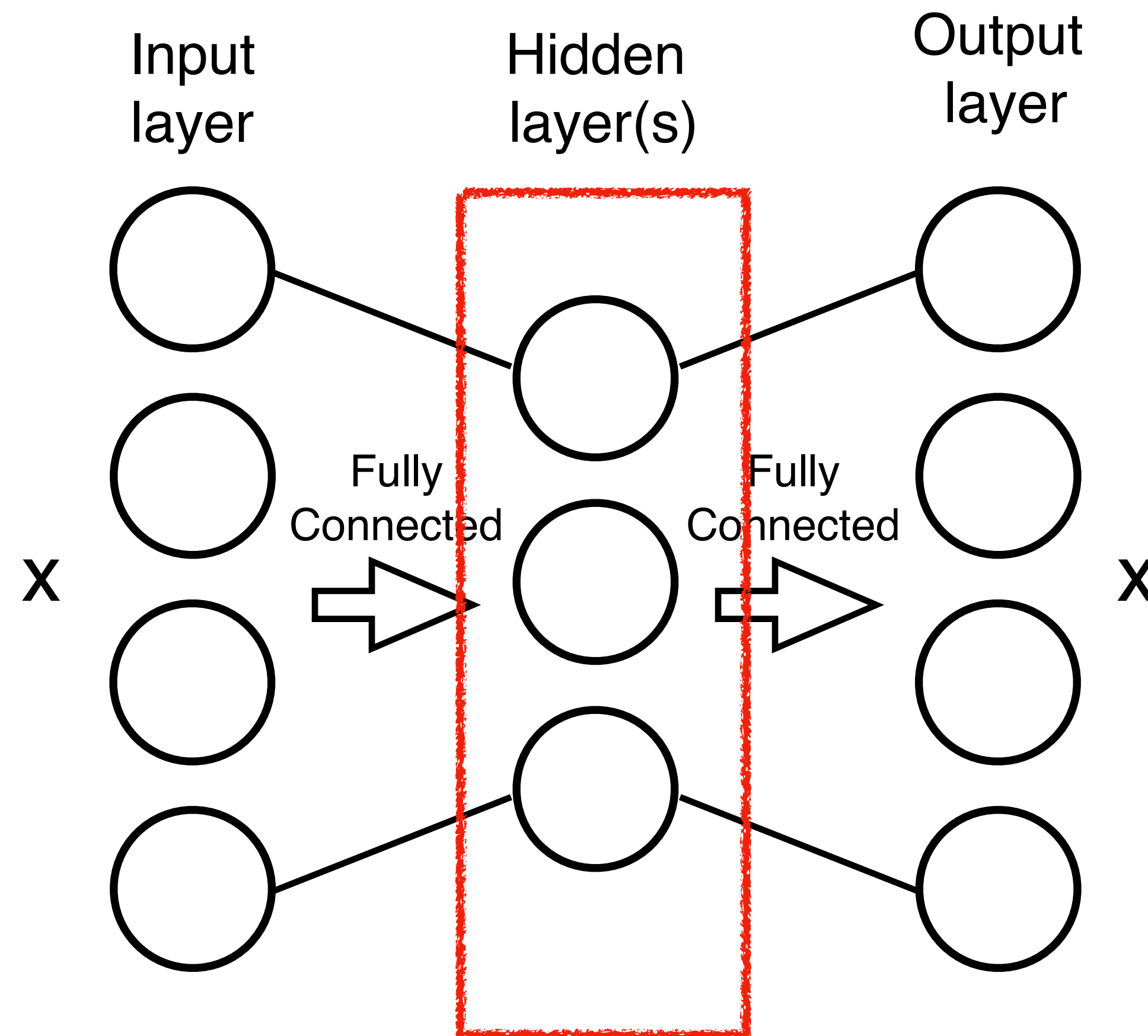
Objective:

How well the network predicts **X** ?

$$\begin{aligned} \text{Loss} &:= \sum_{i=1}^N (\mathbf{x}_i - \hat{\mathbf{x}}_i)^2 \\ &= \sum_{i=1}^N (\mathbf{x}_i - \mathbf{f}_2(\mathbf{f}_1(\mathbf{x})))^2 \end{aligned}$$

Autoencoders

- A neural network architecture for unsupervised learning



Lower dimensional representation

Objective:

How well the network predicts X ?

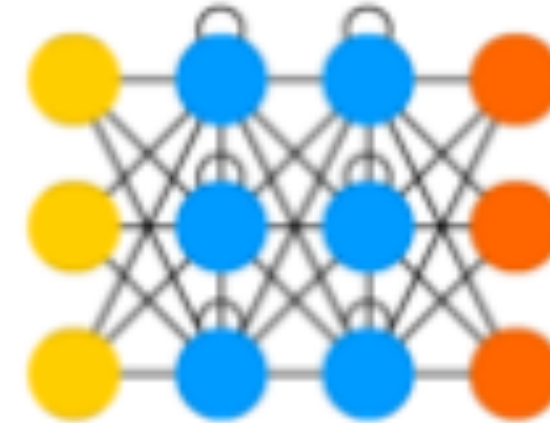
$$\begin{aligned} \text{Loss} &:= \sum_{i=1}^N (\mathbf{x}_i - \hat{\mathbf{x}}_i)^2 \\ &= \sum_{i=1}^N (\mathbf{x}_i - \mathbf{f}_2(\mathbf{f}_1(\mathbf{x})))^2 \end{aligned}$$

Various Architectures

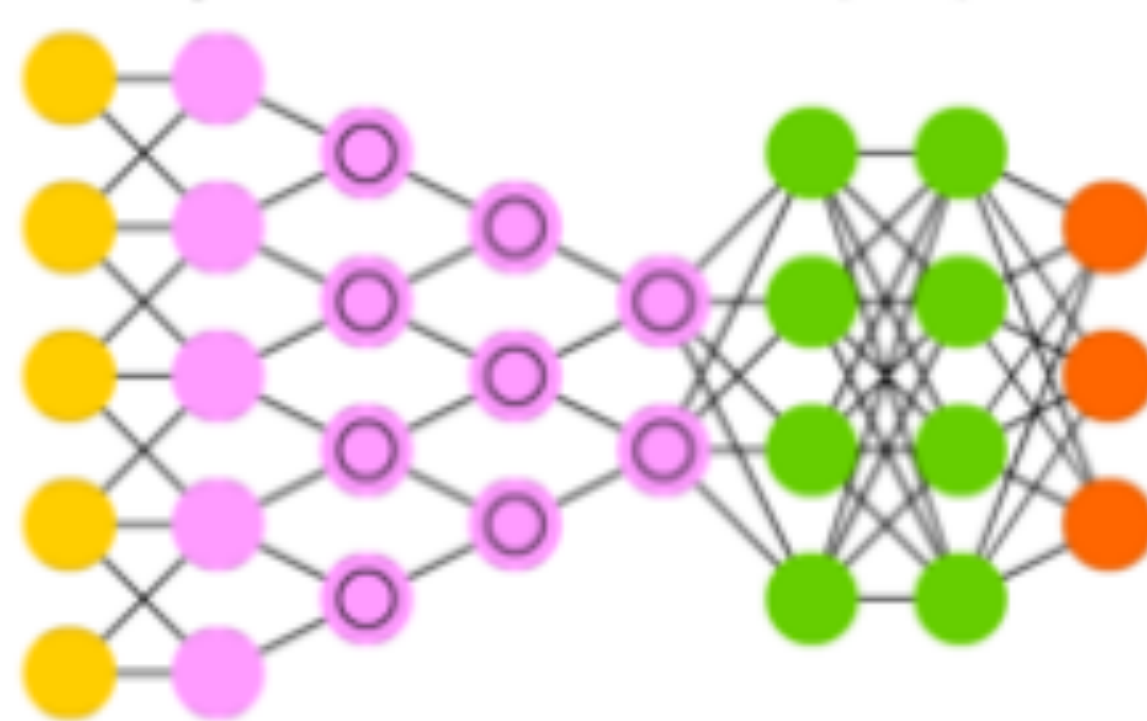
Deep Feed Forward (DFF)



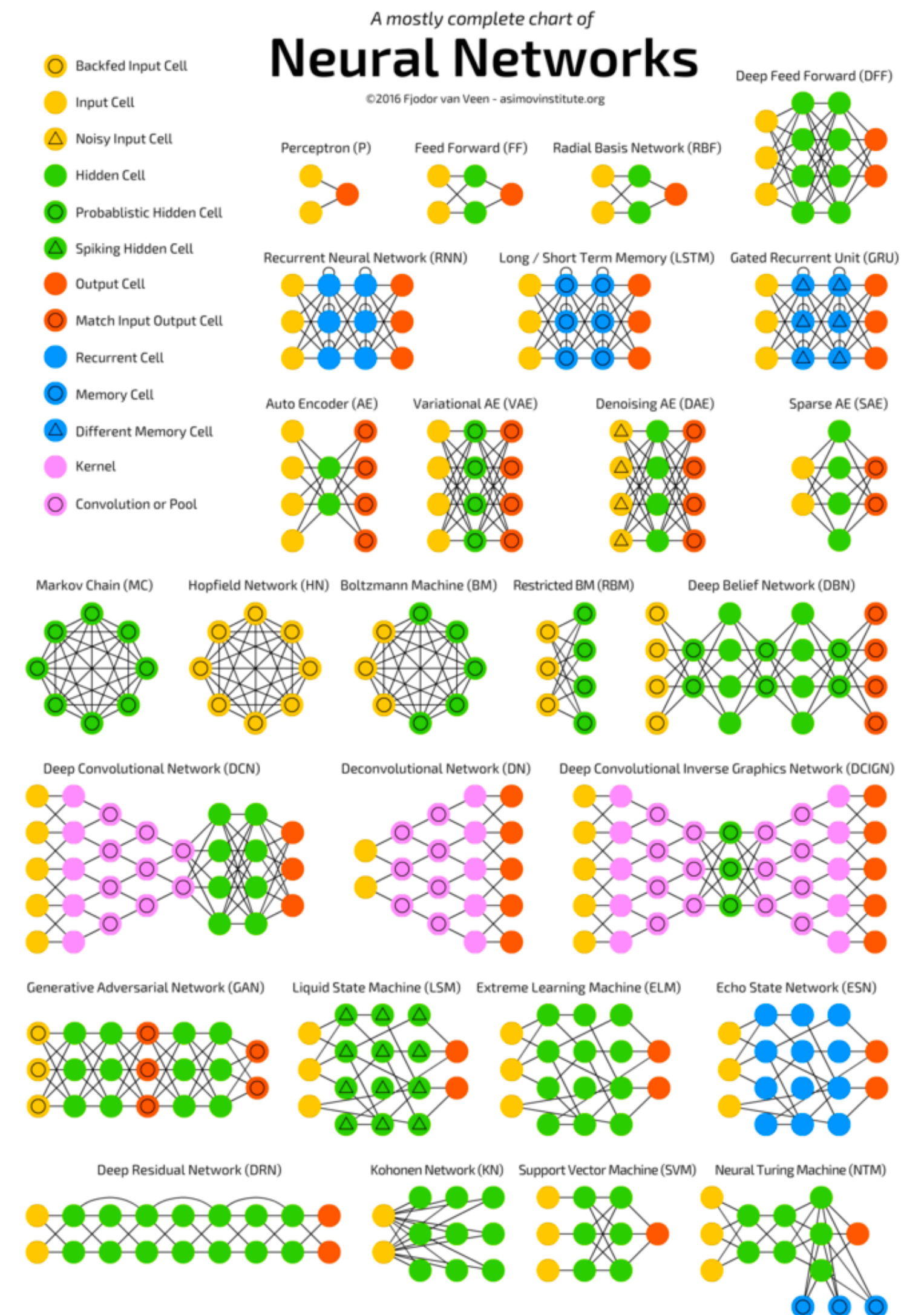
Recurrent Neural Network (RNN)



Deep Convolutional Network (DCN)



Auto Encoder (AE)



Leijnen, Stefan, and Fjodor van Veen. "The neural network zoo." *Multidisciplinary Digital Publishing Institute Proceedings*. Vol. 47. No. 1. 2020.



Quiz 4

Login to your Gradescope account