

Apprentissage Automatique I 60629

Sommaire

Apprentissage non supervisé
— Semaine #7

1. Apprentissage non supervisé

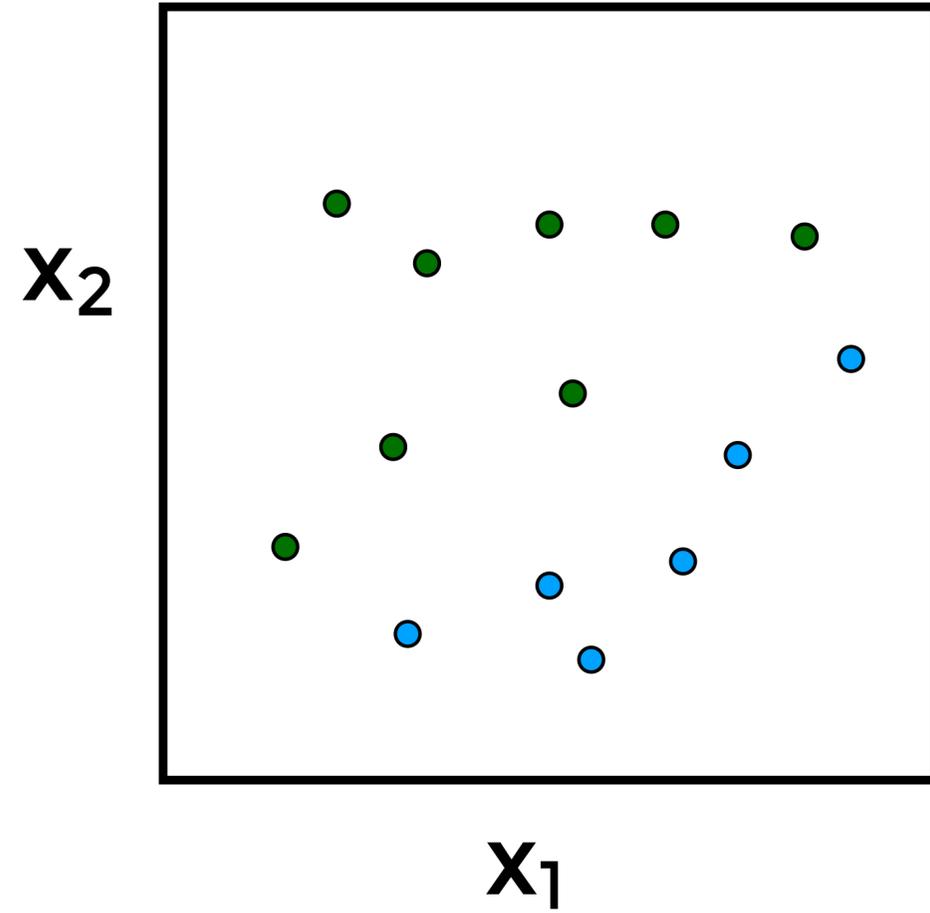
$$\{\mathbf{x}_i\}_{i=0}^n$$

- L'expérience vient sans cible
- La tâche est donc d'apprendre “des propriétés reliées à la structure des données”

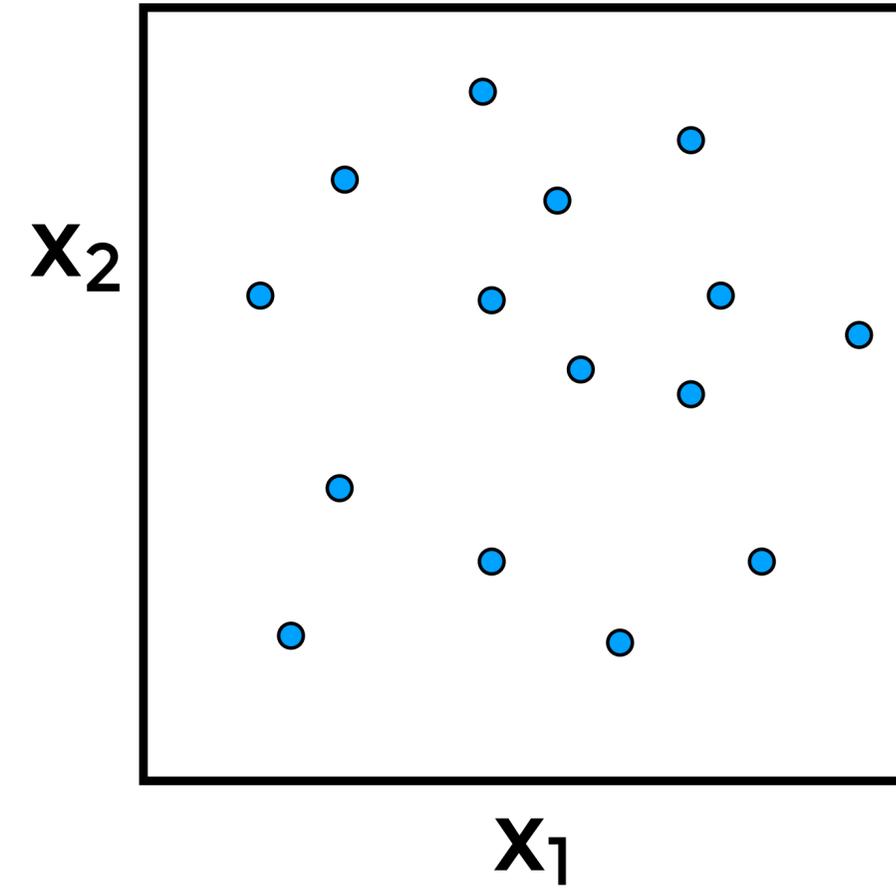
Différentes tâches

- Finding patterns
 - Regroupement $f : X \rightarrow \{1, 2, \dots, K\}$ (K clusters)
 - Réduction de la dim. $f : X^p \rightarrow X^k, k \ll p$
 - Modélisation de la densité $f : X \rightarrow [0, 1]$
 - ...

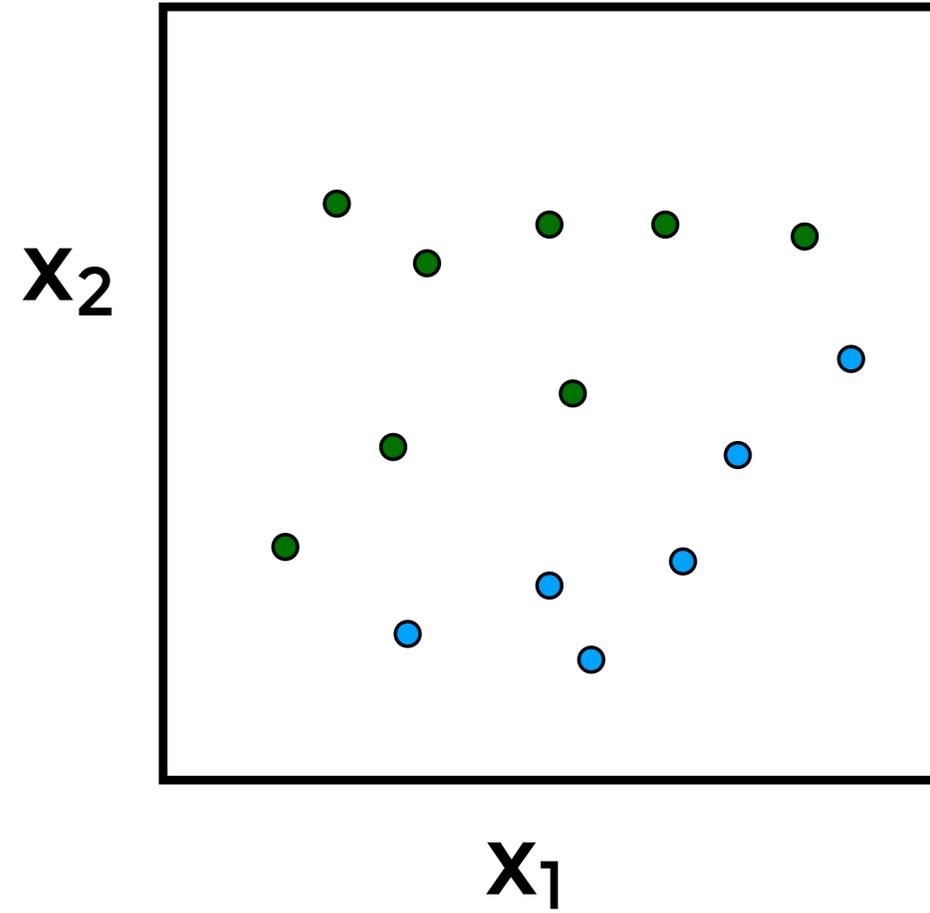
Supervisé



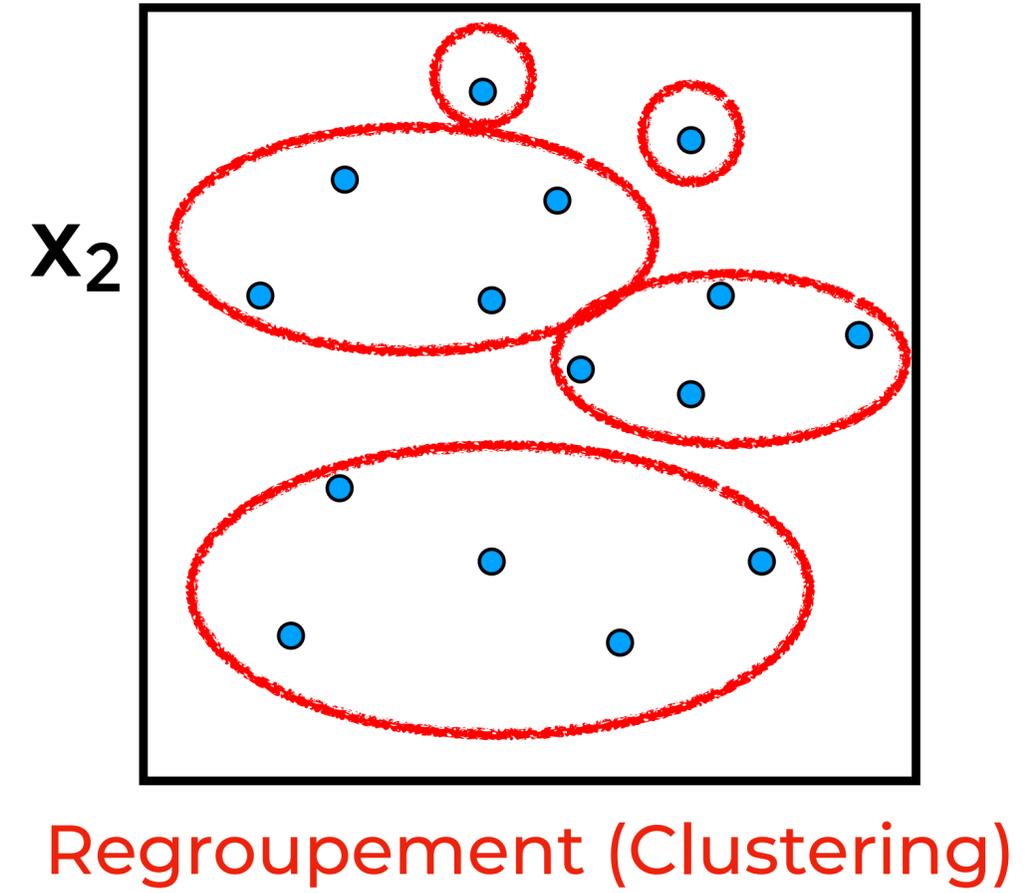
Non supervisé



Supervisé

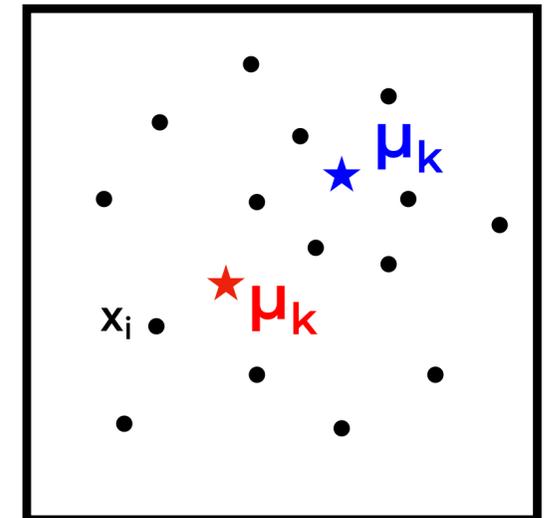


Non supervisé



K-means clustering

- Modèle de regroupement (et un algorithme pour apprendre ses paramètres)
 - Il y a K groupe. Chaque point appartient à un groupe. Les groupes ont un centre: μ
- Objectif: Trouver les centres μ_k qui minimise la distance intragroupe

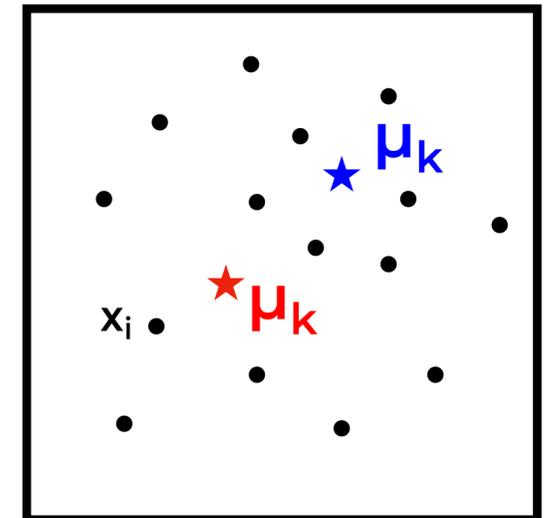


K-means clustering

- Modèle de regroupement (et un algorithme pour apprendre ses paramètres)
 - Il y a K groupe. Chaque point appartient à un groupe. Les groupes ont un centre: μ
- Objectif: Trouver les centres μ_k qui minimise la distance intragroupe

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

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



K-means clustering

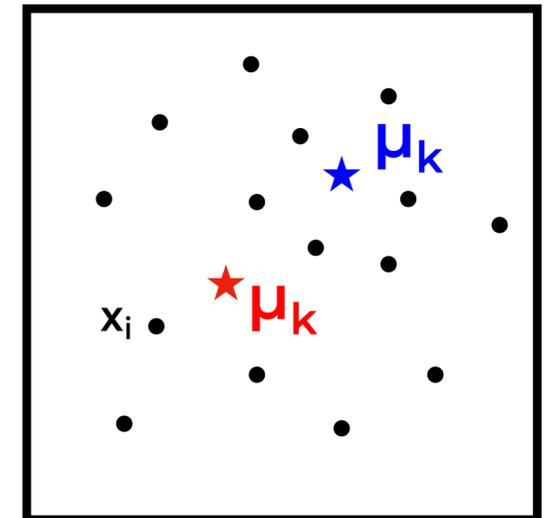
- Modèle de regroupement (et un algorithme pour apprendre ses paramètres)
 - Il y a K groupe. Chaque point appartient à un groupe. Les groupes ont un centre: μ
- Objectif: Trouver les centres μ_k qui minimise la distance intragroupe

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

- Algorithme:

- Initialisation des centres
- On répète jusqu'à convergence:
 1. Mise à jour des responsabilités: r
 2. Mise à jour des centres des groupes: $\mu_k \quad \forall k$

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



K-means clustering

- Modèle de regroupement (et un algorithme pour apprendre ses paramètres)
 - Il y a K groupe. Chaque point appartient à un groupe. Les groupes ont un centre: μ
- Objectif: Trouver les centres μ_k qui minimise la distance intragroupe

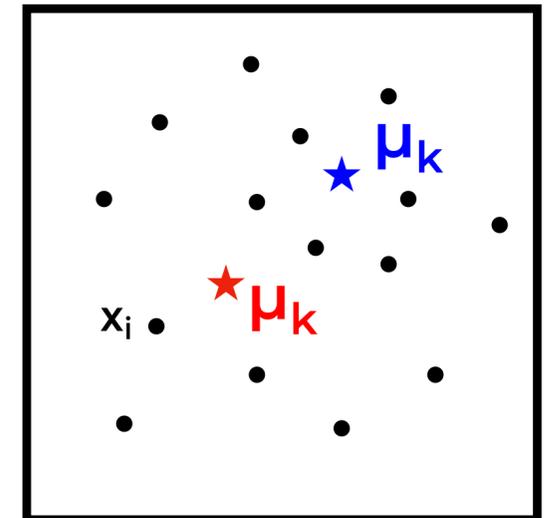
$$\text{Objective} := \sum_{i=1}^N \sum_{k=1}^K r_{ik} \|x_i - \mu_k\|^2$$

- Algorithme:
 - Initialisation des centres
 - On répète jusqu'à convergence:

1. Mise à jour des responsabilités: r

2. Mise à jour des centres des groupes: $\mu_k \quad \forall k$

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



K-means clustering

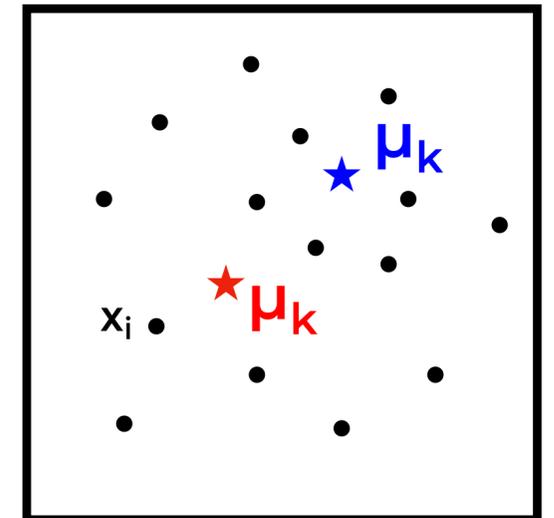
- Modèle de regroupement (et un algorithme pour apprendre ses paramètres)
 - Il y a K groupe. Chaque point appartient à un groupe. Les groupes ont un centre: μ
- Objectif: Trouver les centres μ_k qui minimise la distance intragroupe

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

- Algorithme:

- Initialisation des centres
- On répète jusqu'à convergence:
 1. Mise à jour des responsabilités: r
 2. Mise à jour des centres des groupes: $\mu_k \quad \forall k$

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



K-means clustering

- Modèle de regroupement (et un algorithme pour apprendre ses paramètres)
 - Il y a K groupe. Chaque point appartient à un groupe. Les groupes ont un centre: μ
- Objectif: Trouver les centres μ_k qui minimise la distance intragroupe

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

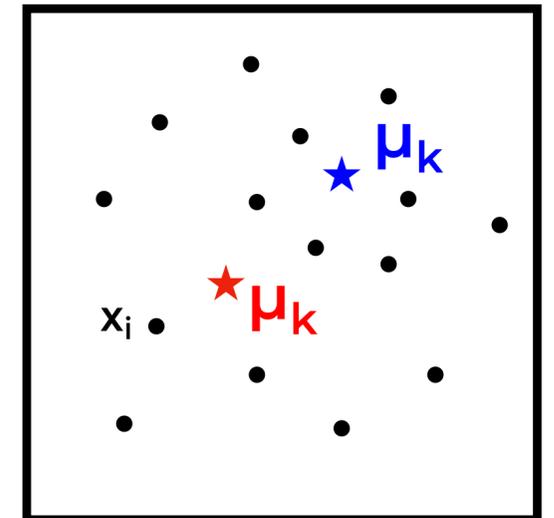
- Algorithme:

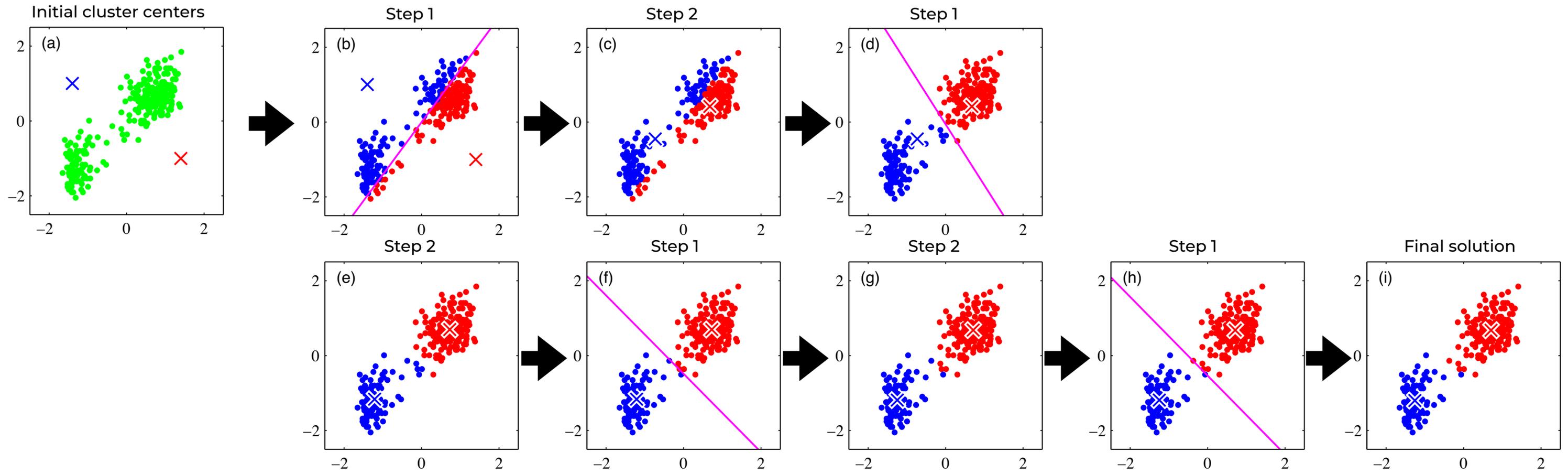
- Initialisation des centres
- On répète jusqu'à convergence:

1. Mise à jour des responsabilités: r

2. Mise à jour des centres des groupes: $\mu_k \quad \forall k$

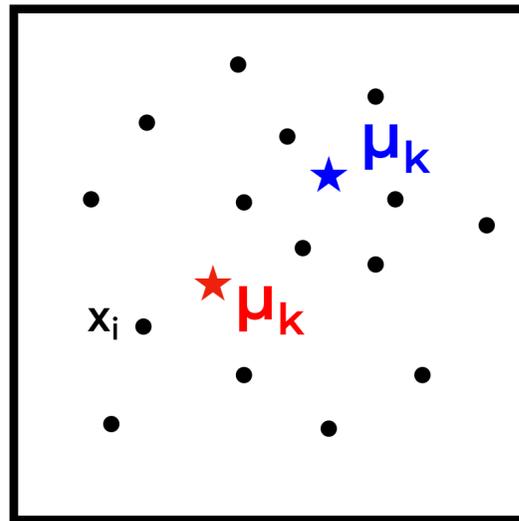
$$r = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}_{N \times 2}$$





Une approche probabiliste pour K-means

Regroupement K-means

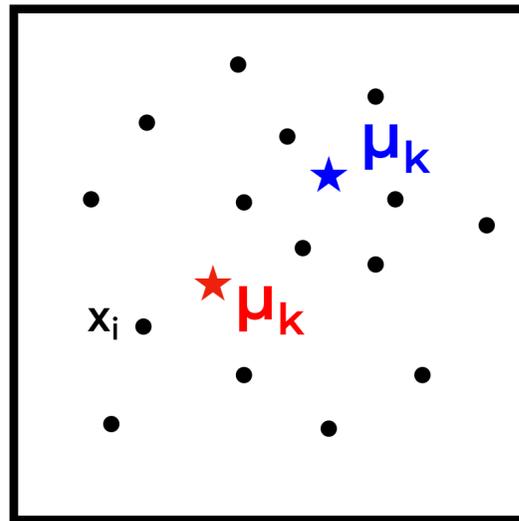


$$\text{Objective} := \sum_{i=1}^N \sum_{k=1}^K r_{ik} \| \mathbf{x}_i - \mu_k \|^2$$

- Les responsabilités sont continues $[0, 1]$
 - Chaque groupe en a une : π_k
- Chaque groupe modélise les données avec une
- gaussienne: $\mathcal{N}(\mathbf{x}_i | \mu_k, \Sigma_k)$

Une approche probabiliste pour K-means

Regroupement K-means

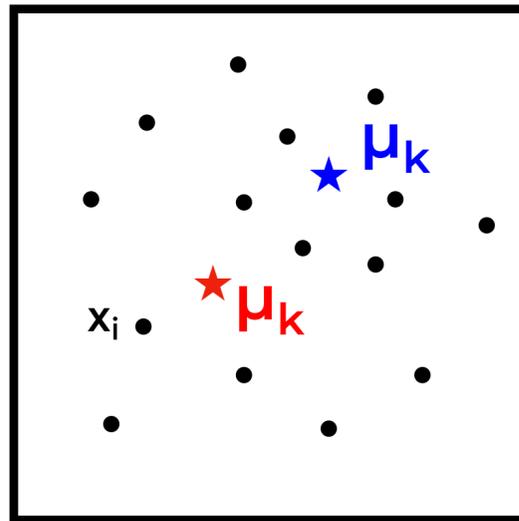


$$\text{Objective} := \sum_{i=1}^N \sum_{k=1}^K \overset{\text{Responsabilité}}{r_{ik}} \left\| \mathbf{x}_i - \overset{\text{Centre}}{\mu_k} \right\|^2$$

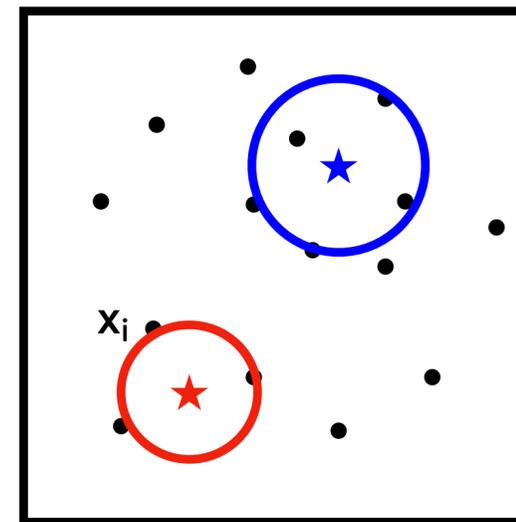
- Les responsabilités sont continues $[0, 1]$
 - Chaque groupe en a une : π_k
- Chaque groupe modélise les données avec une
- gaussienne: $\mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$

Une approche probabiliste pour K-means

Regroupement K-means



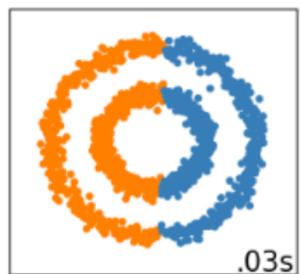
Regroupement "Soft" K-means



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

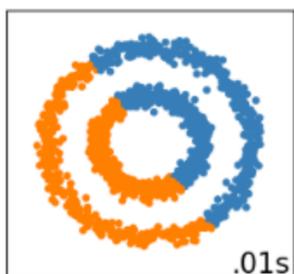
- Les responsabilités sont continues [0, 1]
 - Chaque groupe en a une : π_k
- Chaque groupe modélise les données avec une
- gaussienne: $\mathcal{N}(\mathbf{x}_i | \mu_k, \Sigma_k)$

K-means

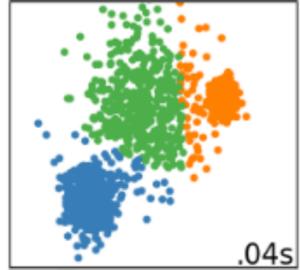


Similaire

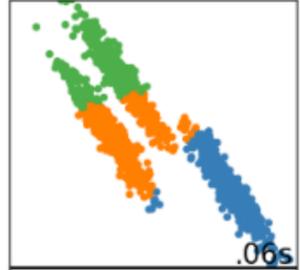
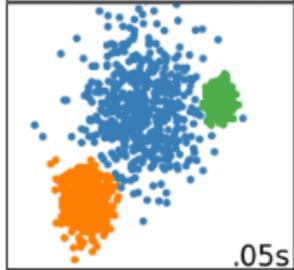
GMMs



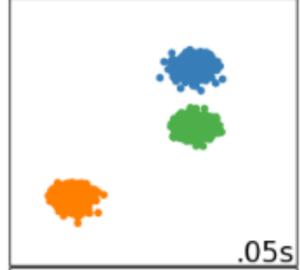
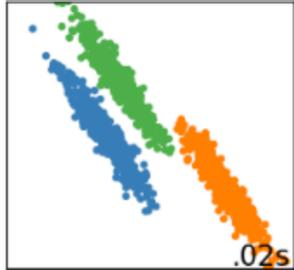
Similaire



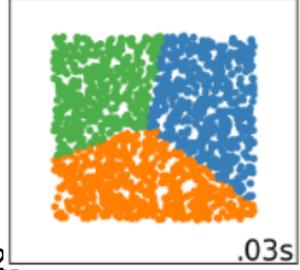
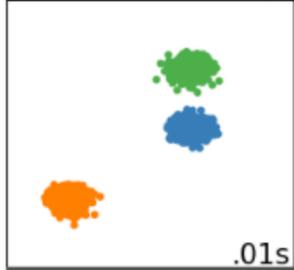
GMM est mieux



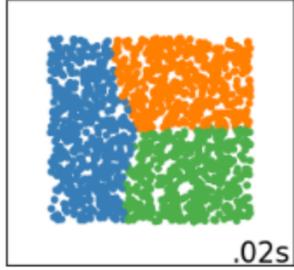
GMM est mieux



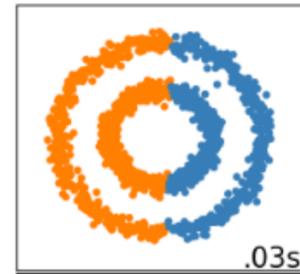
Similaire



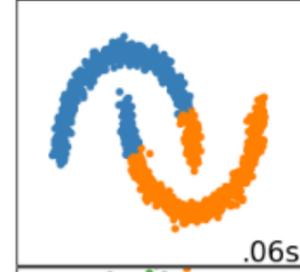
Similaire



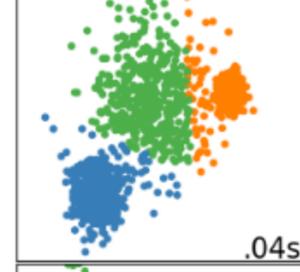
K-means



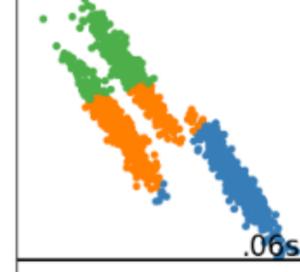
Similaire



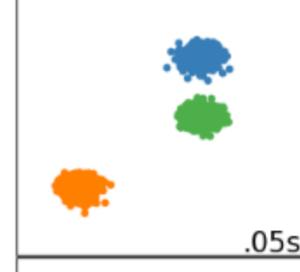
Similaire



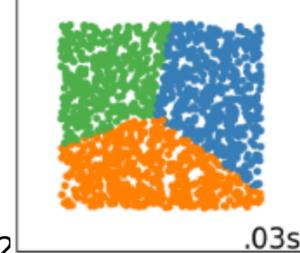
GMM est mieux



GMM est mieux

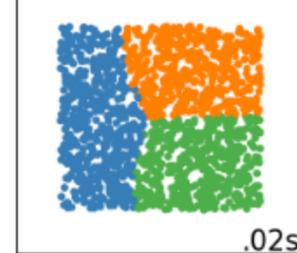
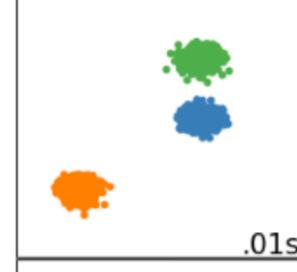
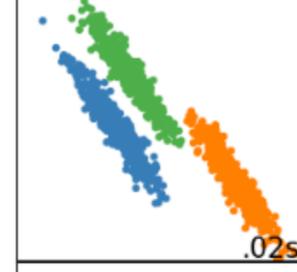
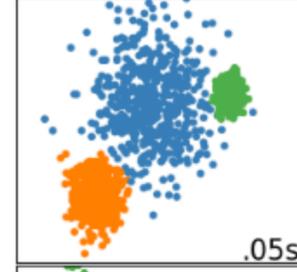
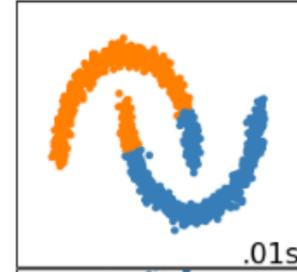
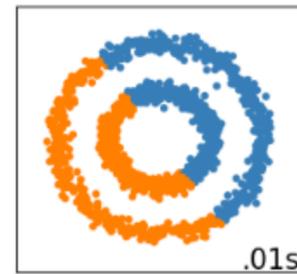


Similaire



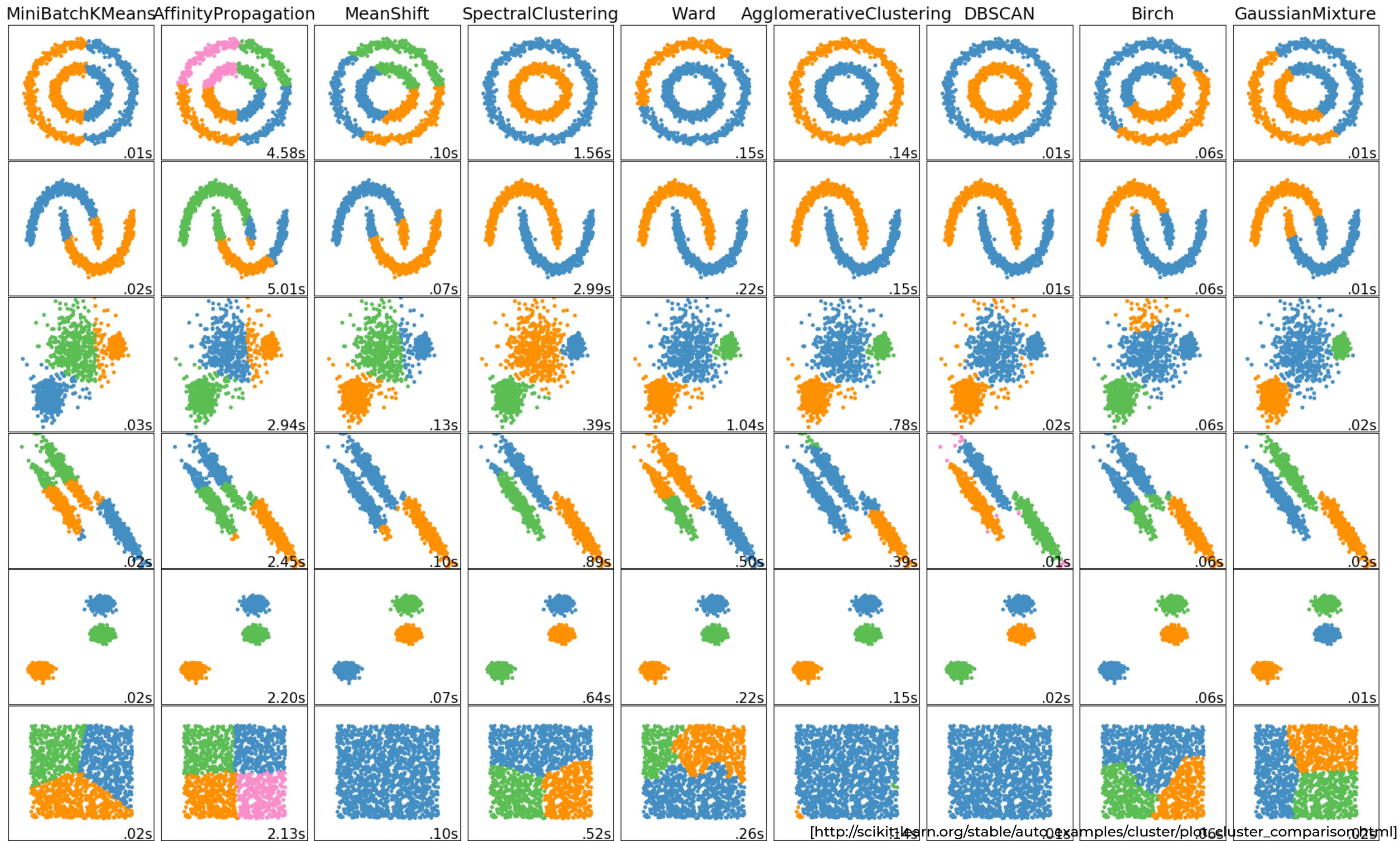
Similaire

GMMs



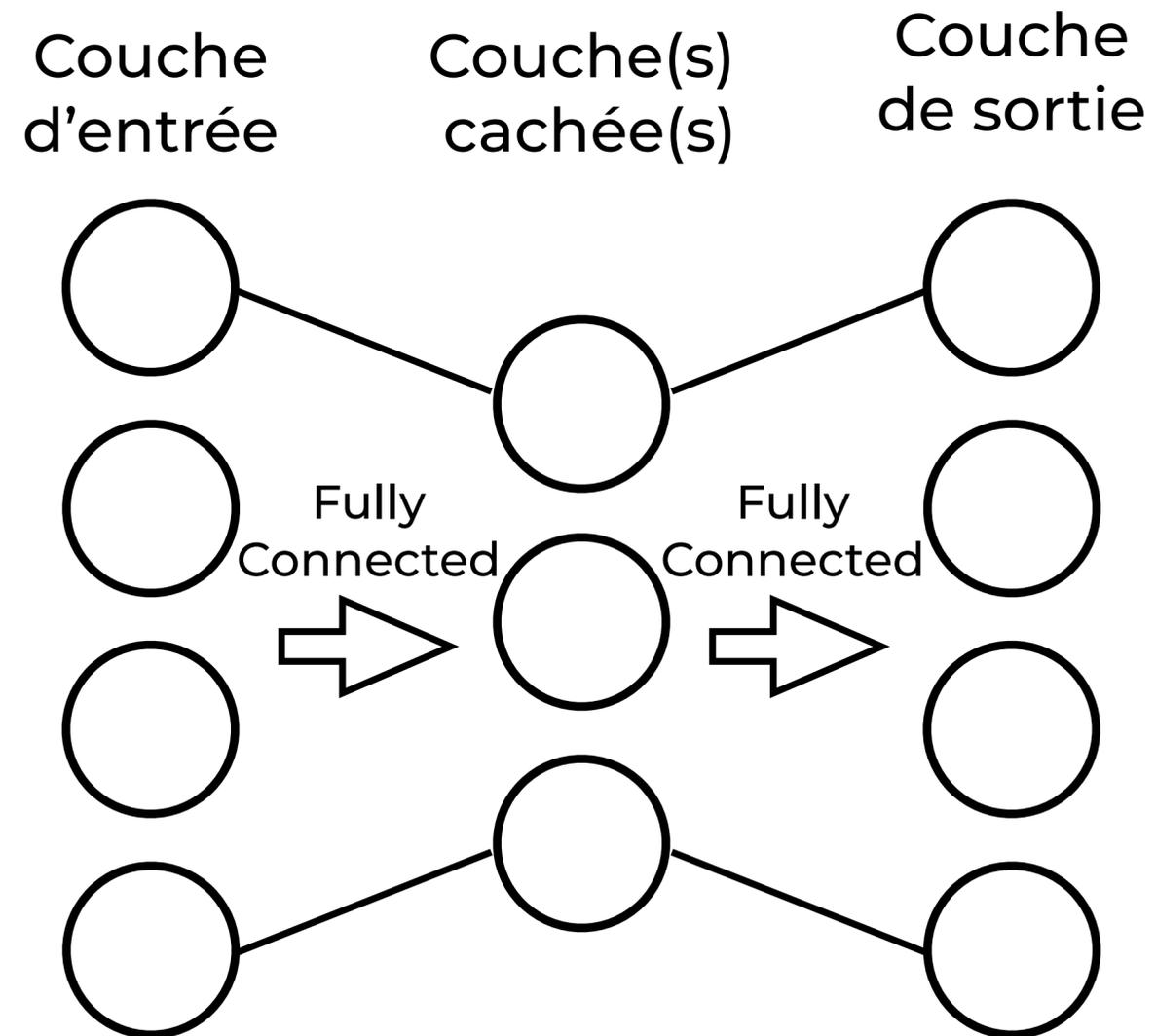
Comparaison de K-means au GMMs

- Les GMMs apprennent une matrice de covariance
- Variance par groupe
- Covariance entre les groupes
- GMMs a bien plus de paramètres
- Matrice de covariance (MxM)



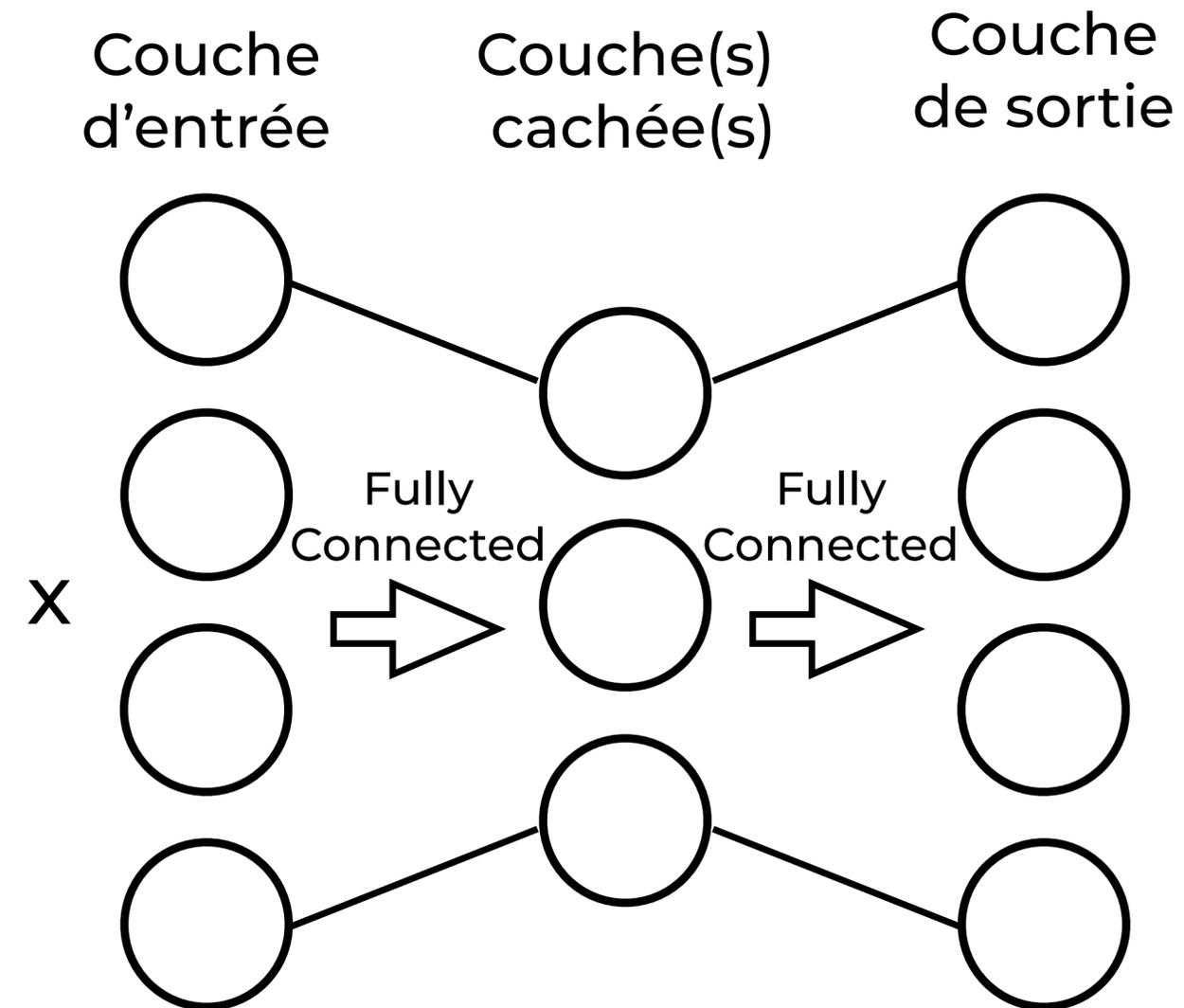
Autoencodeurs

- Une architecture de réseaux de neurones pour l'apprentissage non supervisé



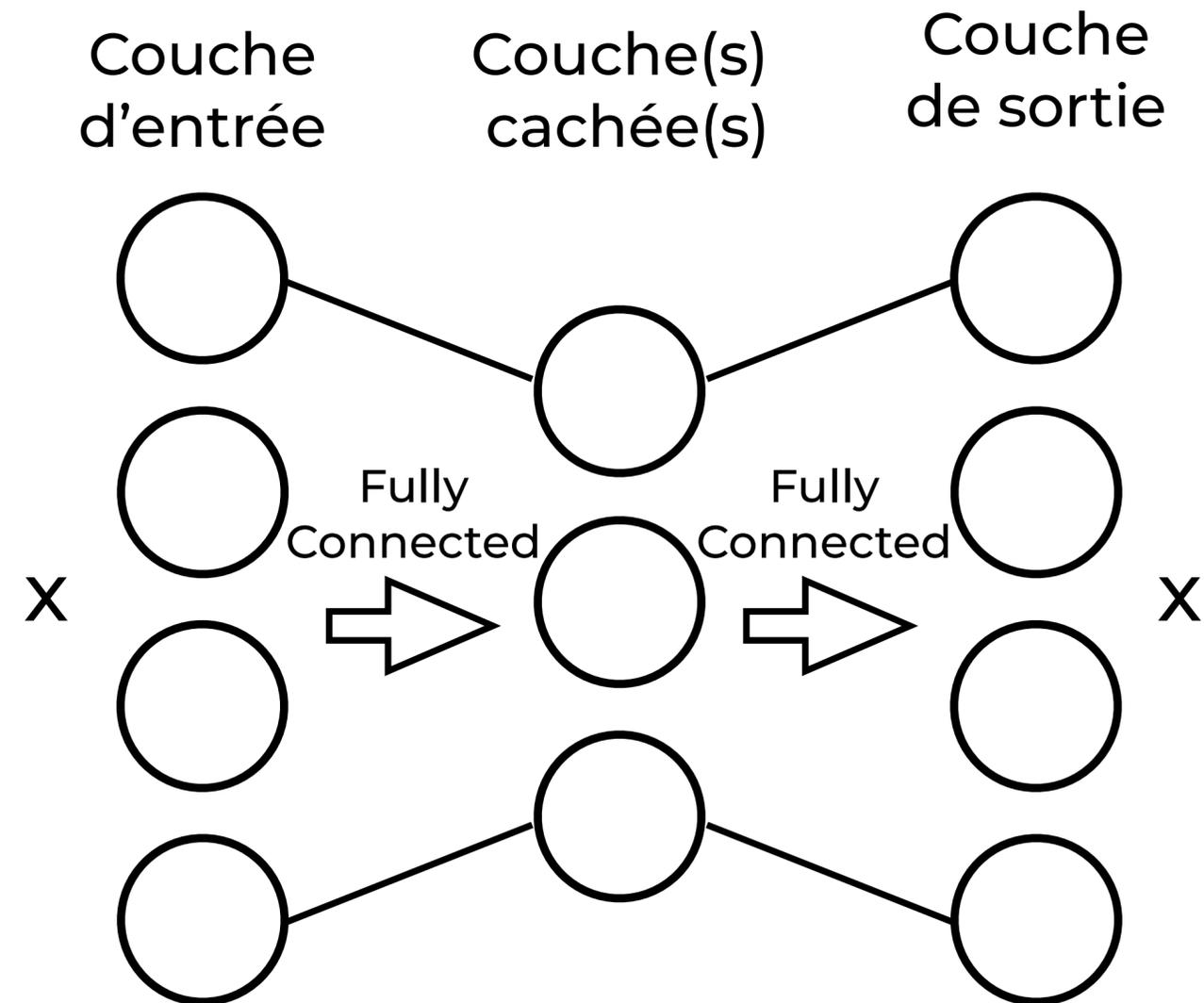
Autoencodeurs

- Une architecture de réseaux de neurones pour l'apprentissage non supervisé



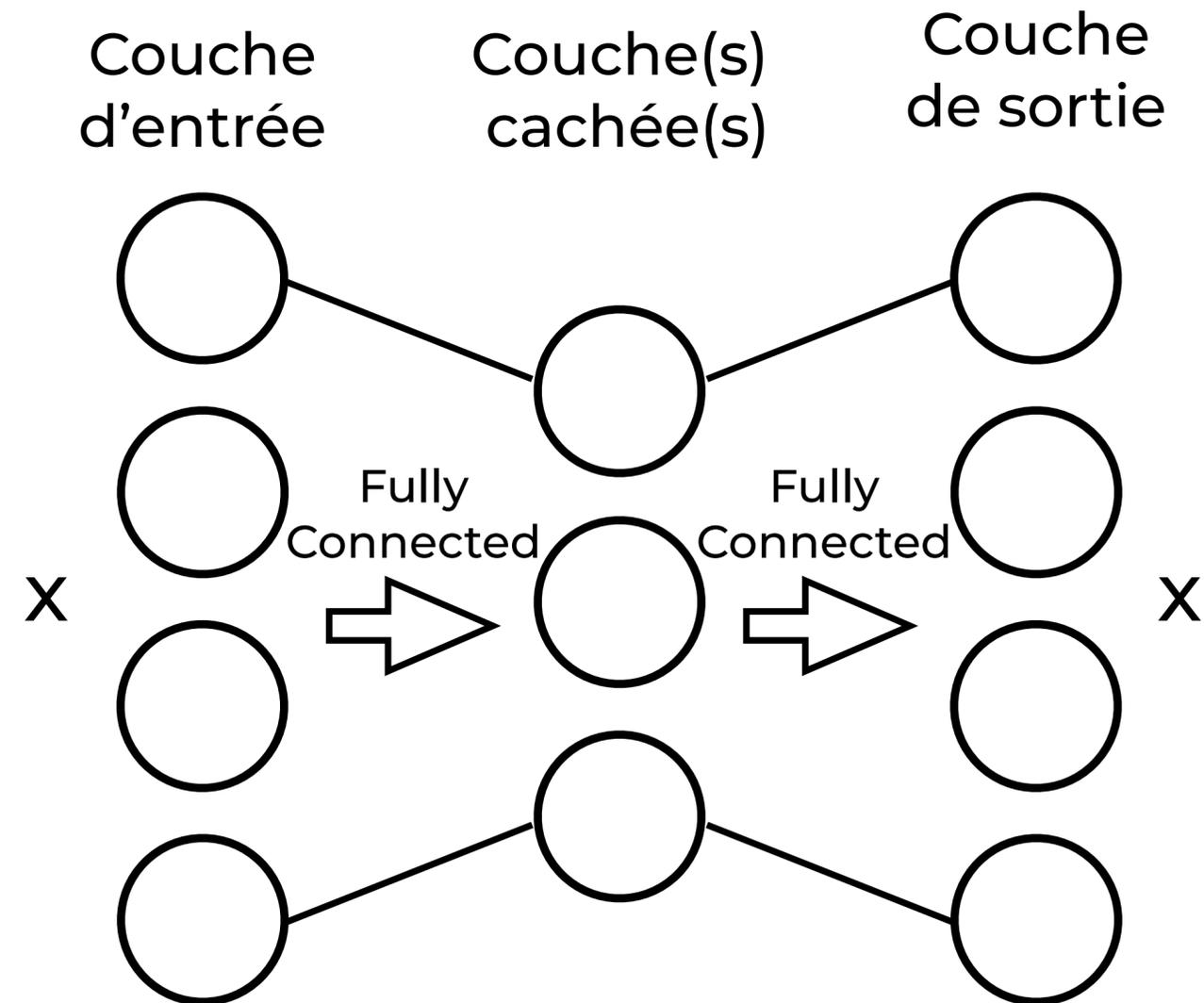
Autoencodeurs

- Une architecture de réseaux de neurones pour l'apprentissage non supervisé



Autoencodeurs

- Une architecture de réseaux de neurones pour l'apprentissage non supervisé

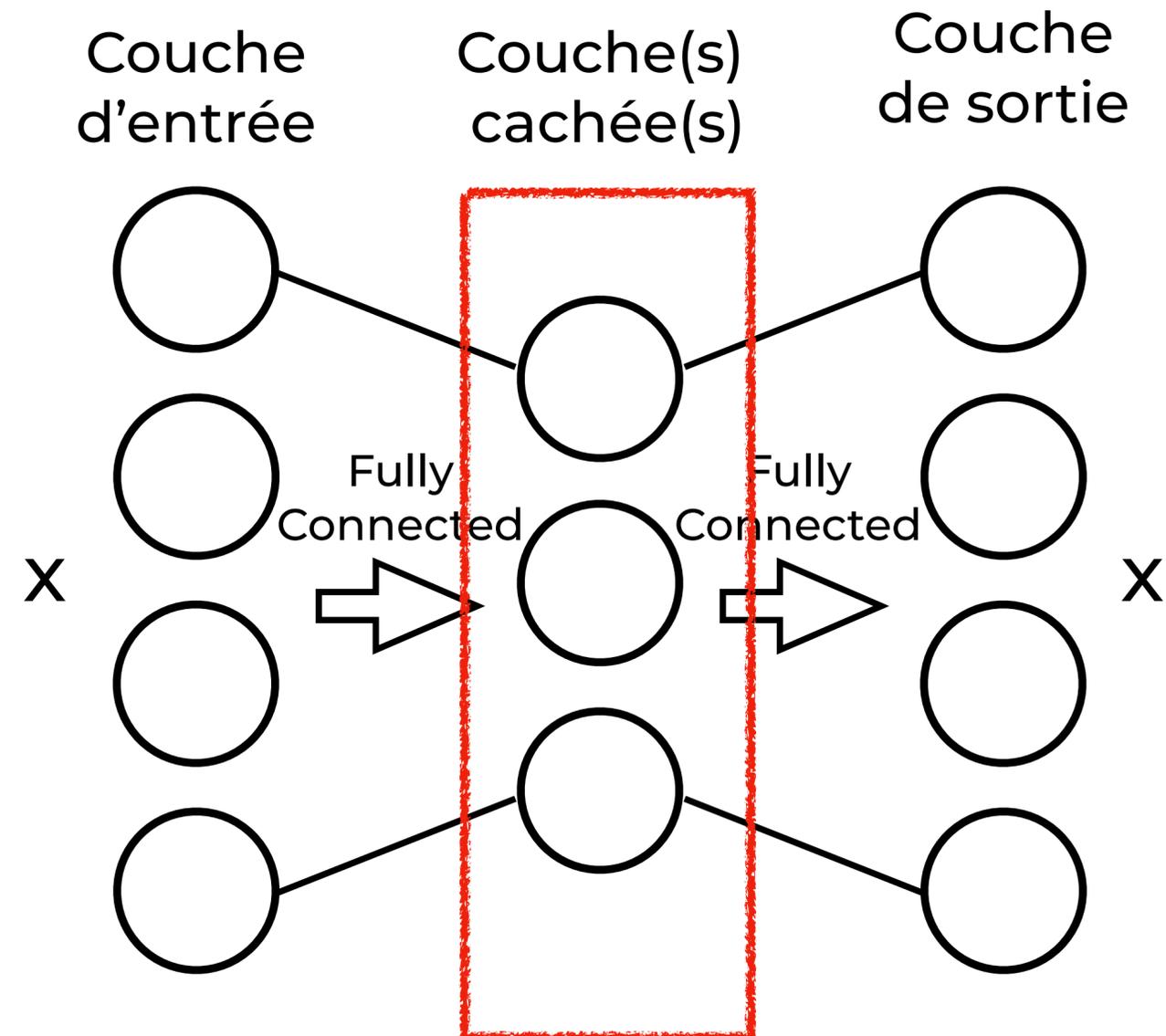


Objectif:

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

Autoencodeurs

- Une architecture de réseaux de neurones pour l'apprentissage non supervisé



Objectif:

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

Représentation en plus petite dimensionnalité