

Introduction à l'Intelligence Artificielle

Cours 4 – mercredi 19 novembre 2022

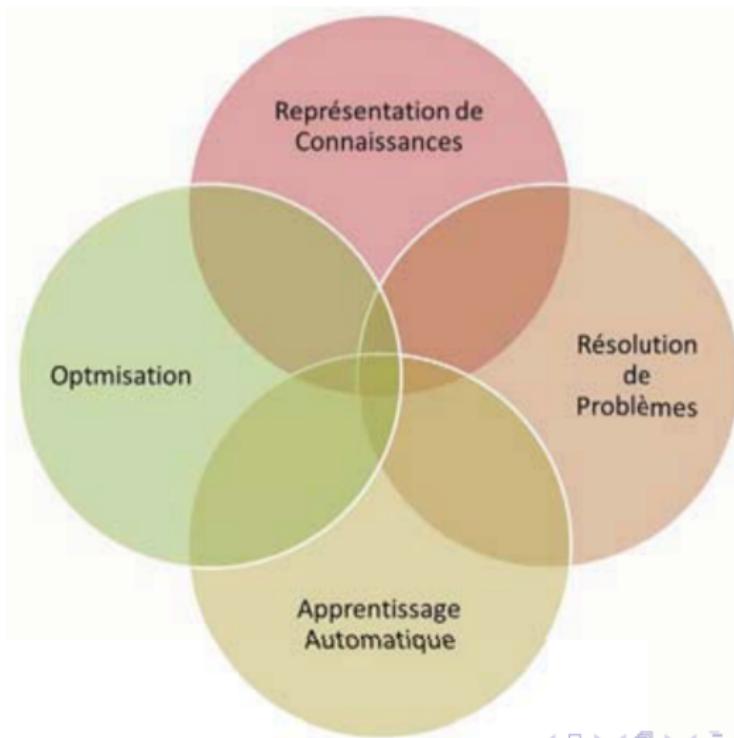
Adrien Revault d'Allonnes

ara@up8.edu

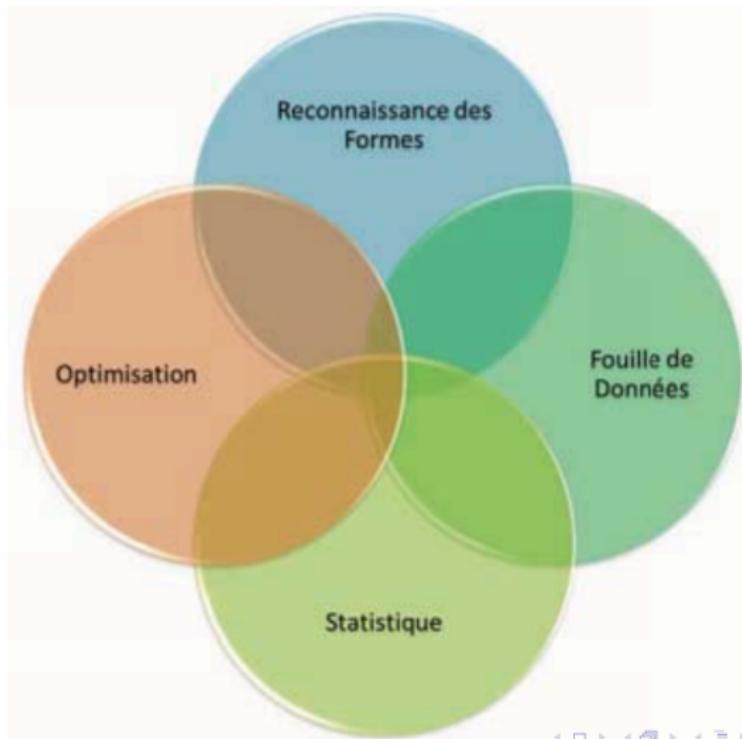
Université Paris 8 – Vincennes à Saint-Denis

FIA – sept. à déc., 2022

Intelligence artificielle



Apprentissage automatique



Plusieurs définitions

- Learning is making useful changes in mind – [Marvin Minsky, 1985]
- Learning is any change in a system that allows it to perform better the second time on repetition of the same task or another task drawn from the same population – [Herbert Simon, 1983]
- Learning is the organisation of experience – [Scott, 1983]
- Learning is constructing or modifying representations of what is being experienced – [Riszard Michalski, 1986]

Apprentissage

- Plusieurs définitions (écoles) de l'apprentissage
 - cognitivistes
 - postulat : assimilation de l'organisme à un ordinateur, à une machine à traiter les informations
 - apprentissage : l'apprentissage vise à créer de nouveaux programmes
 - Théories écologiques
 - postulat : le comportement dynamique traduit l'adaptation du système aux contraintes qui pèsent sur lui
 - apprentissage : intégrer de nouveaux états stables

Allez voir les spécialistes... .

Apprentissage

- C'est quoi la science de l'apprentissage ?
 - on étudie les apprentissages
 - naturels
 - artifiels
 - des théories
 - des modèles
 - des algorithmes
- Différentes écoles
 - les symbolistes
 - les connexionnistes (rejoins par les statisticiens)
 - statistical machine learning

Ordinateurs intelligents (par apprentissage)

• Problématique

- nous souhaitons avoir des ordinateurs
 - intelligents
 - adaptatifs
 - avec un comportement robuste
- programmer de tels comportements est souvent impossible
 - i.e. intelligence artificielle dans les jeux (scripts)

• Solution

- faire un ordinateur capable de se programmer lui-même
 - à partir d'exemples (apprentissage classique / par imitation)
 - à partir de son « expérience » (apprentissage par renforcement)

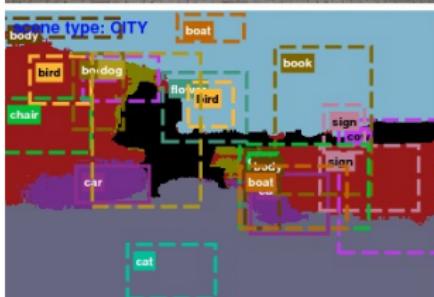
Exemples

Écrire un programme permettant de ...



Exemples

Écrire un programme permettant de ...



Exemples

Écrire un programme permettant de ...



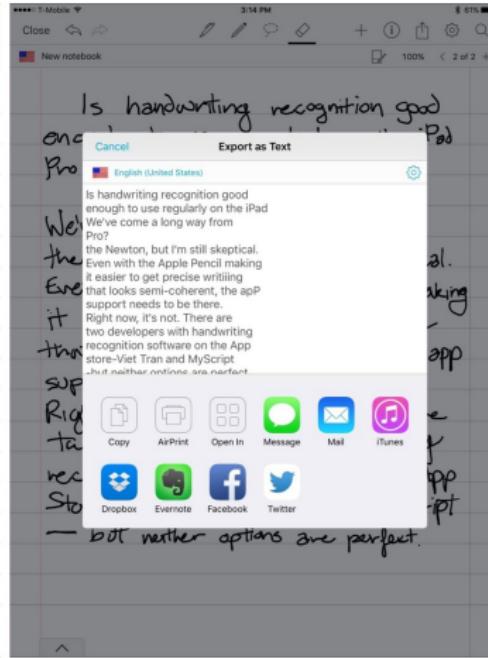
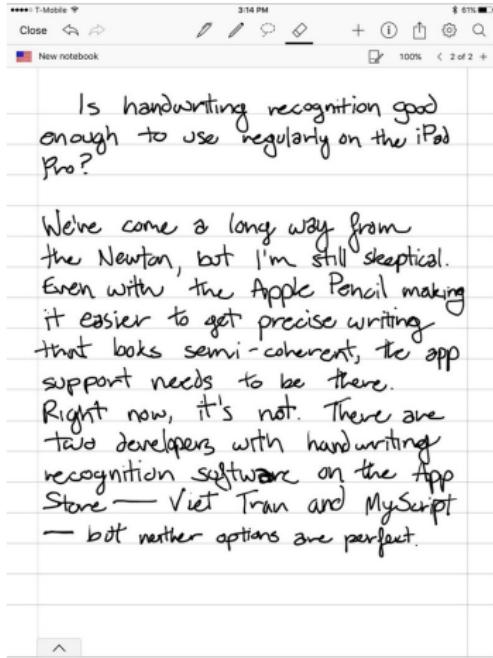
Exemples

Écrire un programme permettant de ...



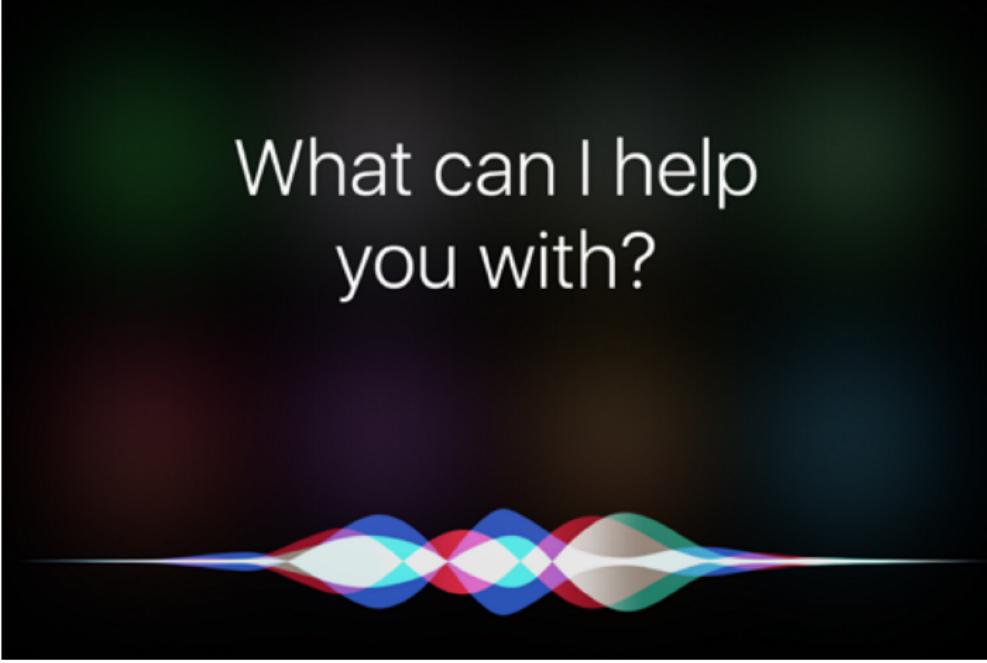
Exemples

Écrire un programme permettant de ...



Exemples

Écrire un programme permettant de ...



What can I help
you with?

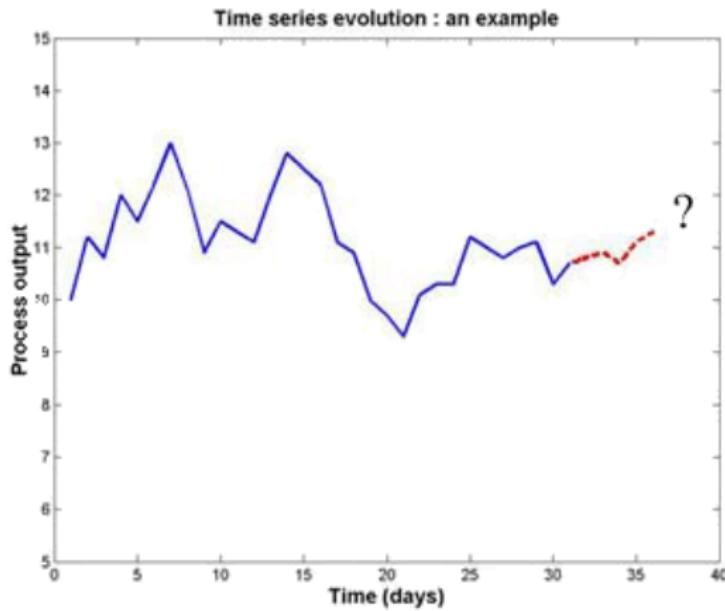
Exemples

Écrire un programme permettant de ...



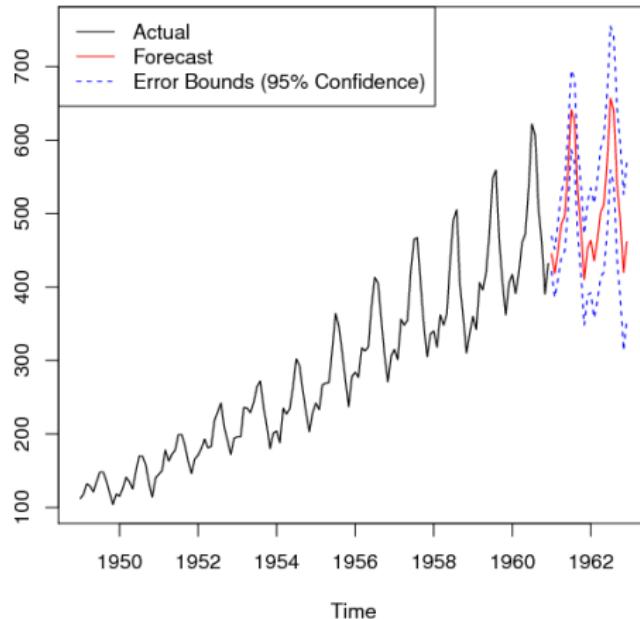
Exemples

Écrire un programme permettant de ...



Exemples

Écrire un programme permettant de . . .



Exemples

Écrire un programme permettant de ...



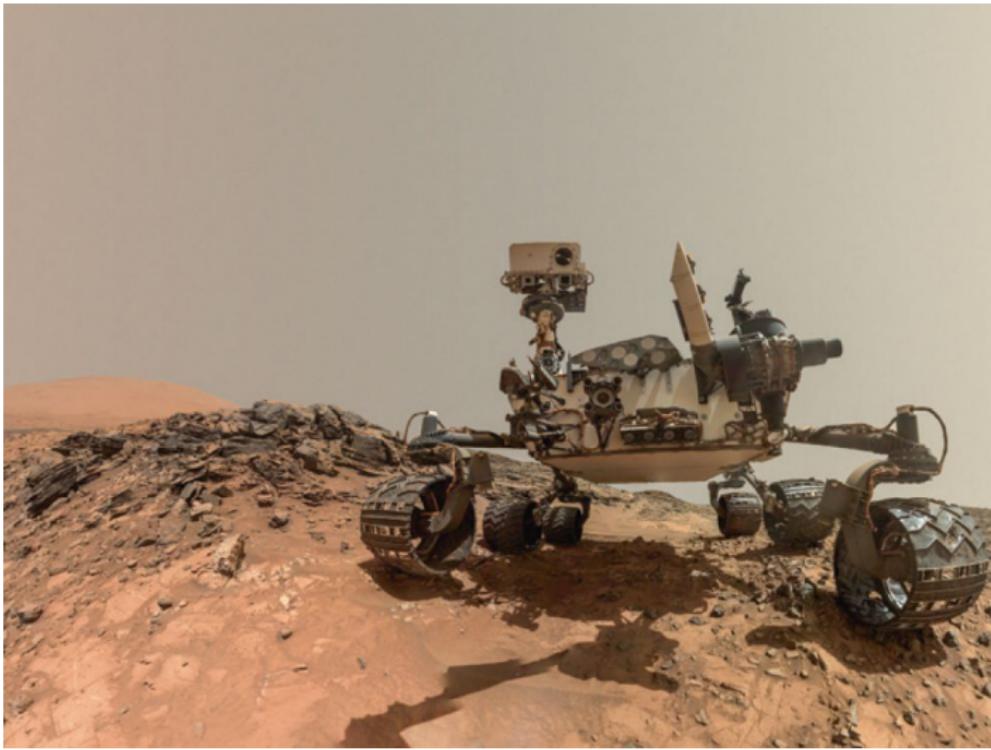
Exemples

Écrire un programme permettant de ...



Exemples

Écrire un programme permettant de ...



Exemples

Écrire un programme permettant de ...



Exemples

Écrire un programme permettant de ...



Quand utiliser l'apprentissage ?

- Il n'y a pas d'expert humain
- L'utilisation d'un expert humain est trop coûteuse ou trop peu performante
- La quantité de données est telle qu'une analyse manuelle est impossible
- Les données évoluent rapidement dans le temps
- Les modèles doivent être adaptés à l'utilisateur
- Les systèmes doivent s'adapter facilement à des conditions opérationnelles différentes

Apprentissage automatique (Machine Learning)

- Différents contextes d'apprentissage
 - **l'apprentissage supervisé** : un expert est employé pour étiqueter correctement des exemples
 - **l'apprentissage non supervisé** : aucun expert n'est disponible, l'algorithme doit découvrir par lui-même la structure des données
 - **l'apprentissage par renforcement** : l'algorithme apprend un comportement
- Exemple de modèles
 - les machines à vecteurs de support
 - les réseaux de neurones
 - les arbres de décision
 - les modèles graphiques
 - les algorithmes génétiques et la programmation génétique

Exemple illustratif – Léon Bottou 2010

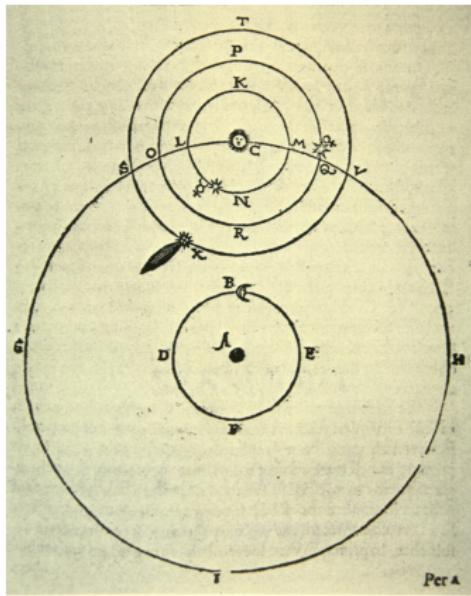
Prolématique : quelles sont les lois explicatives du mouvement des étoiles ?



Observation des étoiles

Exemple illustratif – Léon Bottou 2010

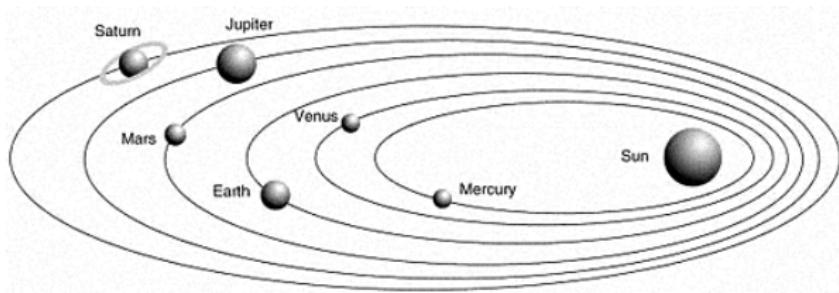
Prolématique : quelles sont les lois explicatives du mouvement des étoiles ?



Émission d'hypothèses – Système de Tycho Brahé (XVI^e)

Exemple illustratif – Léon Bottou 2010

Prolématique : quelles sont les lois explicatives du mouvement des étoiles ?



Émission d'hypothèses – Lois de Kepler (XVI^e)

Exemple illustratif – Léon Bottou 2010

Prolématique : quelles sont les lois explicatives du mouvement des étoiles ?

- Validation
 - par Galilée pour la planète Vénus en 1610
 - explication donnée par Newton en 1687

Apprentissage supervisé : notations

- On considère
 - un espace de représentation \mathcal{X}
 - un espace de sortie \mathcal{Y}
 - un ensemble d'apprentissage $X = \{(x_1, y_1), \dots, (x_n, y_n)\}$
- On cherche à apprendre, à partir de X , une fonction f_θ telle que
 - $f_\theta : \mathcal{X} \longrightarrow \mathcal{Y}$
 - θ est l'ensemble des **paramètres** de la fonction (**à apprendre**)
 - $f_\theta(x)$ prédit l'étiquette y associée à x
 - sur l'ensemble d'apprentissage X
 - sur l'ensemble des x qui n'appartiennent pas à X (généralisation)

Apprentissage supervisé : formalisation

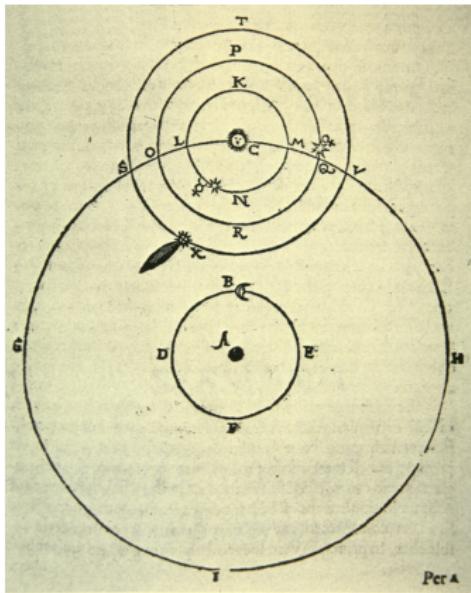
- On considère
 - un espace de représentation \mathcal{X}
 - un espace de sortie \mathcal{Y}
 - un ensemble d'apprentissage $X = \{(x_1, y_1), \dots, (x_n, y_n)\}$



Observation des étoiles

Apprentissage supervisé : formalisation

- On cherche à apprendre, à partir de X , une fonction f_θ telle que
 - $f_\theta : \mathcal{X} \rightarrow \mathcal{Y}$



Quelle fonction f ? Quels paramètres θ

Apprentissage supervisé : formalisation

- $f_\theta(x)$ prédit l'étiquette y associée à x
 - sur l'ensemble d'apprentissage X
 - sur l'ensemble des x qui n'appartiennent pas à X (généralisation)
- Validation
 - par Galilée pour la planète Vénus en 1610
 - explication donnée par Newton en 1687

Les deux problèmes clefs de l'apprentissage

- Les modèles d'apprentissage posent deux problèmes principaux
 - **inférence** : comment calculer $f_\theta(x)$ connaissant x (efficacement, si possible) ? Quel modèle utiliser ?
 - **apprentissage** : soit un ensemble de n exemples notés $\{x_1, x_2, \dots, x_n\}$ i.i.d. selon une distribution inconnue (appelé ensemble d'apprentissage). Quelles sont les valeurs des paramètres θ qui permettent la meilleure prédiction (sur des données inconnues) ?

La classification

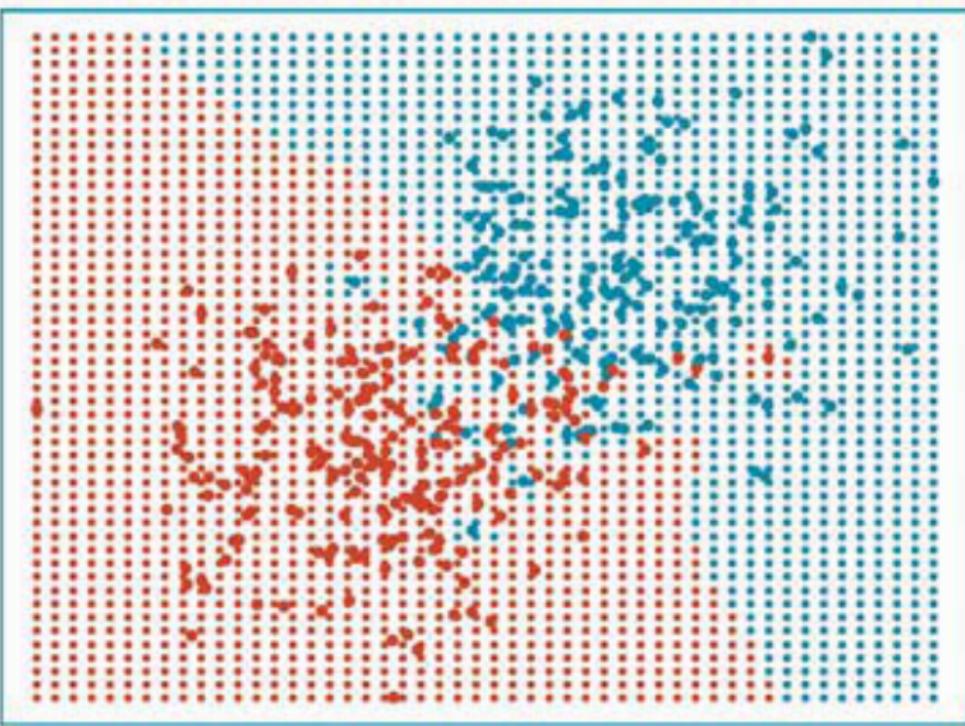
- Hypothèses

- x est un vecteur de taille n (2, dans nos illustrations)
- y est une catégorie : soit -1 , soit $+1$
- on appelle cela de la classification binaire
 - classification de spams dans les emails
 - classification à la hausse ou à la baisse de la bourse
 - le bien, le mal, ...

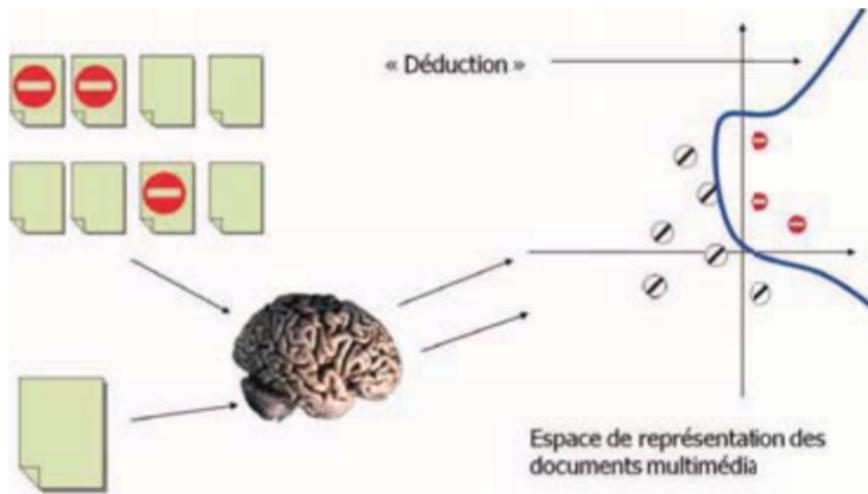
Ensemble d'apprentissage



Frontière de décision



La classification : processus



La classification

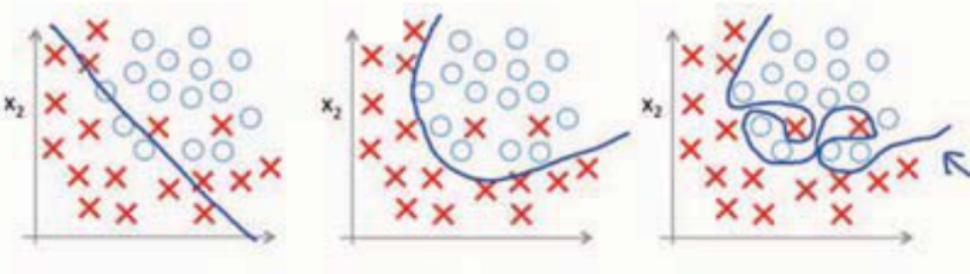
- Apprentissage
 1. transformer les données étiquetées en vecteurs (étiquetés)
 2. trouver une frontière de décision f_θ
- Inférence
 - pour chaque nouvelle donnée, regarder de quel côté de la barrière elle se trouve en calculant $f_\theta(x)$

La classification

- Question 1 : quelle fonction f_θ utiliser – quelle hypothèse ?
- Question 2 : comment trouver les meilleures valeurs de θ ?
- Question 3 : quelle est la qualité du modèle trouvé ?

La classification

- Question 1 : quelle fonction f_θ utiliser – quelle hypothèse ?



La classification

- Question 1 : quelle fonction f_θ utiliser – quelle hypothèse ?



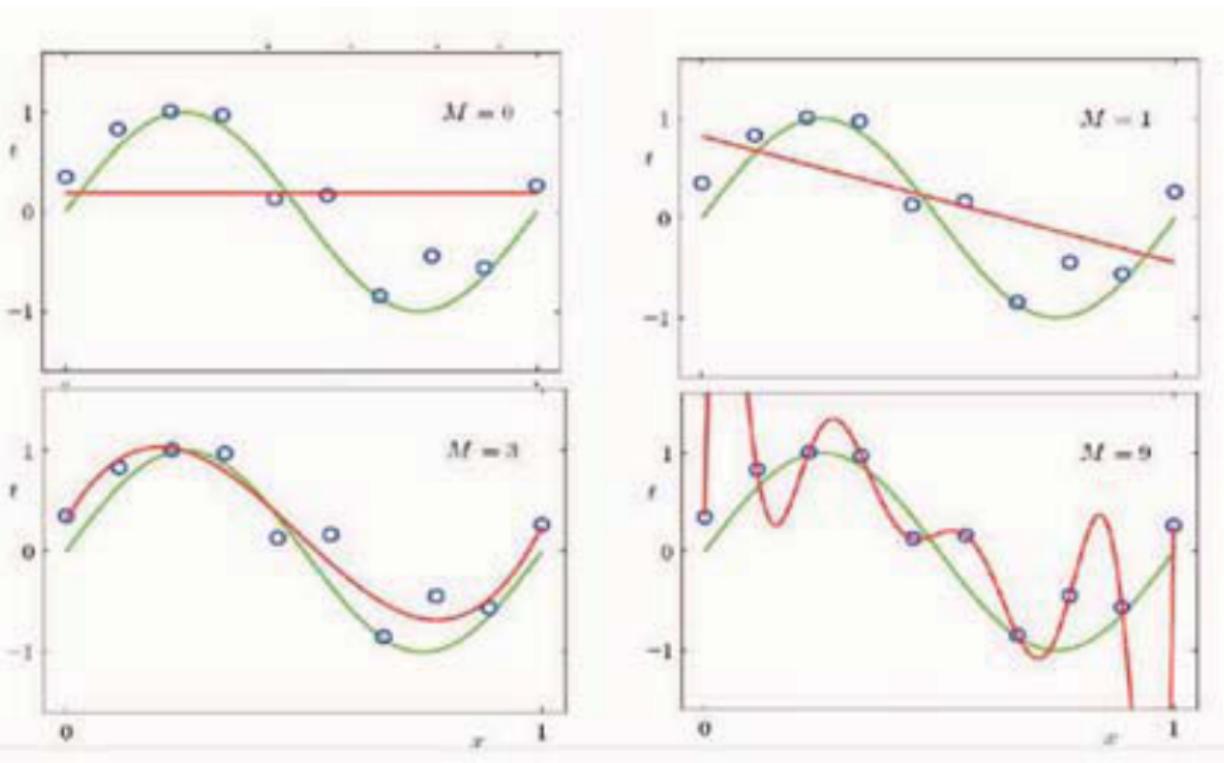
La classification

- Question 1 : quelle fonction f_θ utiliser – quelle hypothèse ?
- Sous-apprentissage
 - le fait qu'un modèle n'a pas la capacité à bien modéliser les données d'apprentissage
- Sur-apprentissage
 - le fait que, par sa trop grande capacité à stocker les informations, une structure une structure dans une situation de **surapprentissage**¹ aura de la peine à généraliser les caractéristiques des données

1. ou overfitting ou apprentissage par cœur

La classification

- Question 1 : quelle fonction f_θ utiliser – quelle hypothèse ?



Apprentissage

- Question 2 : comment trouver les meilleures valeurs de θ ?
- Critère d'apprentissage
Principe : Trouver les paramètres qui, sur l'ensemble d'apprentissage, minimisent/maximisent un critère donné (problème d'optimisation)

$$\theta^* = \operatorname{argmin}/\operatorname{argmax} F(\theta, \mathcal{X})$$

- quel critère ?
 - maximum de vraisemblance
 - risque empirique
 - risque empirique régularisé
 - ...

Apprentissage

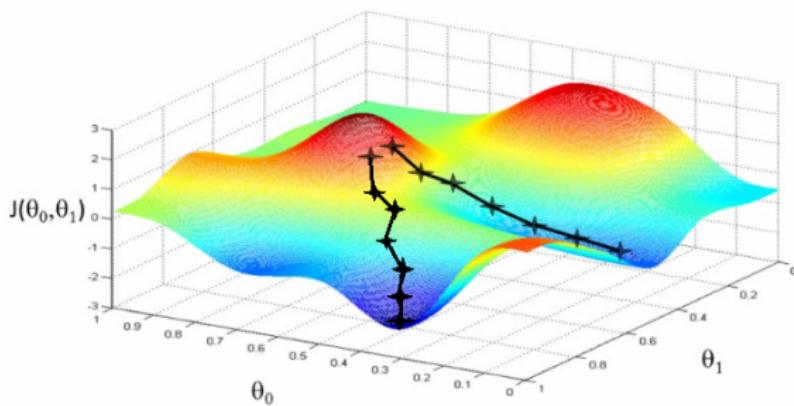
- Question 2 : comment trouver les meilleures valeurs de θ ?
- Plein de modèles existent
 - réseaux de neurones
 - machines à vecteurs de support
 - HMM
 - réseaux bayésiens
 - KNN
 - arbres de décision

Apprentissage

- Question 2 : comment trouver les meilleures valeurs de θ ?
- Plein de modèles existent
 - les différences entre les modèles existant sont
 - complexité (mémoire/CPU)
 - type d'entrées/sorties (vecteurs, séquences, structures, ...)
 - méthode d'optimisation sous-jacente (optimisation convexe, non convexe)
 - performance

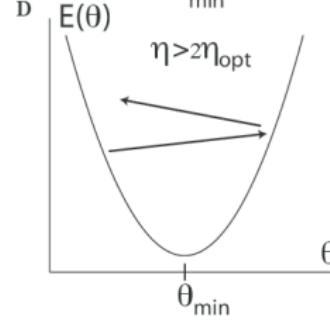
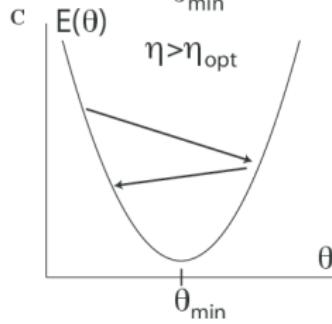
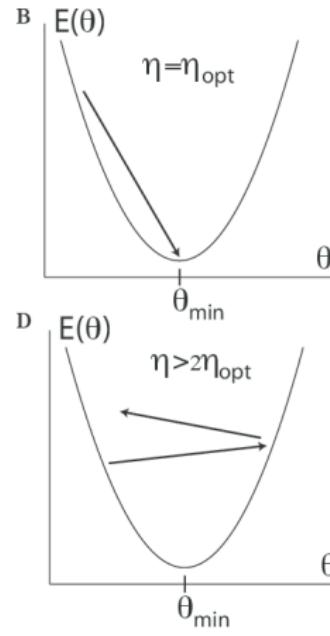
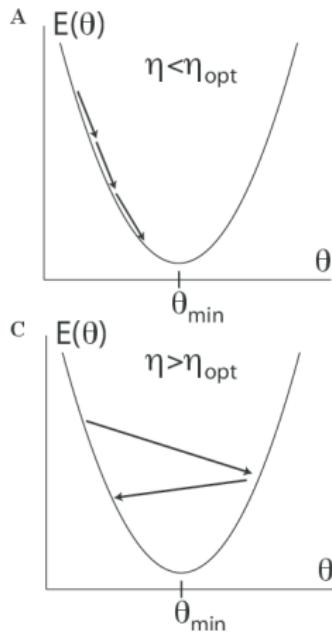
Apprentissage

- Question 2 : comment trouver les meilleures valeurs de θ ?



Apprentissage

- Question 2 : comment trouver les meilleures valeurs de θ ?



Apprentissage

- Question 2 : comment trouver les meilleures valeurs de θ ?

Évaluation

- Question 3 : quelle est la qualité du modèle trouvé ?
- Problème non trivial
 - évaluation sur les données d'apprentissage ? **NON !**
 - évaluation sur de nouvelles données étiquetées
 - prendre la base d'apprentissage
 - la couper en deux
 - entraîner sur une partie
 - tester sur l'autre
 - problème : réduction du nombre d'exemples d'apprentissage

Évaluation

- Question 3 : quelle est la qualité du modèle trouvé ?
- Validation croisée²

Méthode d'estimation de fiabilité d'un modèle fondée sur une technique d'échantillonnage

