

# Chapitre1: Introduction

# Contenu de la matière

- Chapitre 01: Introduction
- Chapitre 02: Concepts d'apprentissage
- Chapitre 03: Arbres de décision
- Chapitre 04: Classificateur de Bayes
- Chapitre 05: La régression logistique
- Chapitre 06: Machine à vecteurs de supports
- Chapitre 07: Les réseaux de neurones
- Chapitre 08: Clustering et groupement de données
- Chapitre 09: Réduction de la dimensionnalité de données
- Chapitre 10: Applications

# Préalables

- Connaissances de base en programmation.
- Connaissances de base en probabilités et statistiques.

## Quelques références:

- S. Rogers et M Girolami. A first Course in machine learning, CRC press, 2012.
- C. Bishop. Pattern Recognition and Machine learning. Springer 2006.
- R. Duda, P. Storck et D. Hart. Pattern Classification. Prentice Hall, 2002.

# **Motivation de l'apprentissage et concept du Big data.**

# Motivation

« We are drowning in information and starving in knowledge »

John Naisbitt,  
(auteur américain)

Nous entrons à l'aire du « **big data** ». Il existe, par exemple:

- 1 trillion ( $10^{12}$ ) de pages Web.
- 1 heure de vidéos chargée chaque seconde sur YouTube.
- $10^6$  transactions/heure chez Wal-Mart. La compagnie possède une base de données de 2.5 peta-bytes ( $10^{15}$ ) d'information.

# Motivation

- Les **expérimentations**, les **observations** et les **simulations** dans différents domaines (science, économie, médecine, Web, etc.) génèrent des grandes masses de données.
- Le **commerce électronique** (ex. Amazon, eBay, etc.), les **moteurs de recherche** (ex. Google, Yahoo!, Microsoft, etc.) et les **medias sociaux** (ex. Facebook, YouTube, Twitter, etc.) génèrent des quantités massives de données.
- **L'analyse de l'information** contenue dans ces données a déjà permis de réaliser des percées dans les domaines de la **génétique**, le **Web**, **l'économie**, la **sécurité**, etc.

# Motivation

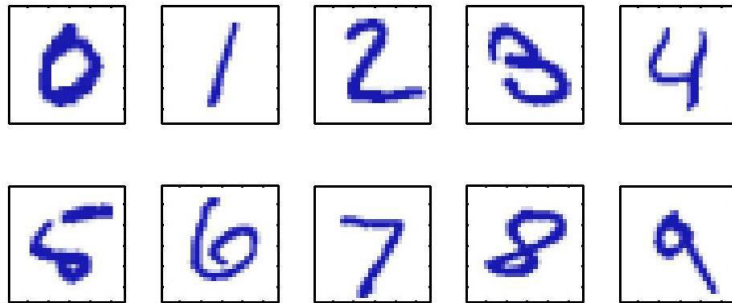
- **L'apprentissage automatique (machine Learning)** est une discipline émergente pour développer une intelligence artificielle capable:
  - *d'analyser automatiquement des données;*
  - *de détecter des motifs et associations de données;*
  - *d'utiliser des motifs pour la prédiction de données;*
  - *de prendre des décisions automatiques;*
- Cette matière permettra de connaître des outils développés dans cette discipline, avec des applications en **forage de données**, la **reconnaissance des formes**, la **vision artificielle**, etc.

# Apprentissage automatique et applications



# Applications

- **Exemple 1:** Reconnaissance des caractères manuscrits.



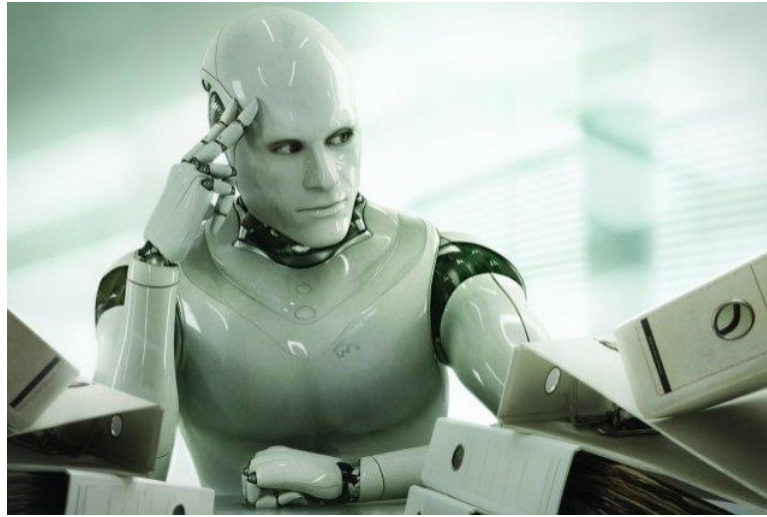
Images are 28 x 28 pixels

Représenter une image comme un vecteur  $X \in \mathbb{R}^{784}$ .

Apprendre un classificateur  $f(X): f(X) \rightarrow \{0,1,2,3,4,5,6,7,8,9\}$ . **Applications**

**Approche par règles (classique):** si la valeur du pixel (15,24) est supérieur a 50 et celle sur pixel a la position...., alors c'est un 8.

# Applications



## **Une meilleure approche:**

En donnant à l'ordinateur la capacité d'apprendre à partir de cas et des erreurs.

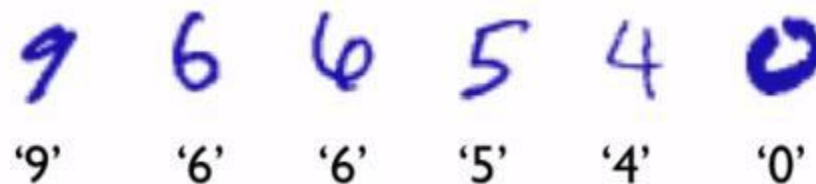
**=> L'apprentissage automatique (machine Learning)**

# Apprentissage automatique

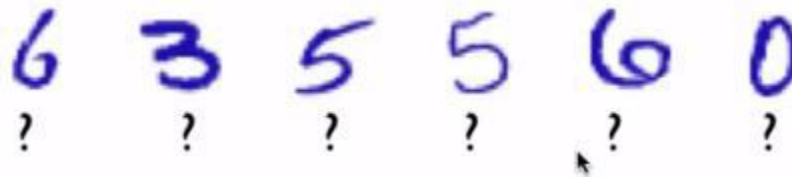
**Principe d'apprentissage:** données d'entrées vs. généralisation

Les algorithmes d'apprentissage précèdent comme suit:

- On fournit à l'algorithme des **données d'entraînement**:

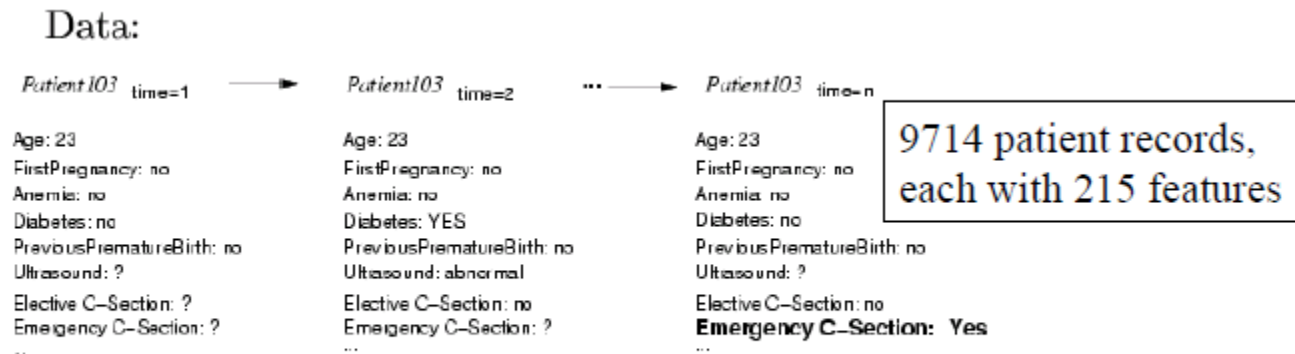


- Et l'algorithme retourne un «programme» capable de **se généraliser** à de nouvelles données



# Applications

- **Exemple 2: (biomédical)** prédiction les urgences de césariennes.



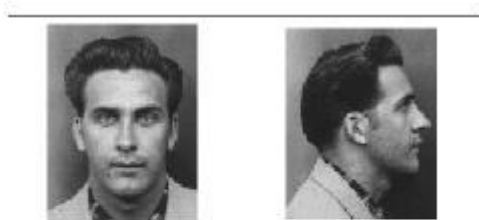
---

One of 18 learned rules:

**If (no previous birth delivery)  
and ( abnormal 2nd semester ultrasound)  
and (mal-presentation at admission)  
Prob. (emergency C-Section) = 0.6.**

# Applications

- **Exemple 3:** Détection de faces dans les images.

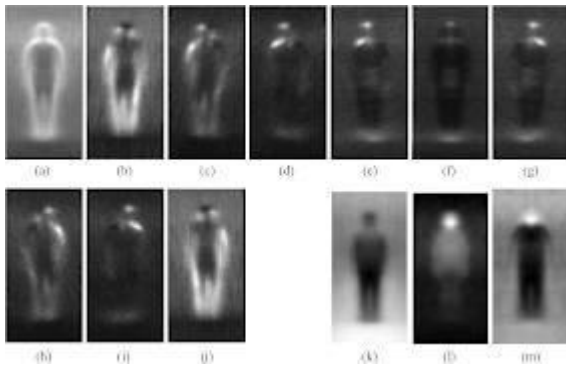


Example training images  
for each orientation



# Applications

- **Exemple 4:** Détection de piétons dans les images.



Example of training images



# Applications

- **Exemple 5:** classification de documents textuels.



Company home page

VS

Personal home page

VS

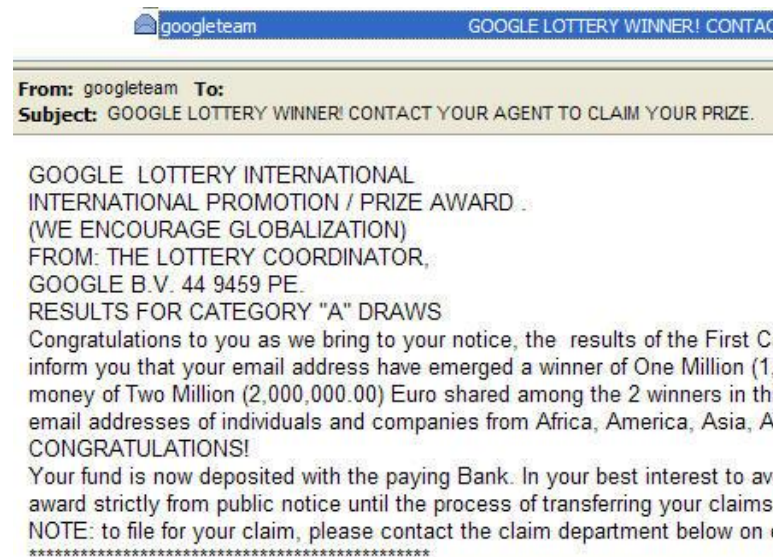
University home page

VS

...

# Applications

- **Exemple 6: Détection de spams dans les courriels.**

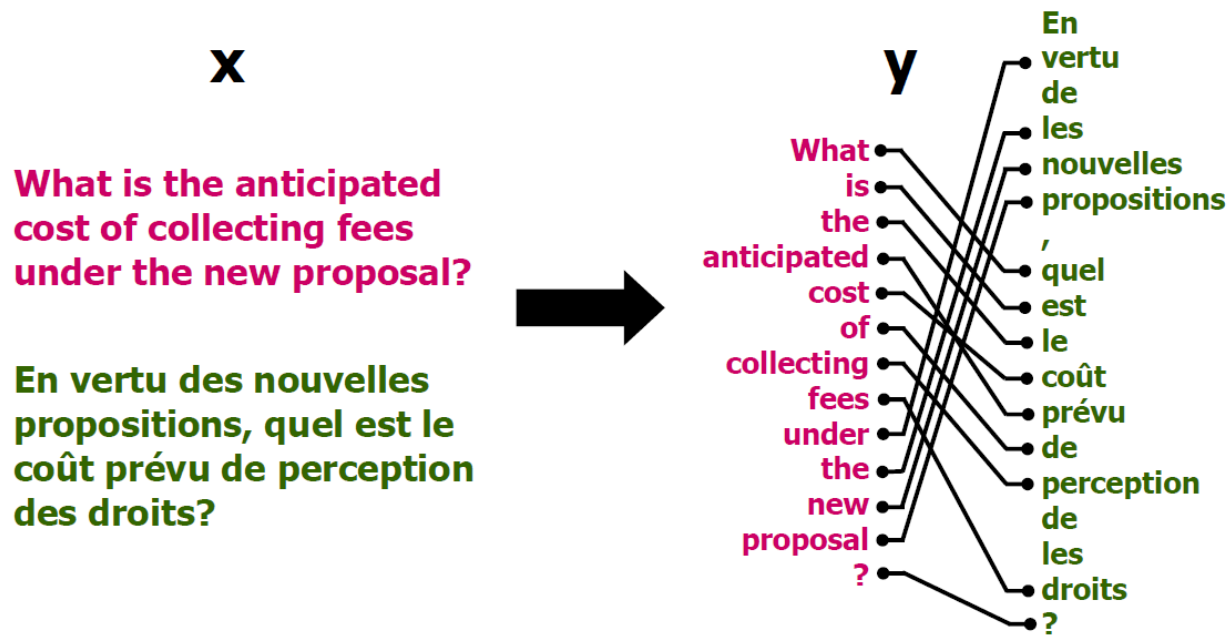


Compter le nombre de **mots clés**, ex. of achat, crédit, etc. • Requiert un **système d'apprentissage**, comme les spammeurs sont en perpétuelle innovation...



# Applications

- Exemple 7: Traduction automatique.



ex. Google translate

# Applications

- **Exemple 8:** Systèmes de recommandation.

## Frequently Bought Together

Customers buy this book with [Pattern Recognition and Machine Learning \(Information Science and Statistics\) \(Information Science and Statistics\)](#) by Christopher M. Bishop



Price For Both: **£104.95**

Add both to Basket

## Customers Who Bought This Item Also Bought

Page 1



[Pattern Recognition and Machine Learning \(Infor...](#)  
by Christopher M. Bishop  
★★★★☆ (4) £48.96

[Show related items](#)



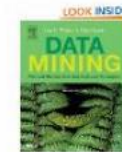
[MACHINE LEARNING \(Mcgraw-Hill International Edit\)](#)  
by Thom M. Mitchell  
★★★★★ (3) £42.74

[Show related items](#)



[Pattern Classification, Second Edition: 1 \(A Wi...](#)  
by Richard O. Duda  
★★★★★ (1) £78.38

[Show related items](#)



[Data Mining: Practical Machine Learning Tools a...](#)  
by Ian H. Witten  
★★★★★ (1) £37.04

[Show related items](#)

# Applications

- La pratique de l'apprentissage automatique

Data:

| PatientID name                | PatientID name-2              | PatientID name                |
|-------------------------------|-------------------------------|-------------------------------|
| Age 23                        | Age 23                        | Age 23                        |
| Profession no                 | Profession no                 | Profession no                 |
| Accident no                   | Accident no                   | Accident no                   |
| Diabetes no                   | Diabetes YES                  | Diabetes no                   |
| PreviousPrenatalUltrasound no | PreviousPrenatalUltrasound no | PreviousPrenatalUltrasound no |
| Ultrasound ?                  | Ultrasound observed           | Ultrasound ?                  |
| EmergencyC-Section-1          | EmergencyC-Section no         | EmergencyC-Section no         |
|                               | EmergencyC-Section-2          | EmergencyC-Section-Yes        |

One of 18 learned rules:

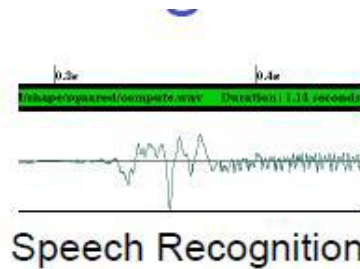
If No previous vaginal delivery, and Abnormal 2nd Trimester Ultrasound, and Malpresentation at admission  
Then Probability of Emergency C-Section is 0.6

Over training data: 26/41 = .63,  
Over test data: 12/20 = .60

Mining Databases

Text analysis

**Peter H. van Oppen**, Chairman of the Board & Chief Executive Officer  
 Mr. van Oppen has served as Chairman of the Board and Chief Executive Officer of ADIG since its acquisition by Interpoint in 1994 and a director of ADIG since 1996. Until its acquisition by Crane Co. in October 1996, Mr. van Oppen served as Chairman and Chief Executive Officer of ADIG. Prior to 1985, Mr. van Oppen worked as a consultant at Price Waterhouse LLP and at Bain & Company in Boston and London. He has additional experience in medical electronics and venture capital. Mr. van Oppen also serves as a director of Medtronic and Spacelabs Medical, Inc. He holds a B.A. from Whitman College and an M.B.A. from Harvard Business School, where he was a Baker Scholar.



Control learning



Object recognition

- Supervised learning
- Bayesian networks
- Hidden Markov models
- Unsupervised clustering
- Reinforcement learning
- ....

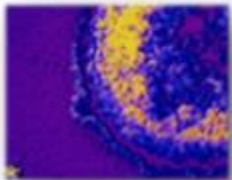
# Applications

- L'apprentissage automatique est actuellement l'approche préférée pour les domaines de:
  - **Parole:** reconnaissance de parole, personne.
  - **Vision artificielle:** reconnaissance d'objets, segmentation, etc.
  - **Robotique:** estimation de positions, de cartes, d'état, etc.
  - **Bio-informatique:** alignement de séquences, analyse de données génétiques.
  - **E-commerce:** commerce automatique, forage de données, spams.
  - **Analyse financière:** allocation de portfolio, crédits, bourses, etc.
  - **Médecine:** diagnostique, traitement, conception de thérapies.
  - **Graphisme:** conception et simulations réalistes.
  - **Web:** Gestion du contenu, entrepôts de données, réseaux sociaux, etc.

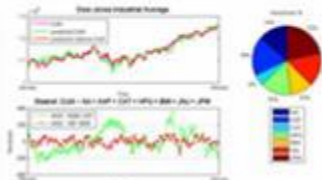
# Applications

- L'apprentissage est un domaine en expansion continue avec **l'acquisition des données** et **l'amélioration des capacités de calcul**.

**Biology**



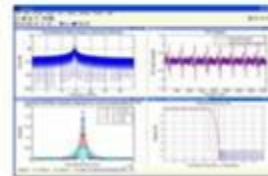
**Financial Services**



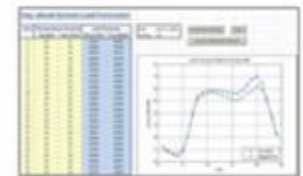
**Image & Video Processing**



**Audio Processing**



**Energy**



# Motivation

- **Pourquoi étudier l'apprentissage automatique?**
  - Création de systèmes adaptatifs.
  - Création d'une intelligence artificielle (IA).
  - Prédiction de nouvelles données, etc.
- **"If an expert-system brilliantly designed, engineered and implemented cannot learn not to repeat its mistakes, it is not as intelligent as a worm or a sea anemone or a kitten."**

**Oliver Selfridge 1926-2008 (pionnier de l'IA)**

# Types d'apprentissage

# Types d'apprentissage

Il existe trois types d'apprentissage:

- **Apprentissage supervisé:** Il y a une cible à prédire.

$$D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(N)}, y^{(N)})\}$$

- **Apprentissage non-supervisé:** cible n'est pas fournie.

$$D = \{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$$

- **Apprentissage par renforcement:** n'est pas couvert dans ce cours.



# Types d'apprentissage

**Apprentissage supervisé:** Classification, régression.

**Classification:** La cible est un indice de classe  $y \in \{1, \dots, K\}$ .

**Ex:** reconnaissance de caractères manuscrits.

$x$ : valeurs des intensités des pixels de l'image.

$y$ : identité du caractère (classe).

**Régression:** La cible est un nombre réel  $y \in \mathbb{R}$ .

**Ex:** Prédiction de la valeur d'une action de bourse.

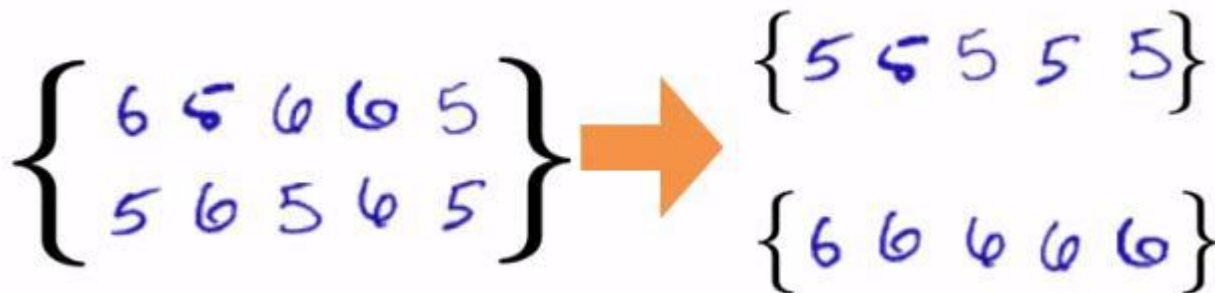
$x$ : Informations économiques de la journée.

$y$ : Valeur de l'Actions en bourse (nombre réel).

# Types d'apprentissage

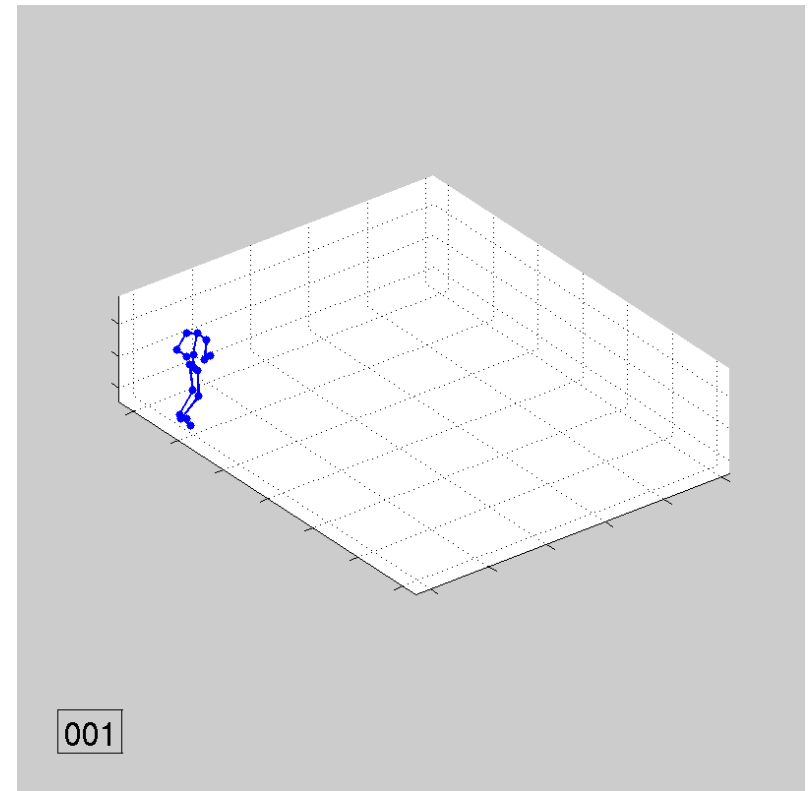
**Apprentissage non-supervisé:** lorsque la cible n'est pas explicitement connue.

- *Partitionnement de données / clustering*



# Types d'apprentissage

- **Estimation de densités:** apprendre la loi de probabilité ayant généré les données.
  - Pour générer de nouvelles données réalistes.
  - Pour distinguer les «vraies» données des «fausses» données (ex. spam filtering).
  - Compression de données.



# Références

1. M. S. Allili. Techniques d'apprentissage automatique (Cours de 2e cycle). Université du Québec en Outaouais (UQO), Québec, Canada. Hivers 2015.
2. S. Rogers et M Girolami. A first Course in machine learning, CRC press, 2012.
3. C. Bishop. Pattern Recognition and Machine learning. Springer 2006.
4. R. Duda, P. Storck et D. Hart. Pattern Classification. Prentice Hall, 2002.