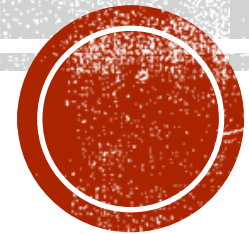


INTRODUCTION

APPRENTISSAGE AUTOMATIQUE

Hadjadj Abdelhalim



■ 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



MOTIVATIONS : L'ÈRE DU BIG DATA ET DE L'APPRENTISSAGE AUTOMATIQUE



CONTEXTE : L'EXPLOSION DES DONNÉES

- « Nous sommes noyés dans l'information et affamés de connaissances »

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.



SOURCES PRINCIPALES DES DONNÉES MASSIVES

- Dans divers secteurs tels que la science, l'économie, la médecine et le Web, une immense quantité de données est produite en continu grâce aux expérimentations, aux observations et aux simulations.
- Les plateformes numériques jouent un rôle clé dans cette explosion des données. Le commerce en ligne (Amazon, eBay), les moteurs de recherche (Google, Yahoo!, Bing) et les réseaux sociaux (Facebook, YouTube, Twitter) accumulent des volumes considérables d'informations générées par les interactions des utilisateurs.
- 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.



APPRENTISSAGE AUTOMATIQUE ET APPLICATIONS



APPRENTISSAGE AUTOMATIQUE

- L'apprentissage automatique (*machine learning*) est **une branche** de l'intelligence artificielle

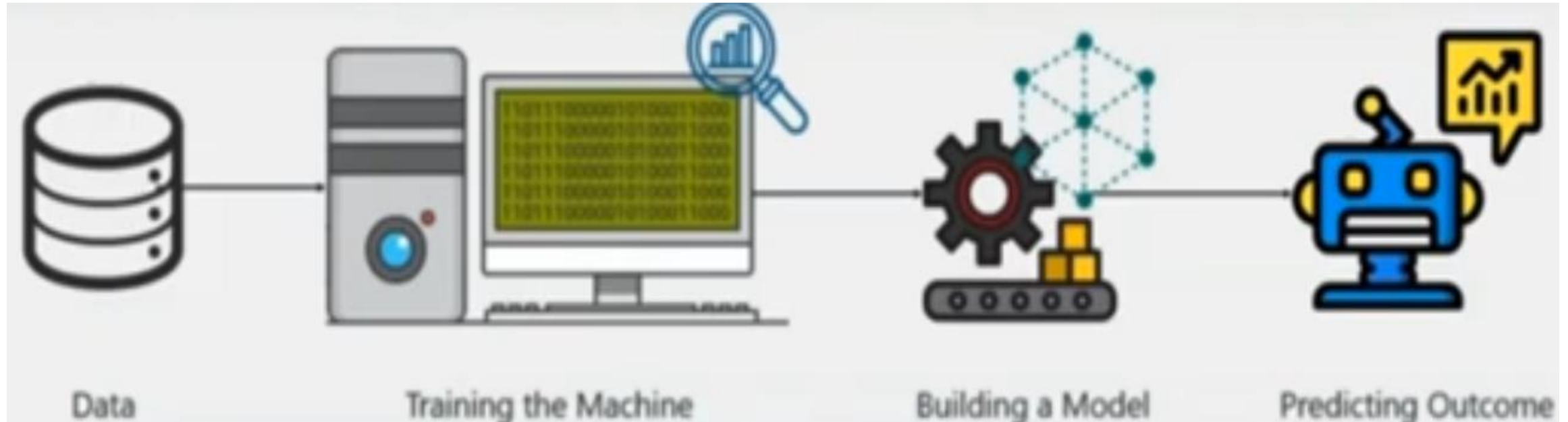


OBJECTIFS PRINCIPAUX :

- L'apprentissage automatique vise à permettre aux systèmes informatiques **d'apprendre à partir de données sans être explicitement programmés pour développer** une intelligence artificielle capable:
 1. Examiner et analyser de manière autonome de vastes ensembles de données
 2. Identifier des schémas récurrents et des relations entre les données ;
 3. Exploiter ces motifs pour prédire des événements futurs ou des comportements
 4. Automatiser certaines prises de décisions sur la base des analyses effectuées.



Apprentissage automatique



L'apprentissage automatique vise à permettre aux machines d'analyser les données afin de fournir une sortie précise. Il est préoccupé par les modèles et la précision



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:

9 6 6 5 4 0
'9' '6' '6' '5' '4' '0'

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

6 3 5 5 6 0
? ? ? ? ? ?



TYPE D'APPRENTISSAGE

 Supervisé	 Non supervisé	 Renforcement
Il y a une cible à prédire. $D = \{x(1), y(1), x(2), y(2), \dots, x(N), y(N)\}$	cible n'est pas fournie. $D = \{x(1), x(2), \dots, x(N)\}$	n'est pas couvert dans ce cours.



TYPE D'APPRENTISSAGE

Apprentissage
supervisé



, CHAT



, CANARD



, CHIEN

Algorithme
d'apprentissage

Classifieur

- Chat
- Canard
- Chien

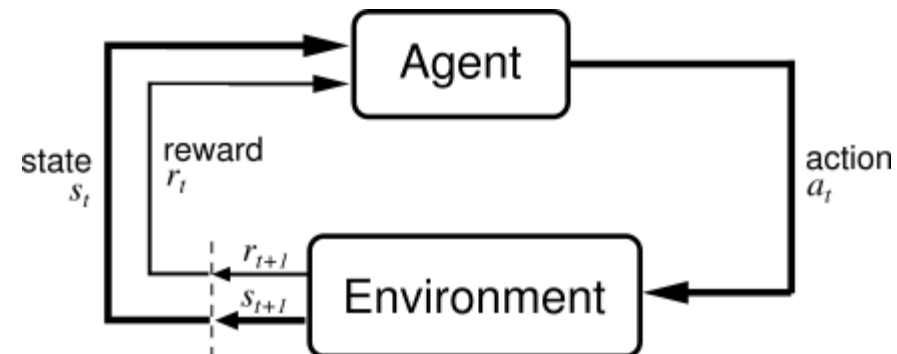
Apprentissage
non supervisé



Algorithme
d'apprentissage

- Groupe 1 (chats)
- Groupe 2 (canards)
- Groupe 3 (chiens)

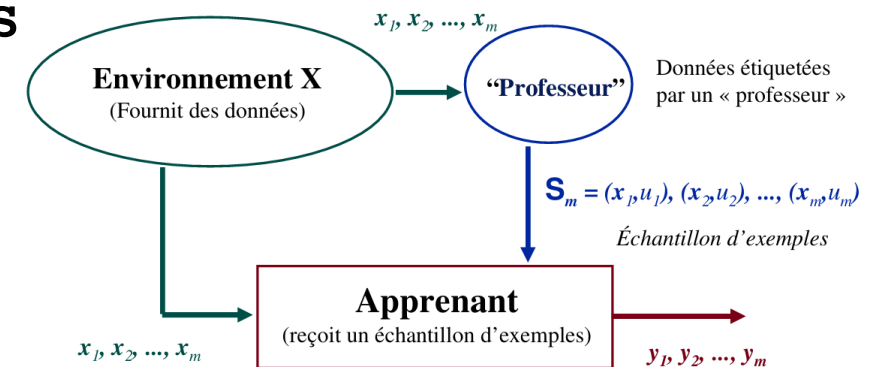
Apprentissage
par renforcement



APPRENTISSAGE SUPERVISE

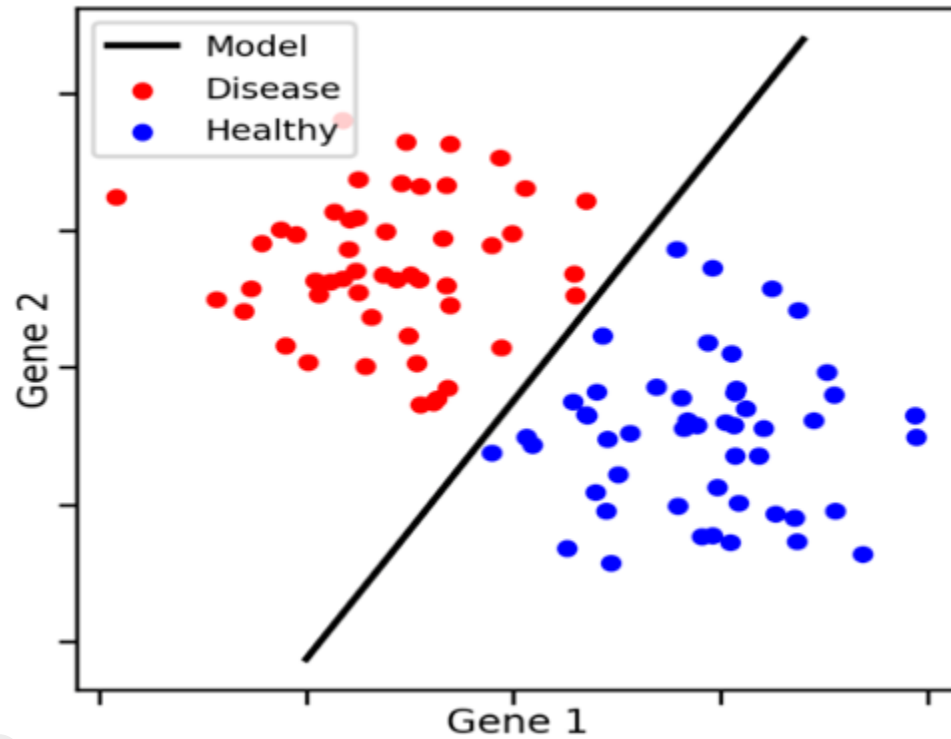
Apprentissage supervisé: Classification, régression.

- **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 bours
 - X: Informations économiques de la journée.
 - Y: Valeur de l'Actions en bourse (nombre réel).

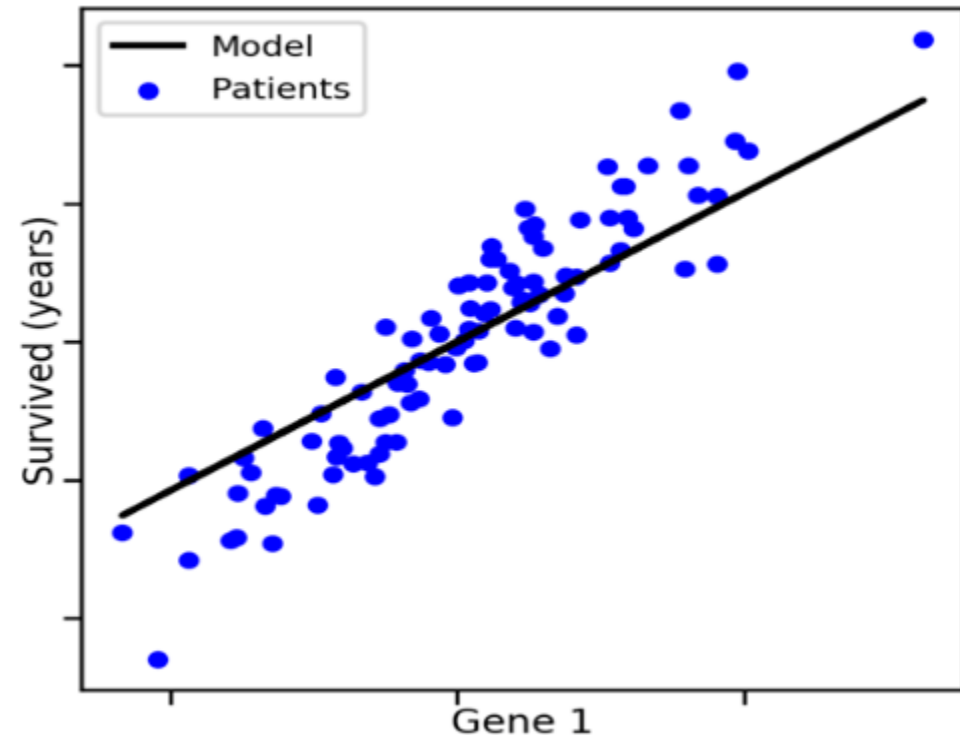


APPRENTISSAGE SUPERVISE

Classification



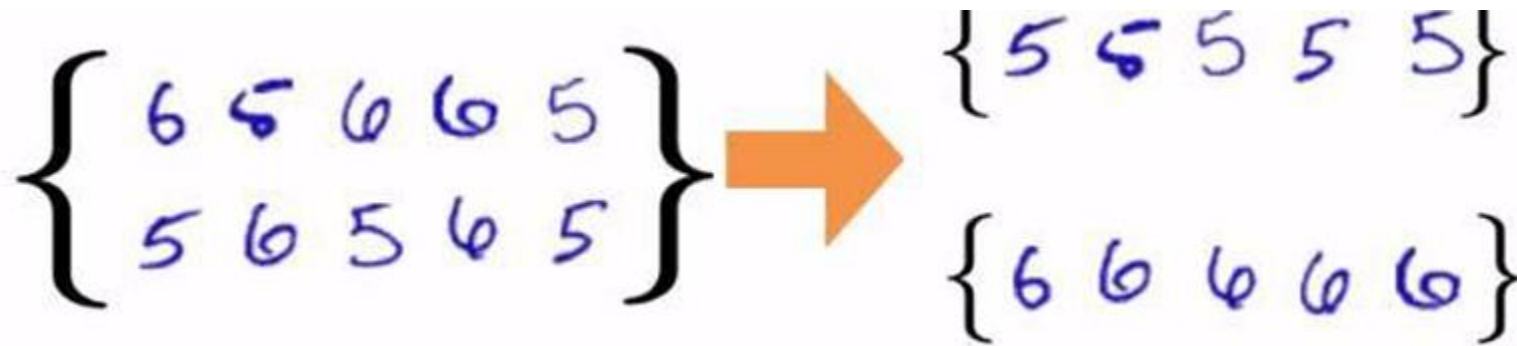
Regression



APPRENTISSAGE NON-SUPERVISE

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

- **Partitionnement de données / clustering**



APPRENTISSAGE NON-SUPERVISE

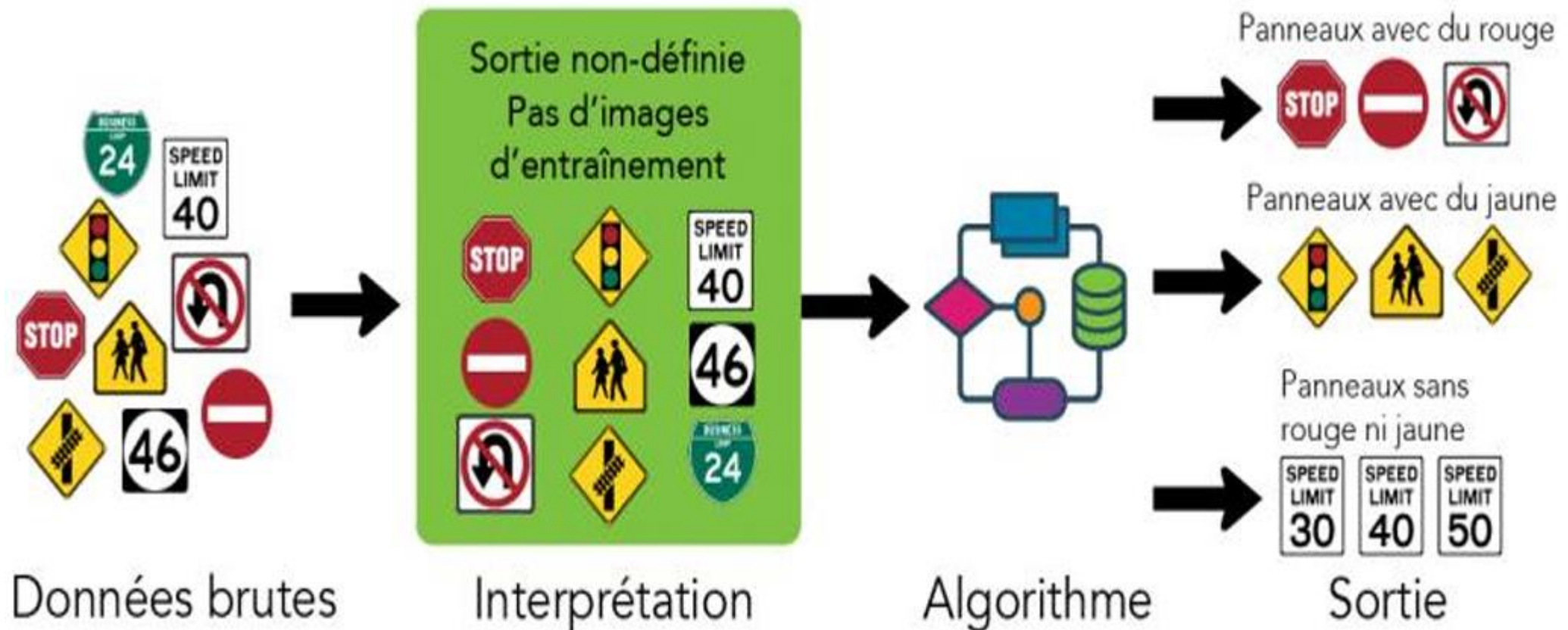
Estimation de densités:

- 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.



APPRENTISSAGE NON-SUPERVISE

EXEMPLE



APPRENTISSAGE PAR RENFORCEMENT



POURQUOI ÉTUDIER L'APPRENTISSAGE AUTOMATIQUE?

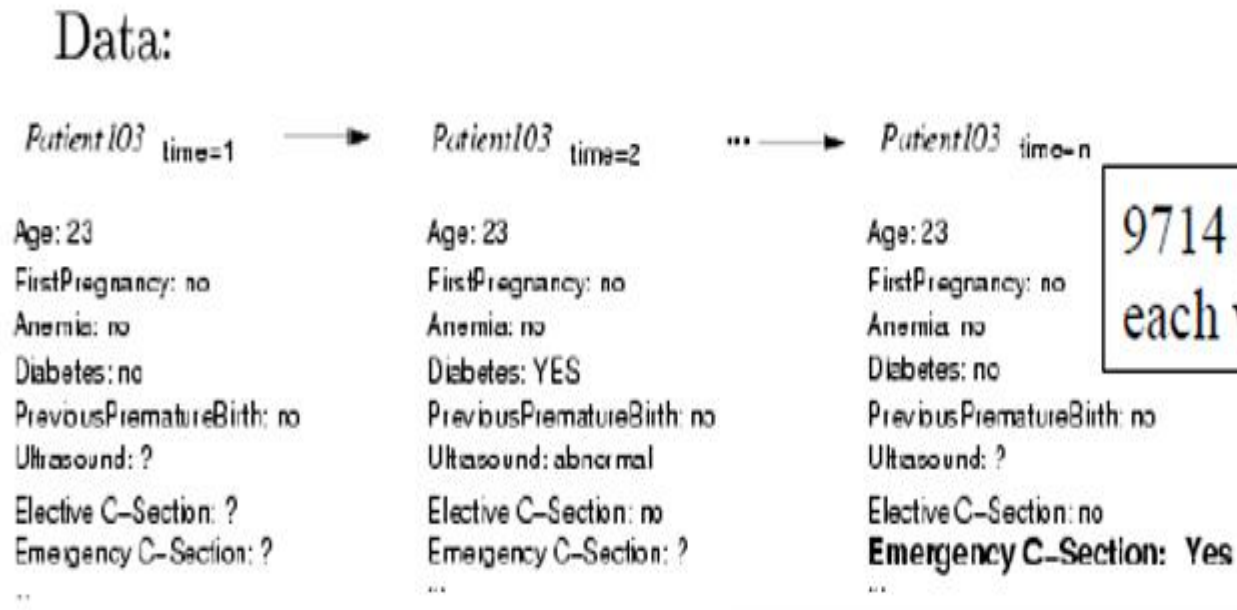
- **Développement de systèmes capables de s'adapter aux changements,**
- **Conception d'une intelligence artificielle autonome et évolutive**
- **Anticipation et prédiction basées sur l'analyse des données**



EXEMPLES CONCRETS

EXEMPLE 1

(biomédical)
prédiction les
urgences
de césariennes.



9714 patient records,
each with 215 features

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.



EXEMPLES CONCRETS

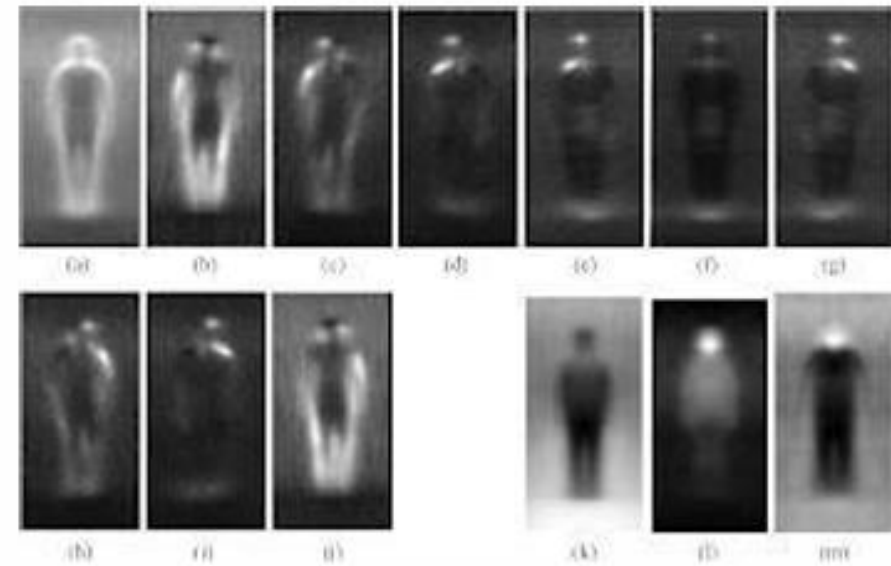
EXEMPLE 2



EXEMPLES CONCRETS

EXEMPLE 3

- **Détection de piétons dans les images.**



Example of training images



EXEMPLES CONCRETS

EXEMPLE 4

classification de documents textuels.



- PERSONNEL
- UNIVERSITAIRE
- , , , ,



EXEMPLES CONCRETS

EXEMPLE 5

Détection de spams dans les courriels.



EXEMPLES CONCRETS

EXEMPLE 6

Traduction automatique.

X

What is the anticipated
cost of collecting fees
under the new proposal?

En vertu des nouvelles
propositions, quel est le
coût prévu de perception
des droits?



Y

What
is
the
anticipated
cost
of
collecting
fees
under
the
new
proposal
?

En
vertu
de
les
nouvelles
propositions
/
quel
est
le
coût
prévu
de
perception
de
les
droits
?



EXEMPLES CONCRETS

EXEMPLE 7

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

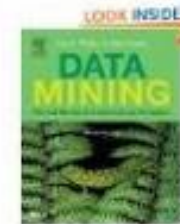
[+ Show related items](#)



[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](#)



EXAMPLES CONCRETS

EXAMPLE 8

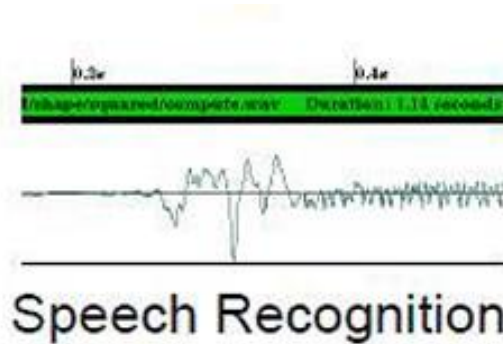
Data:		
Patient13 (age=1)	Patient20 (age=2)	Patient22 (age=3)
Age: 28	Age: 28	Age: 28
PrePregnancy Wt: 60	PrePregnancy Wt: 60	PrePregnancy Wt: 60
Arrested: 0	Arrested: 0	Arrested: 0
Delivered: 0	Delivered: YES	Delivered: 0
PreviousPrenatalUltrasound: 0	PreviousPrenatalUltrasound: 0	PreviousPrenatalUltrasound: 0
Ultrasound: ?	Ultrasound: abnormal	Ultrasound: ?
DeliveryC-Section: 1	DeliveryC-Section: 0	DeliveryC-Section: 0
EmergencyC-Section: 1	EmergencyC-Section: ?	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



Control learning

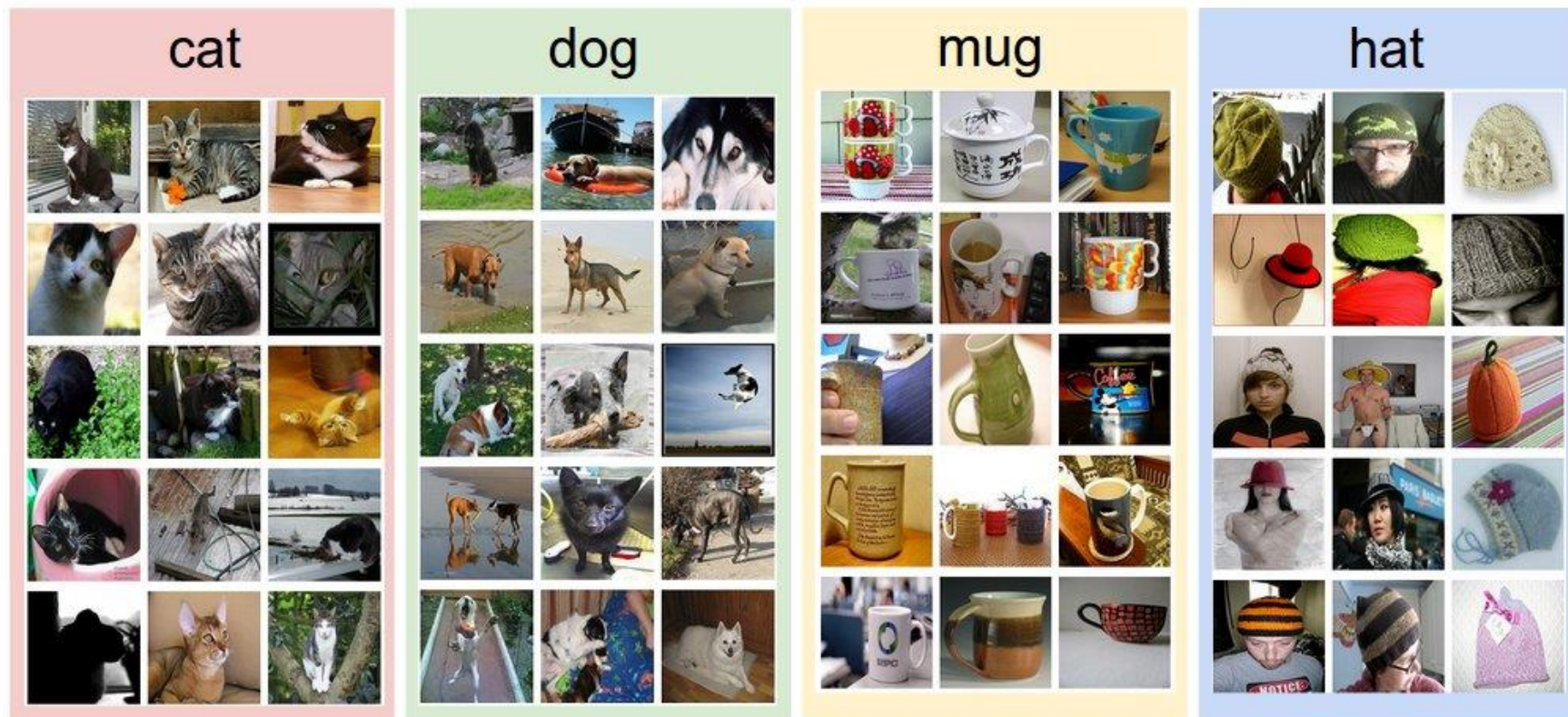
Text analysis

Peter H. van Oppen, Chairman of the Board & Chief Executive Officer. Mr. van Oppen has served as Chairman of the Board since its acquisition by Interpoint in 1994 and a director of ADIC since 1985. Until its acquisition by Crane Co. in October 1995, Mr. van Oppen served as President of the company. Prior to 1985, Mr. van Oppen worked as a consulting engineer 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 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.



EXEMPLES CONCRETS

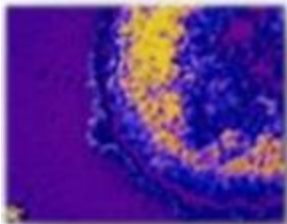
EXAMPLE 8



DOMAINE D'APPLICATION

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

Biology



Financial Services

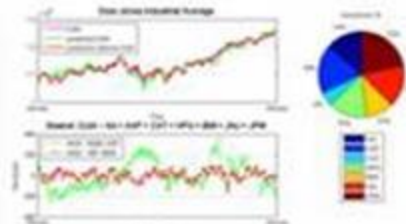
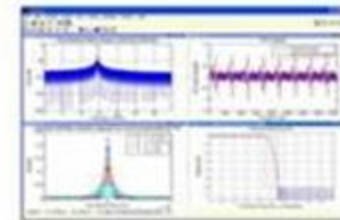


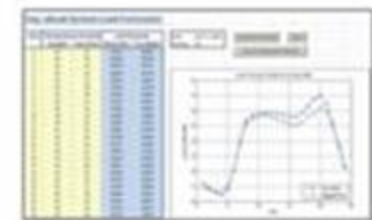
Image & Video Processing



Audio Processing



Energy



DOMAINE D'APPLICATION

L'apprentissage automatique : une approche privilégiée dans divers domaines

- **Vision par ordinateur** : Détection et classification d'objets, segmentation d'images.
- **Graphisme** : Création d'images et de simulations réalistes
- **Robotique** : Estimation de la position, modélisation des environnements, analyse des états.
- **Bio-informatique** : Alignement de séquences ADN, analyse des données génétiques.
- **E-commerce** : Automatisation des recommandations, analyse des données clients, détection de spams.
- **Web et réseaux sociaux** : Gestion du contenu, analyse des interactions, stockage et traitement des données massives.
- **Reconnaissance vocale** : Identification et transcription de la parole, reconnaissance des locuteurs.
- **Finance** : Optimisation de portefeuilles, évaluation des risques de crédit, prédiction des marchés boursiers.
- **Médecine** : Diagnostic assisté, personnalisation des traitements, conception de nouvelles thérapies.

