

IFT3395/6390

# Fondements de l'apprentissage machine

Premier cours:

## INTRODUCTION

Professeur: Pascal Vincent

Institut  
des algorithmes  
d'apprentissage  
de Montréal



# Au programme aujourd’hui

- ⦿ Faire connaissance...
- ⦿ Page Web du cours: sur STUDIUM (Moodle)  
<https://studium.umontreal.ca>
- ⦿ Présentation des objectifs, du plan de cours,  
des modalités d'évaluation, ...
- ⦿ Présentation informelle du domaine de  
l'apprentissage automatique.

# Cours d'apprentissage machine offerts au DIRO

Cours d'introduction: survol large du domaine couvrant de nombreuses approches

## **IFT3395/6390 Fondements de l'apprentissage machine**

Automne

- Offert aux cycles supérieurs et en 3ème année de bac,

Cours avancé présentant un formalisme et des techniques essentielles pour l'apprentissage

## **IFT6269 Modèles graphiques probabilistes et apprentissage**

Automne

Cours avancé présentant la fine pointe de la recherche en réseaux de neurones artificiels et apprentissage profond

## **IFT6266 Algorithmes d'Apprentissage**

## **/ Apprentissage de représentations**

Hiver

## PREMIÈRE PARTIE

Plan de Cours  
et autres informations pratiques

Les informations de plan de cours se trouvent sur

La page Web du cours qui est sur StudiUM:  
<https://studium.umontreal.ca>

Si vous êtes inscrits, dans le menu «mes cours»  
vous devriez voir:

**IFT3395-6390-A-A16**

Fondements de l'apprentissage machine

Sinon, dites-le moi...

# Tonalité du cours

Le cours nécessite d'être assez à l'aise  
en **mathématiques** & **informatique**

- Algèbre linéaire  
(vecteurs, matrices, ...)
- Probabilité et statistique  
(variable aléatoire, distribution, espérance, ...)
- Analyse  
(dérivées partielles...)
- Algorithmes  
(et complexité)
- Structures de données
- Programmation  
(environnement Python +numpy+matplotlib)

Il y aura des rappels pour toutes les notions mathématiques essentielles...

# Tonalité du cours

- Matériel pédagogique provient de sources diverses (attention aux notations mathématiques!).
- Le cours sera donné en Français, mais une partie du matériel pourra être en Anglais (ressources et matériel parfois utilisé comme support de cours).

## DEUXIÈME PARTIE

Présentation informelle du domaine  
de l'apprentissage automatique

# Au programme

- ⦿ Place de l'apprentissage en Intelligence Artificielle.
- ⦿ Les disciplines fondatrices de l'apprentissage.
- ⦿ Les domaines d'application de l'apprentissage.
- ⦿ Ex. de types de problèmes en apprentissage.
- ⦿ La représentation des données.

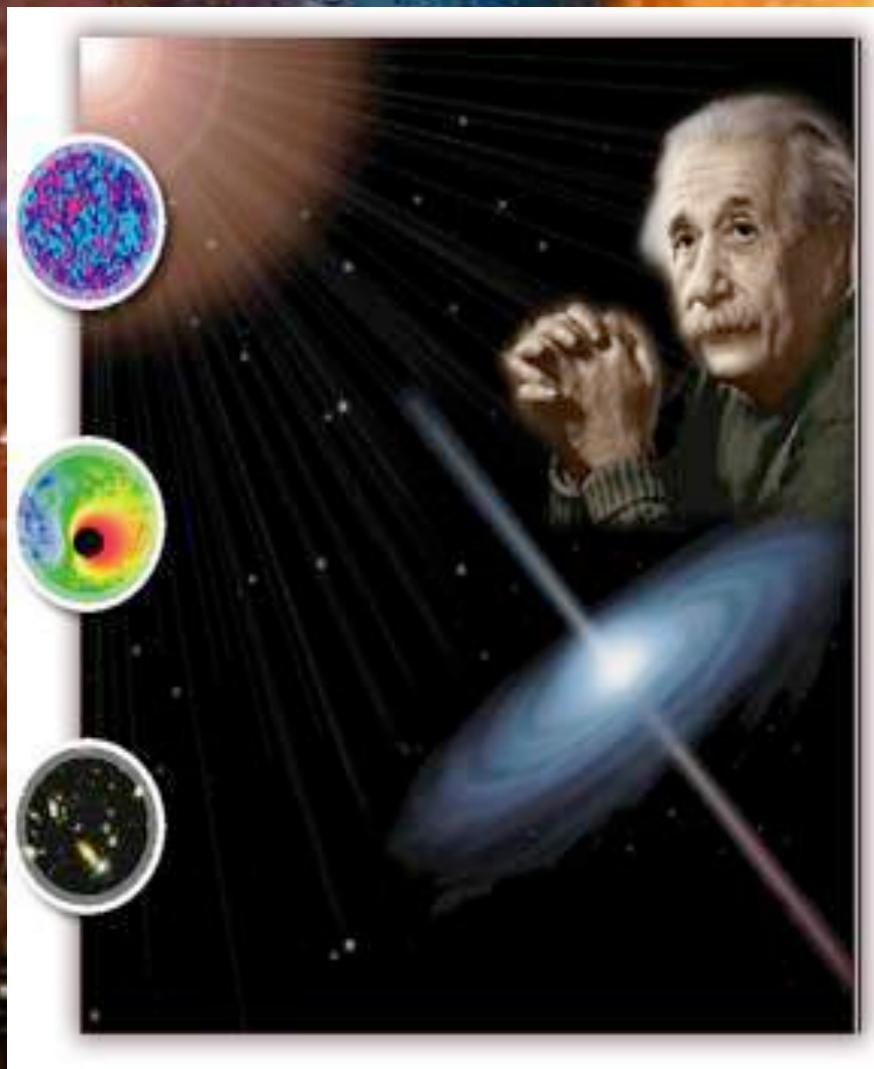
# Curiosité scientifique

# Trois grands mystères

L'univers,  
espace / temps  
énergie / matière

La Vie

L'intelligence  
La conscience



# De l'intelligence naturelle: un cerveau qui apprend, qui s'adapte

- $10^{11}$  neurones,  
 $10^{14}$  synapses
- Complexe réseau de  
neurones
- Apprentissage: modifier  
synapses



# À l'intelligence artificielle...

David is 11 years old.

He weighs 60 pounds.

He is 4 feet, 6 inches tall.

He has brown hair.

His love is real.

But he is not.



A STEVEN SPIELBERG FILM  
**ARTIFICIAL INTELLIGENCE**

WARNER BROS. PICTURES and DREAMWORKS PICTURES

in AMBLIN/STANLEY KUBRICK produced by STEVEN SPIELBERG as AL. ARTIFICIAL INTELLIGENCE HALEY JOEL OSMENT JUDE LAW FRANCES O'CONNOR BRENDAN GLEESON WILLIAM HURT Production Design by STAN WINSTON STUDIO Special Visual Effects & Animation by INDUSTRIAL LIGHT & MAGIC Camera Design BOB RINGWOOD Music by JOHN WILLIAMS Cinematographer MICHAEL KAHN, A.C.E. Production Designer RICK CARTER Director of Photography JANUSZ KAMINSKI, A.S.C. Executive Producers JAN HARLAN WALTER F. PARKES Associate Producer STEVEN SPIELBERG Based on a Screen Story by IAN WATSON Story by BRIAN ALDISS Produced by KATHLEEN KENNEDY STEVEN SPIELBERG BONNIE CURTIS

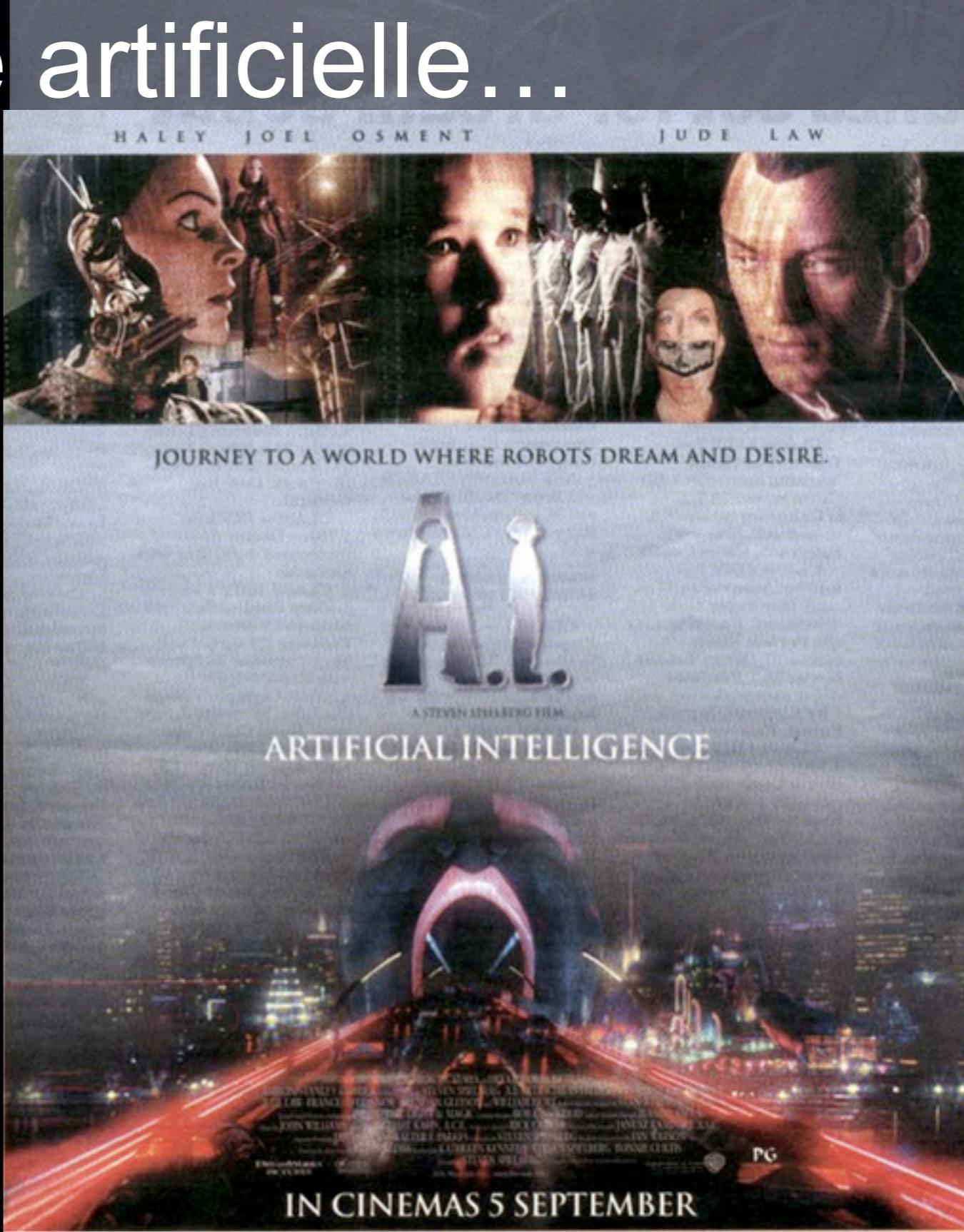
DREAMWORKS  
PICTURES



Directed by STEVEN SPIELBERG

SUMMER 2001

AOL Keyword: A.I. [www.AIMovie.com](http://www.AIMovie.com)



# Origines de l'apprentissage machine

## Perspective historique

- Né d'un objectif ambitieux:  
**l'Intelligence Artificielle**

- Projet fondateur:  
Le **Perceptron** (Frank Rosenblatt 1957)  
Premier neurone artificiel **apprenant** à partir  
d'exemples

- Deux approches historiquement opposées en IA:

**Inspirée du cerveau:**

- ⇒ réseaux de neurones  
apprenant à partir d'exemples
- ⇒ perception artificielle.

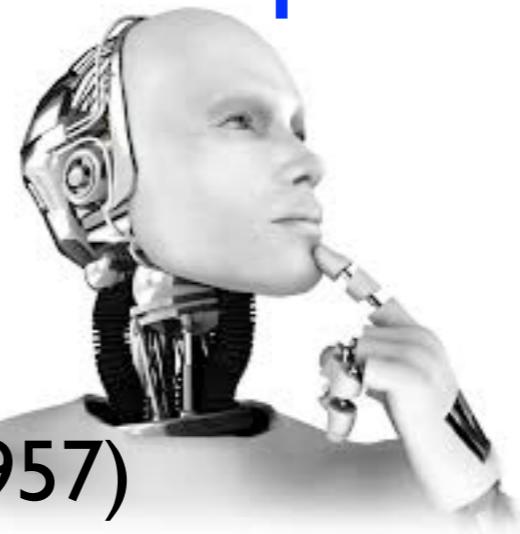
**IA « classique » symbolique:**

- Centré sur capacité de **raisonnement logique**
- ⇒ Pas d'apprentissage (règles codées à la main)
- ⇒ pas de gestion de l'incertain

**Finalement rajoutés (Bayes Nets...)**

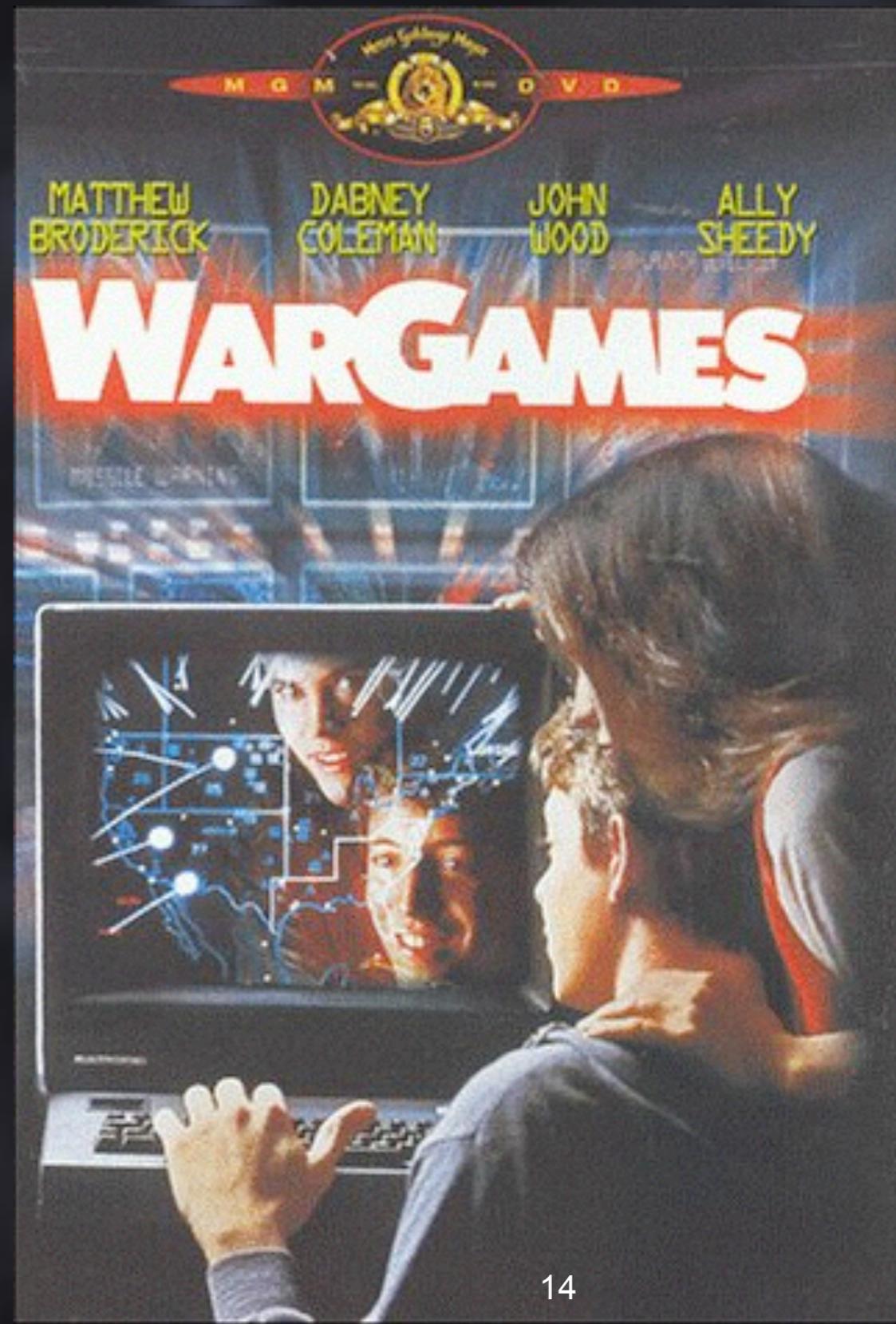
**L'apprentissage et les modèles probabiliste ont largement gagné**

- ⇒ **apprentissage machine / machine learning**



# I.A. de la science-fiction...

- 1983: dans **WarGames**, un ordinateur apprend en jouant contre lui-même à tic-tac-toe et “global thermonuclear war”.



# à la réalité...

## Au Backgammon

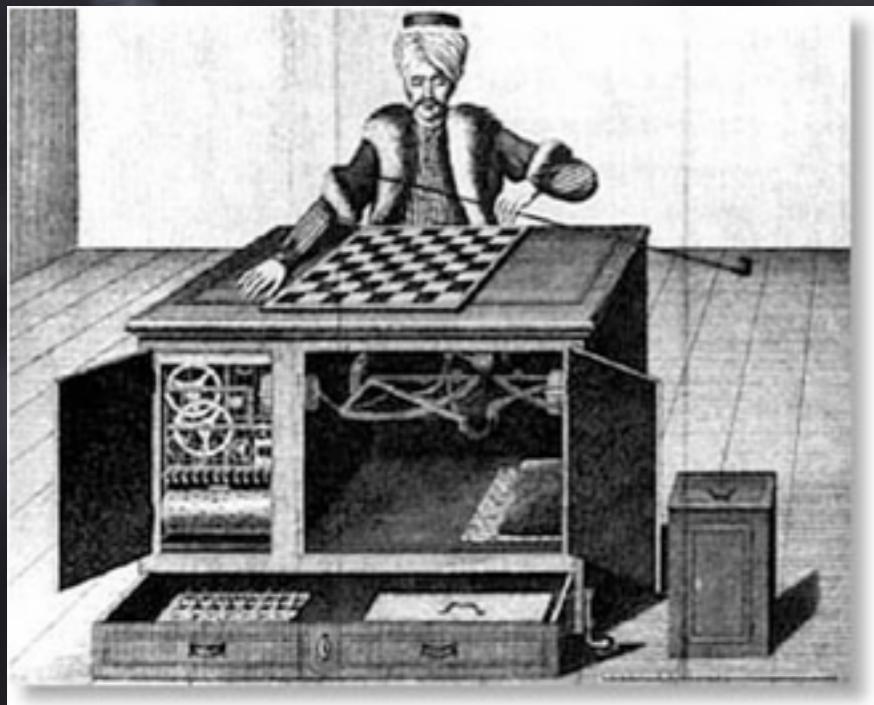
- 1995: TD-gammon, un réseau de neurones artificiels entraîné en jouant 200 000 parties de backgammon contre lui-même, joue à un niveau équivalent aux meilleurs joueurs mondiaux (Tesauro 1995).



# Aux échecs

## (Exemple de succès d'une approche d'IA « classique »)

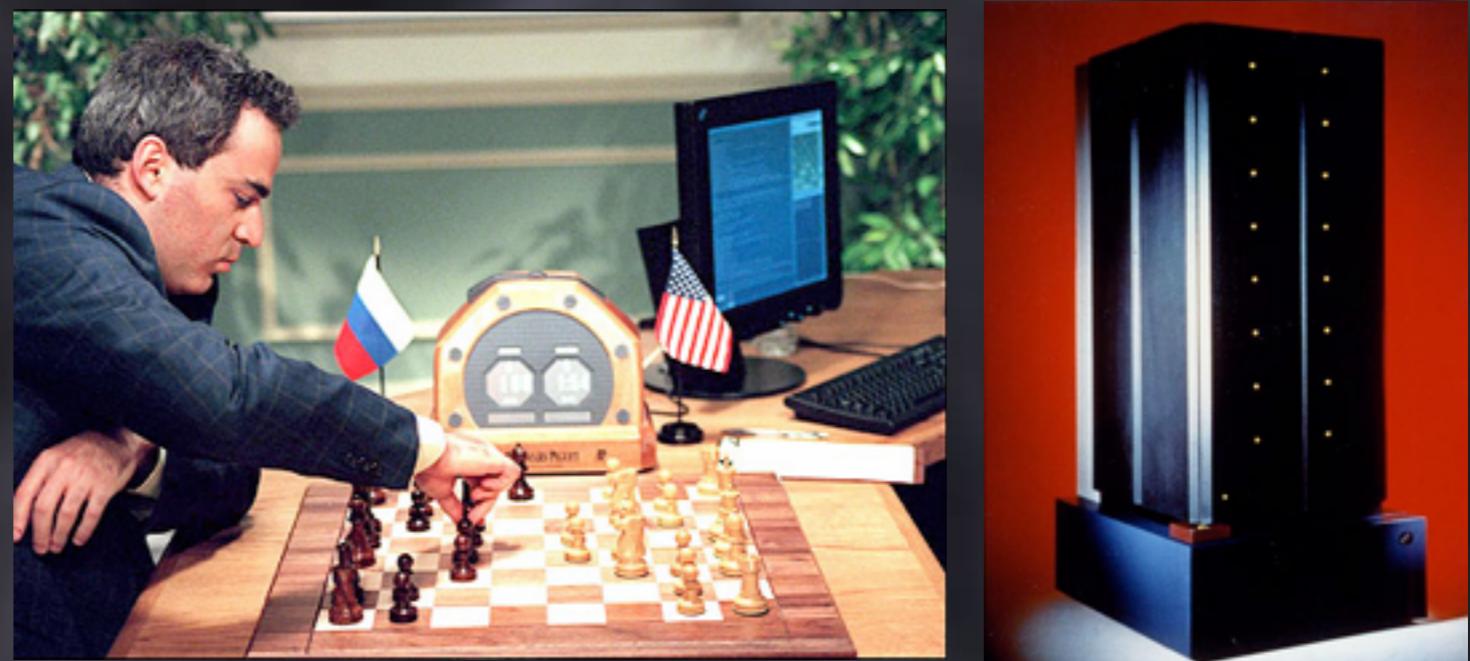
# 1770: «Le Turc mécanique» automate joueur d'échec



# A gagné contre Napoléon Bonaparte et Benjamin Franklin

# Un canular!

## 1997: Garry Kasparov contre «Deep Blue» d'IBM



May 11th, 1997  
**Computer won world champion of chess**  
(Deep Blue) (Garry Kasparov)



(Reuters = Kyodo News)

# À Geopardy

- Février 2011:  
Watson,  
système d'IBM,  
bat les  
champions  
humains de  
**Geopardy.**



- Utilise l'apprentissage automatique à partir de données textuelles.

# Mars 2016



Le GO: un des rares jeux de table où l'être humain avait encore une large avance.  
Plus maintenant!

# L'apprentissage est au cœur des succès modernes de l'IA

Et il n'y a pas que les jeux

- Google search
- Systèmes de vision artificielle
- Reconnaissance de la parole (ex: Siri)
- Recommandations de produits:  
Netflix / Amazon / ...
- Voitures autonomes
- Robots autonomes, etc...



# Applications

- Applications “traditionnelles”: reconnaissance des formes
  - écriture, parole, empreinte digitale. etc.
  - les experts humains les font bien
  - nombre de données, nombre d’attributs, nombre de classes modérés
  - bruit/ambiguïté modéré

# Applications

- Applications “modernes”:
  - forage de données, forage de texte, prédiction financière, ranger des “hits”(Google), analyse des expressions génétiques
  - les experts humains n'existent pas
  - nombre de données, nombre d'attributs, nombre de classes énormes
  - bruit/ambiguïté augmenté

# Applications

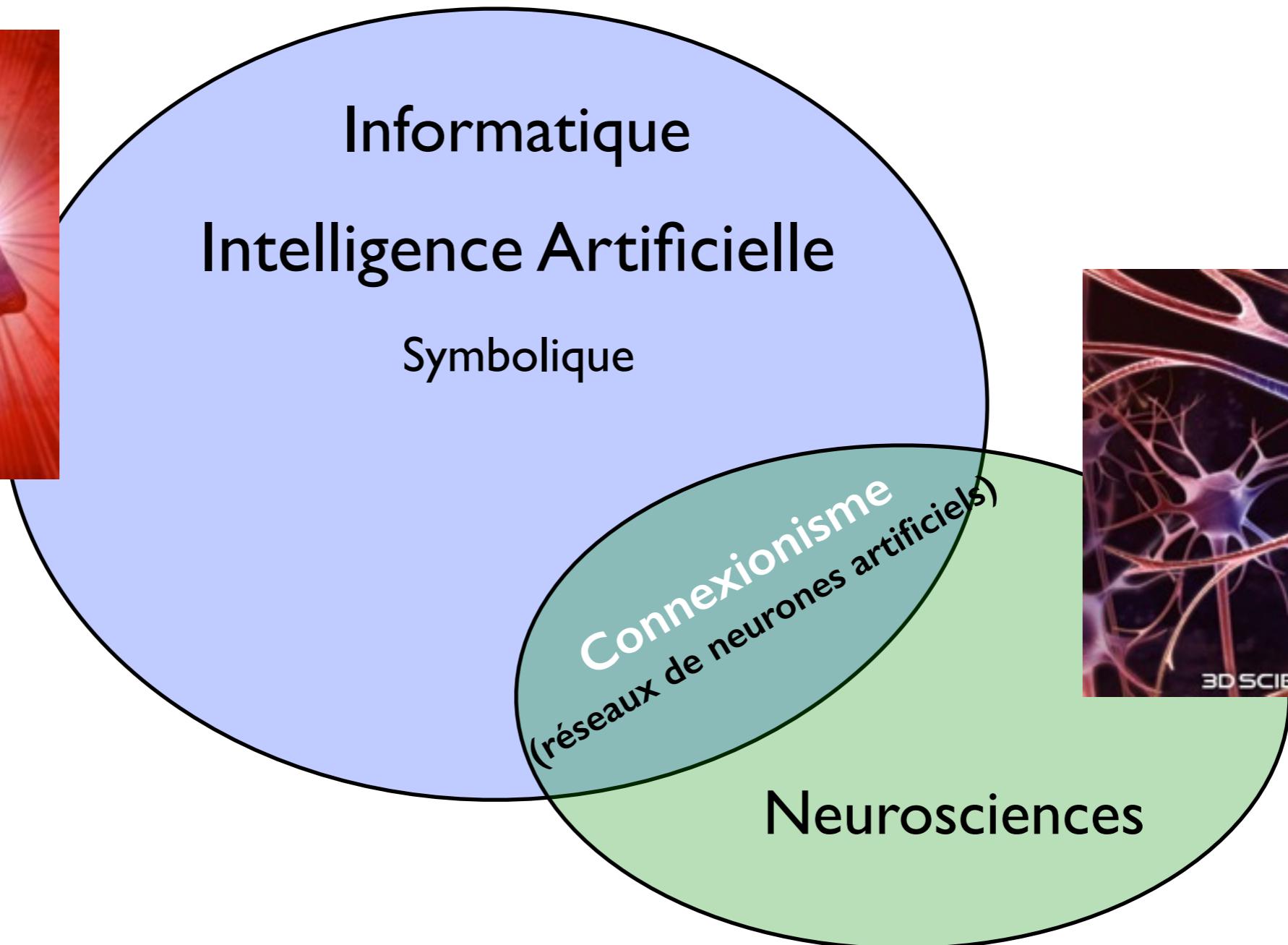
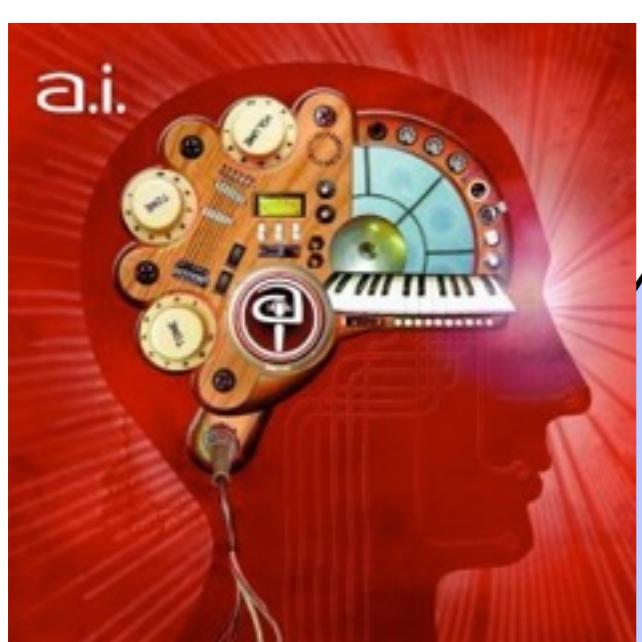
Applica

- Reconnaissance des formes
  - écriture
  - visages (identité et émotions)
- parole
- empreinte digitale
- Forage de texte
  - Google
  - classification de textes
- Traitement de langage
  - prédiction du prochain mot
  - désambiguïsation du sens
  - prédiction du POS  
(*Part-Of-Speech*)
- Génie logiciel
  - prédiction de la stabilité

# Applications

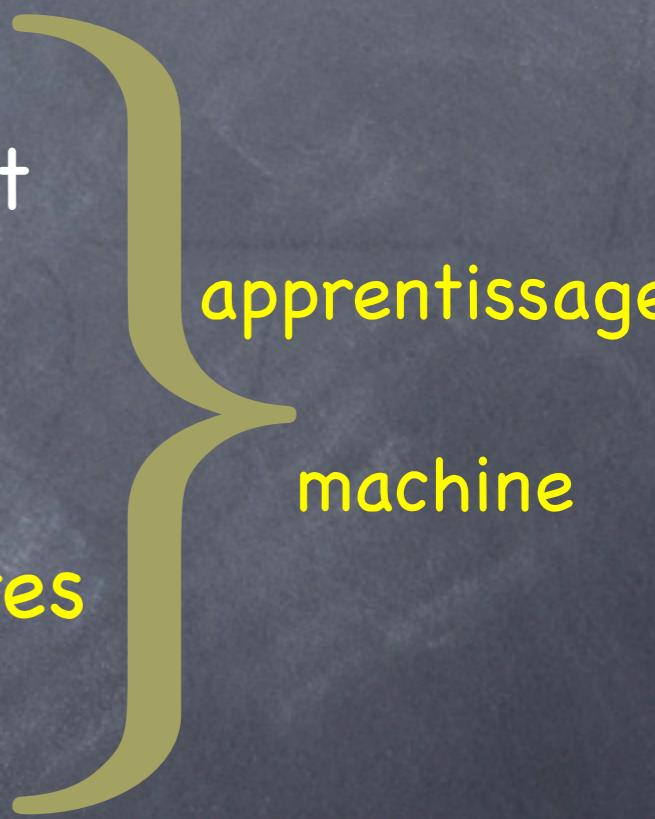
- Forage de données
  - marketing direct
  - prédition des primes
- Bioinformatique
  - prédition du risque de cancer
  - détection de cancer
  - découverte de médicament
- Filtrage collaboratif
  - Recommandation de produits à des usagers:  
Netflix, Amazon, iTunes, ...

# Vision de l'Intelligence Artificielle en 1957 (Rosenblatt, "Perceptron")

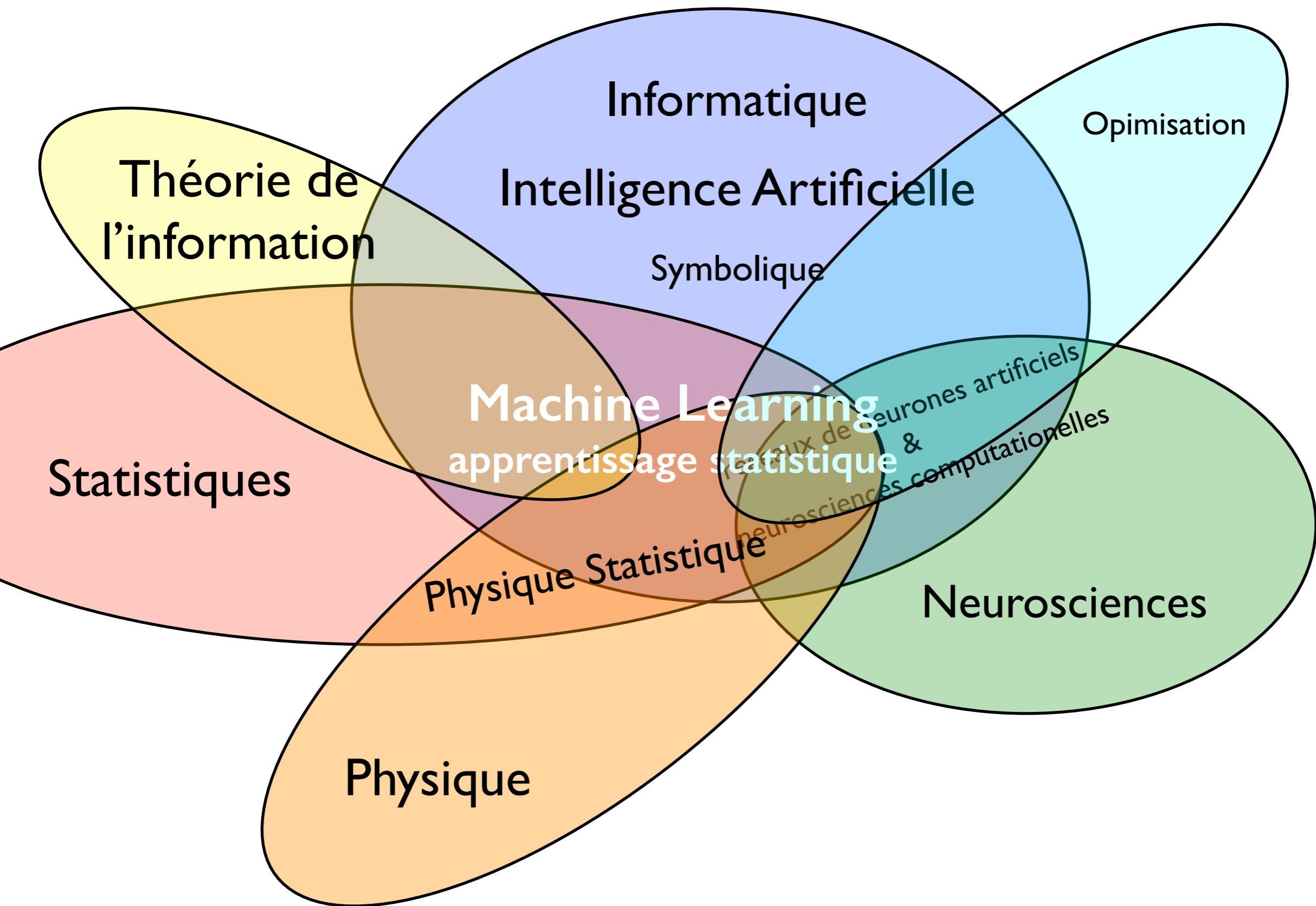


3D SCIENCE.com

# Place de l'apprentissage en I.A. en 2016

- L'I.A. "connexioniste" a mûri, s'est mathématisée, a engendré l'**apprentissage machine** dont les réseaux neuronaux ne sont qu'une sous-partie.
  - L'I.A. classique, en intégrant l'incertain, a engendré les **modèles graphiques probabilistes** (réseaux Bayesiens), dont les paramètres peuvent être appris.
  - Le rôle fondamental de l'**apprentissage** et d'une **approche probabiliste** est largement reconnu.
- 

# Vision actuelle des disciplines fondatrices



# Qu'est-ce que l'apprentissage machine?



Perspective d'un utilisateur (hypnotisé)

- Un champ d'étude scientifique (=sorcellerie) qui
- recherche les principes fondamentaux (des formules)
- et développe des **algorithmes** (incantations) magiques
- capables de d'utiliser des données collectées pour (automagiquement) produire des fonctions prédictives applicables à des données similaires (dans l'avenir!)

# L'ingrédient de base de l'apprentissage machine c'est...

- Collectées de la nature, de l'internet, ou de processus industriels.
- Arrivent stockées sous de nombreuses formes (et formats...), structuré / non structuré, **rarement propres, souvent désordonnées.**
- En apprentissage on aime voir nos données comme une **liste d'exemples** (ou on va les **transformer** en cela!)  

- idéalement **beaucoup d'exemples** de **même nature**.  
→ préférablement avec chaque exemple **un vecteur de nombres** (ou on va les **transformer** en cela!)



# Apprendre à partir d'exemples !



“cheval”



“cheval”



“cheval”

Principe beaucoup plus général que d'écrire à la main, en partant de zéro, un algorithme pour reconnaître un cheval...

Vous savez programmer: comment feriez-vous ?

# Algorithmes “classiques” vs. d'apprentissage

- Approche classique

- **description formelle** des contraintes de l'entrée et de la sortie souhaitée
- **compréhension du problème** computationnel
- design d'une solution algorithmique **basée sur ces connaissances**

- Problèmes

- **Connaissances incomplètes**
- Algorithme trop **coûteux**

# Algorithmes “classiques” vs. d'apprentissage

- Approche d'apprentissage
  - données (exemples) de forme (**entrée,sortie**)
  - compréhension **partielle** du problème: connaissances a-priori
  - apprendre: **chercher dans un ensemble** de fonctions

## Important:

Pour que l'apprentissage machine fonctionne,  
**il faut disposer de (beaucoup de) données:**  
De *nombreux exemples* de la tâche à réaliser.

# Apprentissage

- Une caractéristique essentielle de l'intelligence naturelle
- Apprendre par cœur vs. l'apprentissage inductif
- Mot clé: GÉNÉRALISATION
- Situation d'apprentissage typique:
  - 1 On nous donne des exemples (des données)
  - 2 On nous présente un nouveau cas et il faut prendre une décision

# Ex: la reconnaissance de caractères

Ensemble d'apprentissage

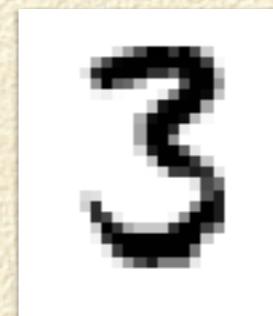
2



3



Point de test



2 ou 3 ?

*Apprendre n'est pas  
simplement  
mémoriser...*

*C'est être capable  
de généraliser!*

à de nouveaux exemples encore  
jamais vus

# Catégories d'apprentissage automatique

- Classification
  - classifier le nouvel exemple
- Régression
  - faire une prédiction à partir du nouvel exemple
- Estimation de densité
  - dire si le nouvel exemple ressemble aux exemples déjà vus

# Apprentissage machine? ou Statistiques?

énormément en commun,  
mais avec une différence de point de vue

► **Statistiques**: branche des **mathématiques**

- Importance de la rigueur et de garanties théoriques
- Hypothèses fortes et tests d'hypothèses

► **Machine learning** (apprentissage machine):  
branche de l'**Intelligence Artificielle** (informatique)

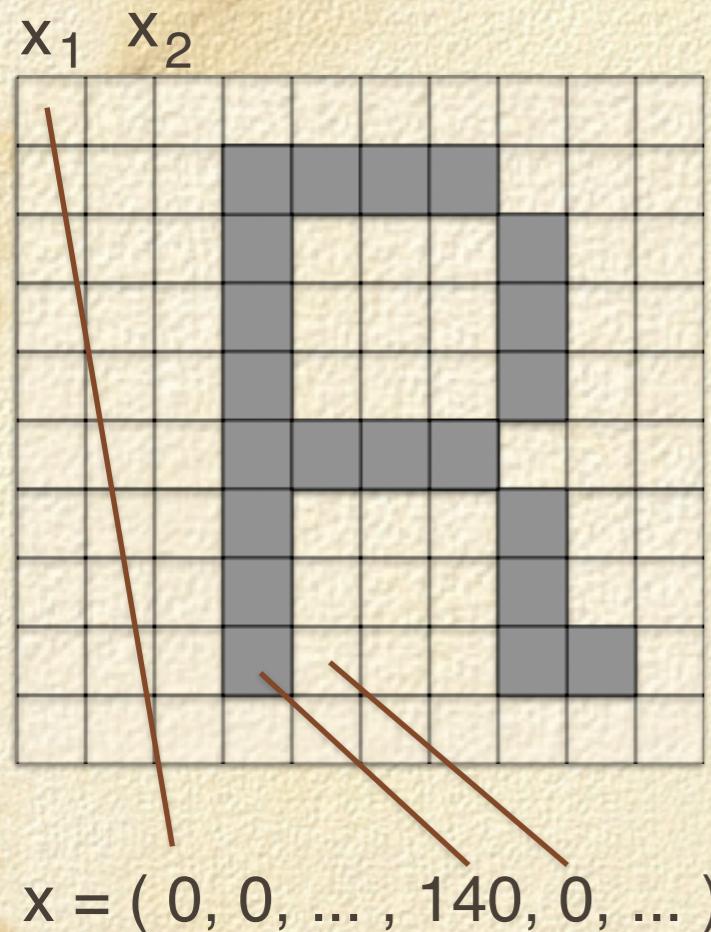
- Grande ambition: l'intelligence!
- On s'inspire de tout ce qu'on peut  
(**neurosciences**, statistiques, physique, théorie de l'information, ...)
- L'important c'est que ça marche! 😊  
Approche pragmatique.

# Data-mining? ou machine learning? ou statistiques? (apprentissage machine)

- ▶ **Statistiques et machine-learning** = études théorique et développements algorithmiques pour l'analyse/apprentissage de données.
- ▶ Le **data-mining** = utilisation de ces techniques sur de gros problèmes industriels (grands ensembles de données).
  - Défis liés à la taille: problèmes de *scaling* (échelle)
  - Approche pratique.

# Représentation vectorielle d'un exemple des données

---



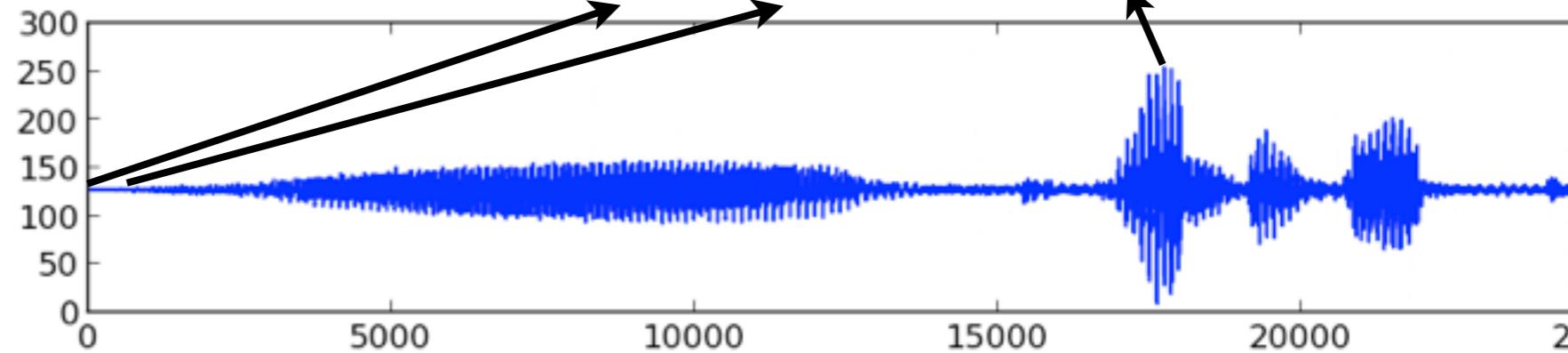
$x$  vecteur de  $R^n$

Transformer un exemple en représentation vecteur  $\mathbf{x} \in \mathbb{R}^d$

Représentation brute:

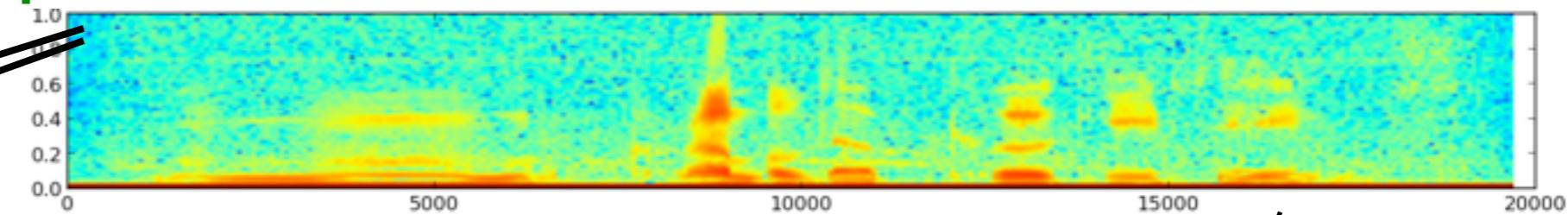
$$\mathbf{x} = (0, 0, \dots, 54, 120, \dots, 0, 0)$$

$$\mathbf{x} = (125, 125, \dots, 250, \dots)$$



OU caractéristiques prétraitement:

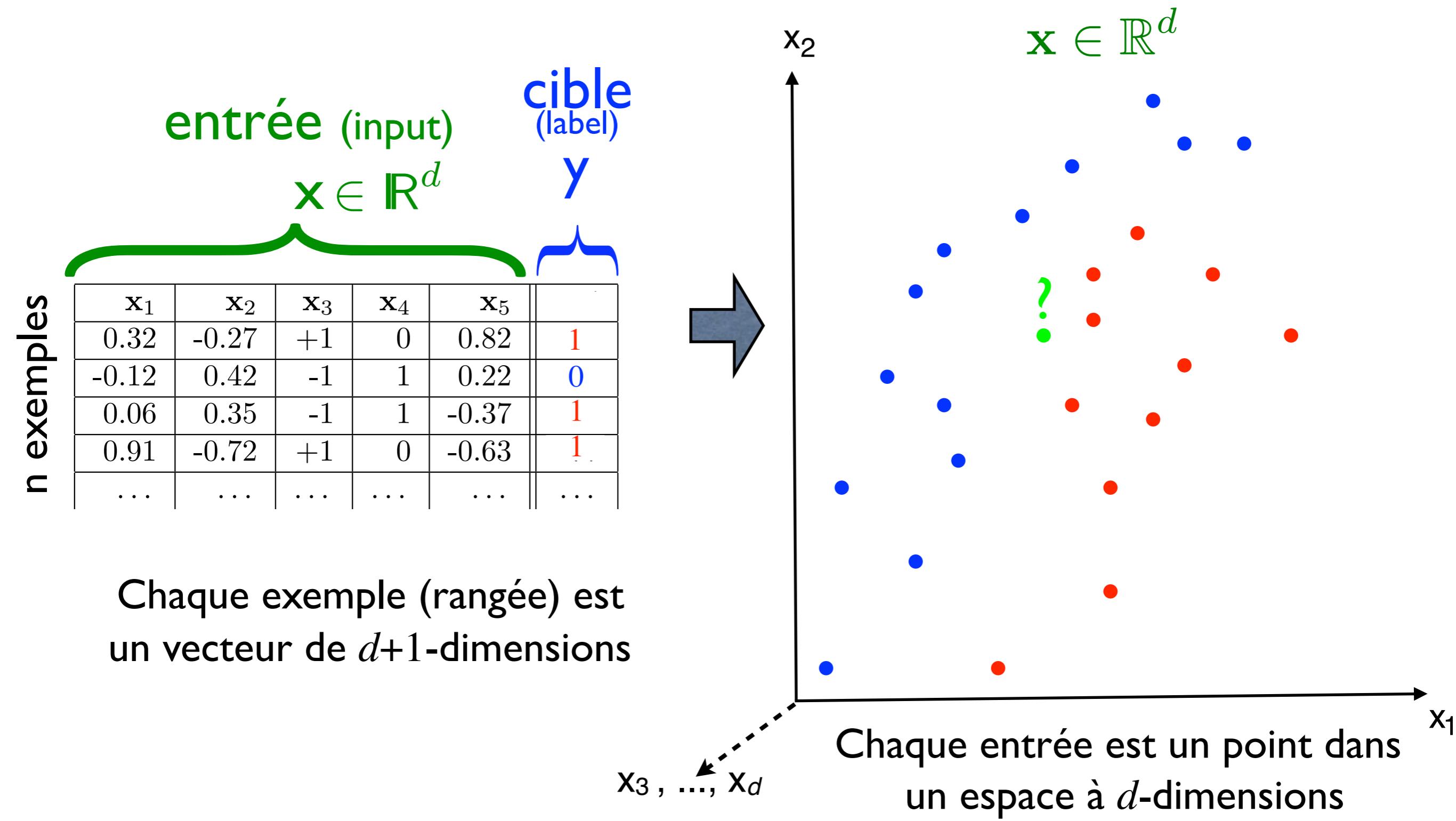
$$\mathbf{x} = (, , , , \dots)$$



Bag of words pour «The cat jumped»:  $\mathbf{x} = (\dots 0 \dots, 0, 1, \dots 0 \dots, 1, 0, 0, \dots, 0, 0, 1, 0, \dots 0 \dots)$

OU vecteur de caractéristiques fait-main:  $\mathbf{x} = (\text{feature } 1, \dots, \text{feature } d)$   
ex: Histograms of Oriented Gradients

# Ensemble de données vue comme un nuage de points dans un espace vectoriel de *haute-dimension*



# Ex: l'algorithme du plus proches voisin

Pour un point test  $x$ :

- On trouve **le plus proche voisin** de  $x$  parmi l'ensemble d'apprentissage selon une certaine mesure de
- On associe à  $x$  la classe de ce plus proche voisin.

