

Introduction à l'apprentissage automatique

Thierry Denœux

Université de Technologie de Compiègne
Département Génie Informatique
Heudiasyc (UMR CNRS 7253)

14 septembre 2018

Plan

1 L'Intelligence Artificielle

- Les origines
- IA forte vs. IA faible

2 L'apprentissage automatique

- Une méthode simple : la régression logistique
- Du linéaire au non linéaire

3 Les réseaux de neurones profonds

- Perceptron
- Perceptrons multicouches
- Réseaux de convolution

Intelligence artificielle (IA)

Définition

Ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence

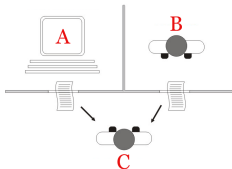
- Une branche de l'**informatique** en interaction avec plusieurs disciplines connexes :
 - Logique mathématique
 - Recherche opérationnelle
 - Neurosciences
 - Linguistique computationnelle
 - Vision par ordinateur
 - Théorie des probabilités, statistique
 - Théorie des jeux
 - Etc.



Plan

- 1 L'Intelligence Artificielle
 - Les origines
 - IA forte vs. IA faible
- 2 L'apprentissage automatique
 - Une méthode simple : la régression logistique
 - Du linéaire au non linéaire
- 3 Les réseaux de neurones profonds
 - Perceptron
 - Perceptrons multicouches
 - Réseaux de convolution

Alan Turing (1912-1954)



- Mathématicien et logicien anglais (King's College, Cambridge ; Univ. of Manchester)
- Fondement théoriques de l'informatique (machine de Turing) et participation au développement des premiers ordinateurs
- L'un des premiers à défendre la thèse que des ordinateurs suffisamment complexes pourraient un jour devenir intelligents (article « Computing Machinery and Intelligence », *Mind*, octobre 1950)
- Inventeur du **test de Turing** (partiellement réussi en 2014 : 33% des juges convaincus au bout de 5 mn lors d'un test réalisé à la Royal Society)

Fondateurs de l'IA symbolique

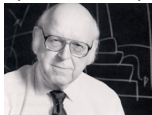
Marvin Minsky
(1927-2016)



Herbert Simon
(1916-2001)



Alan Newell
(1927-1992)



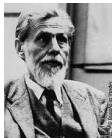
John McCarthy
(1927-2011)



- Fondateurs de l'IA en tant que discipline lors d'un colloque au Dartmouth College (NH, EU) en 1956.
- Initiateurs dans les années 1950 et 1960 de nombreux travaux sur les jeux (dames), la preuve automatique de théorèmes, le traitement du langage naturel, etc.
- **IA symbolique** : l'intelligence se réduit à de la manipulation de symboles. Accent mis sur la logique, le raisonnement, la résolution de problèmes et la représentation des connaissances (systèmes experts).

Fondateurs de l'IA connexionniste

Warren McCulloch
(1898-1969)



Walter Pitts
(1923-1969)



Franck Rosenblatt
(1928-1971)



- Premier modèle de **neurone artificiel** (1943), issue de la rencontre d'un neurophysiologiste (McCulloch) et d'un logicien (Pitts).
- Proposition d'un algorithme d'**apprentissage** et implémentation sur une machine (le perceptron) par Rosenblatt en 1957.
- **IA connexionniste** : pour simuler l'intelligence, il faut s'inspirer de l'architecture du cerveau humain, composé d'un grand nombre d'unités simples (neurones) interconnectées. Utilisation de méthodes numériques (optimisation) et statistiques.

Plan

1 L'Intelligence Artificielle

- Les origines
- IA forte vs. IA faible

2 L'apprentissage automatique

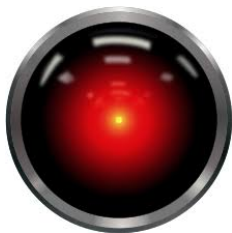
- Une méthode simple : la régression logistique
- Du linéaire au non linéaire

3 Les réseaux de neurones profonds

- Perceptron
- Perceptrons multicouches
- Réseaux de convolution

IA forte

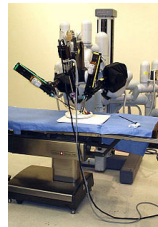
L'**IA forte** se donne pour objectif ultime de développer une machine qui soit non seulement capable de comportements intelligents, mais également douée de **conscience**.



- Théorie de la singularité : la conscience émergera spontanément à partir d'un certain degré de complexité
- Certains chercheurs doutent de la faisabilité d'une telle entreprise, au moins sur des ordinateurs tels qu'ils sont conçus actuellement (Roger Penrose)
- Cet objectif est très loin d'être atteint, même si cette perspective alimente certaines craintes (Bill Gates, Stephen Hawking).

IA faible

L'**IA faible** repose sur une approche pragmatique d'ingénieur : elle consiste à développer des systèmes de plus en plus autonomes permettant de résoudre des problèmes de plus en plus complexes, en simulant l'intelligence humaine.



Grandes étapes de l'IA

1956-1973 Epoque des pionniers de l'IA symbolique. Période de grand optimisme :

- « *Machines will be capable, within twenty years, of doing any work a man can do* » (H. Simon, 1965)
- « *Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved* » (M. Minsky, 1967)

1974-1980 « Hiver de l'IA »

1980-1989 Systèmes experts. En 1985, marché d'un milliard de dollars. Projets ambitieux (ordinateur de 5e génération au Japon)

1985-1995 Réseaux de neurones

2000→2010- Explosion des applications (augmentation de la puissance des ordinateurs, disponibilité de grandes masses de données et progrès réalisés en apprentissage automatique)

Plan

- 1 L'Intelligence Artificielle
 - Les origines
 - IA forte vs. IA faible
- 2 L'apprentissage automatique
 - Une méthode simple : la régression logistique
 - Du linéaire au non linéaire
- 3 Les réseaux de neurones profonds
 - Perceptron
 - Perceptrons multicouches
 - Réseaux de convolution

Apprentissage automatique (*Machine Learning*)

Définition

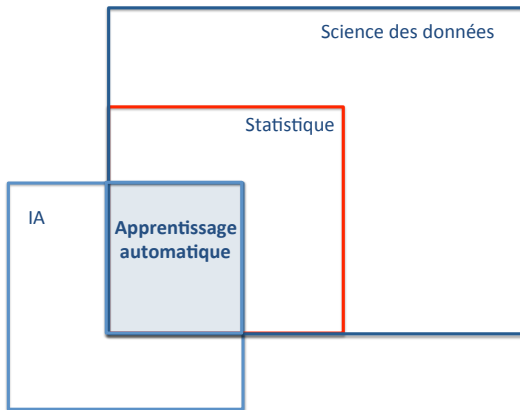
Branche de l'IA qui concerne le développement d'algorithmes permettant de rendre une machine capable d'accomplir des tâches complexes sans avoir été explicitement programmée dans ce but.



- Exemple : comment écrire un programme qui reconnaisse les caractères manuscrits ?
 - Entrer des règles manuellement (difficile et peu fiable)
 - Meilleure méthode : écrire un algorithme (générique) qui génère automatiquement un programme de reconnaissance de caractères à partir d'un **grand nombre d'exemples**.

Science des données

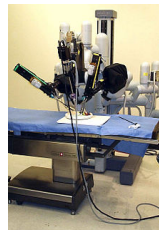
L'apprentissage automatique fait largement appel à des outils et des concepts de la **statistique**, et fait partir d'une discipline plus vaste appelée **science des données**.



Applications de l'apprentissage automatique



	2	2	4	5	
	5	4			1
		5		2	
	1		5		4
		4			2
	4	5	1		



Plan

- 1 L'Intelligence Artificielle
 - Les origines
 - IA forte vs. IA faible
- 2 L'apprentissage automatique
 - Une méthode simple : la régression logistique
 - Du linéaire au non linéaire
- 3 Les réseaux de neurones profonds
 - Perceptron
 - Perceptrons multicouches
 - Réseaux de convolution

Exemple : prédiction de l'infarctus du myocarde

- On sait que la probabilité de développer un infarctus du myocarde (IM) augmente avec l'âge et avec le taux de cholestérol LDL.
- Comment développer un programme qui **prédise le risque d'IM** à partir de ces deux variables ?
- Approche de type « système expert » : modéliser la connaissance d'un médecin par des règles de la forme

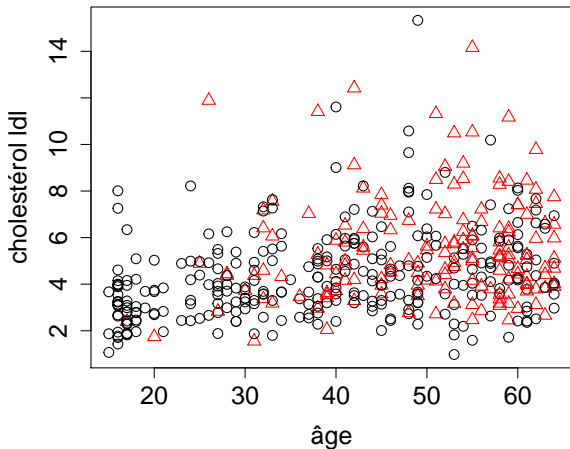
SI âge > 60 ET ldl > 10 ALORS risque élevé

SI $50 < \text{âge} \leq 60$ ET $8 < \text{ldl} \leq 10$ ALORS risque moyen

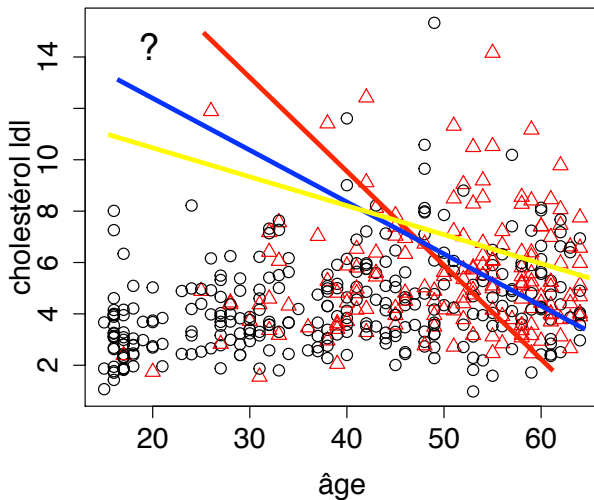
⋮

- Approche peu fiable et difficile à mettre en œuvre, d'autant plus que le nombre de variables explicatives est important.

Ensemble d'apprentissage



Comment trouver une droite séparatrice optimale ?



Modèle

- On ne peut pas prédire à coup sûr l'occurrence d'un IM à partir de l'âge et du taux de cholestérol LDL, mais on peut chercher à estimer la probabilité

$$\mathbb{P}(\underbrace{\text{IM}}_{y=1} \mid \underbrace{\text{âge, ldl}}_x) \text{ notée } p(x)$$

- Modèle linéaire :

$$\ln \frac{p(x)}{1 - p(x)} = w_0 + w_1 \times \text{âge} + w_2 \times \text{ldl}$$

- Formulation équivalente :

$$p(x) = \frac{1}{1 + \exp(-w_0 - w_1 \times \text{âge} - w_2 \times \text{ldl})}$$

Fonction d'erreur

- Si $y = 1$, on veut avoir $p(\mathbf{x})$ aussi grand que possible. On définit l'erreur dans ce cas par $-\ln p(\mathbf{x})$.
- Symétriquement, si $y = 0$, on veut avoir $p(\mathbf{x})$ aussi petit que possible. L'erreur est alors $-\ln(1 - p(\mathbf{x}))$.
- Formule générale :

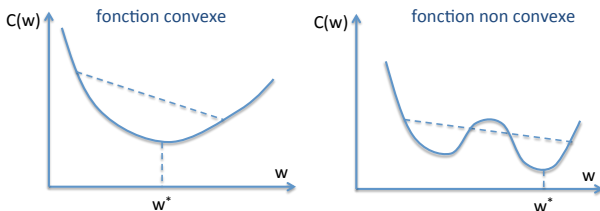
$$\text{erreur} = -y \ln p(\mathbf{x}) - (1 - y) \ln(1 - p(\mathbf{x}))$$

- Erreur totale pour un ensemble d'apprentissage $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ mesurée par l'entropie-croisée :

$$C(\underbrace{w_0, w_1, w_2}_{\mathbf{w}}) = \sum_{i=1}^n \text{erreur}_i$$

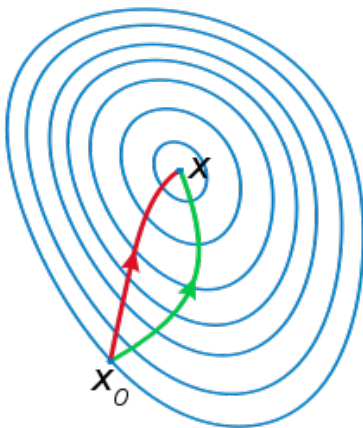
Minimisation de l'erreur

- Une fois définie une fonction d'erreur, le problème de l'apprentissage devient un **problème d'optimisation** : rechercher le vecteur de coefficient w^* qui minimise l'erreur.
- Dans le cas de la régression logistique, ce vecteur est unique car la fonction d'erreur est **convexe**.

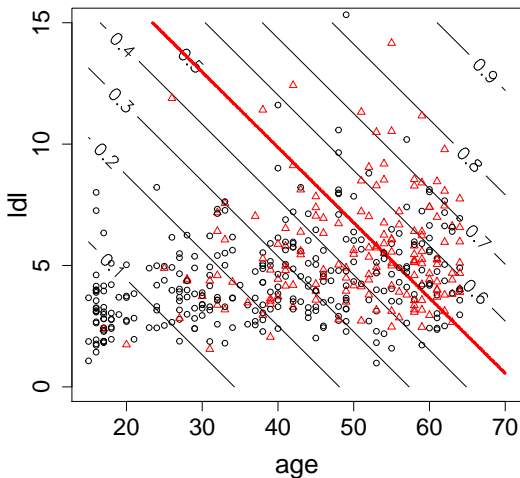


- Le vecteur solution w^* peut être obtenu par un **algorithme itératif**.

Algorithme d'optimisation itératif



Solution



Probabilité d'erreur du classifieur

- Une fois déterminé le vecteur de coefficient \mathbf{w}^* optimal, on dispose d'un programme (classifieur) permettant de classer un nouvel individu.
- Pour estimer la probabilité d'erreur du classifieur, il faut disposer d'un **ensemble de test** indépendant.
- Matrice de confusion pour un ensemble de test de 100 exemples :

		Vraie classe	
		Positif	Négatif
Prediction	Positif	14	10
	Négatif	21	55

Taux d'erreur = $(10+21)/100=31\%$.

Reconnaissance d'expressions

joy



surprise



sadness



disgust



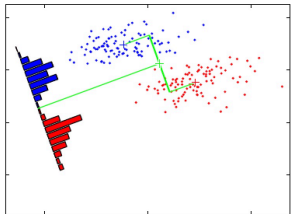
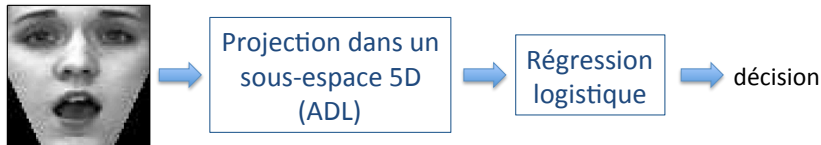
anger



fear

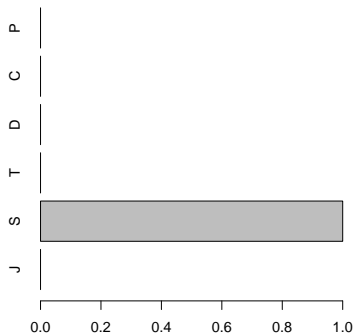
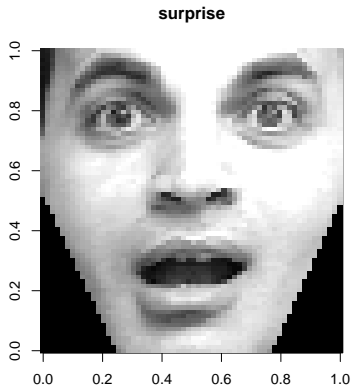


Apprentissage

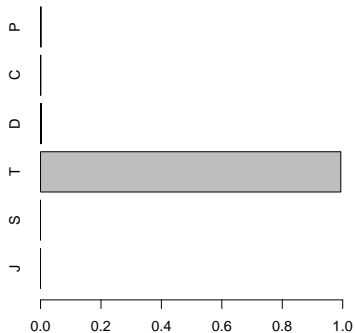
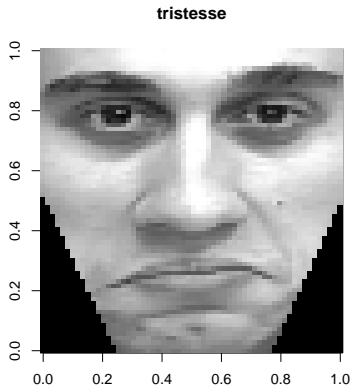


- 216 images 70×60 (36 par expression)
- 144 pour l'apprentissage, 72 pour le test
- 5 caractéristiques extraites par analyse discriminante linéaire
- Taux d'erreur de test : 23.6% (au hasard : 83.3%)

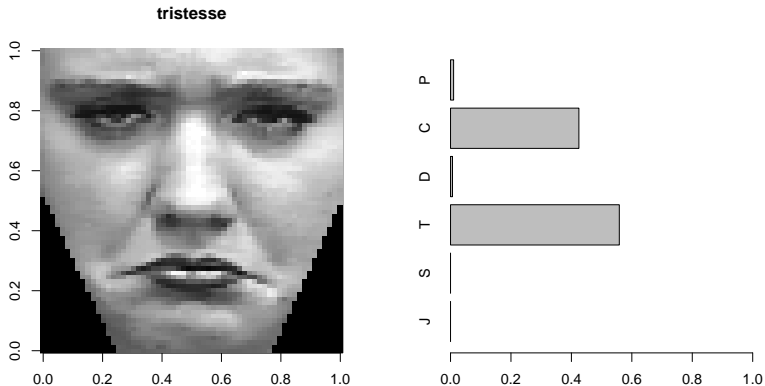
Résultats



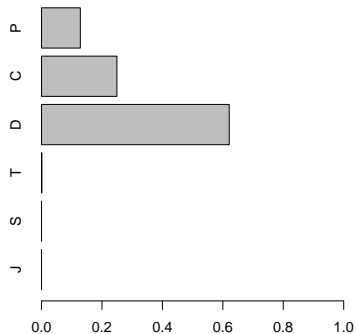
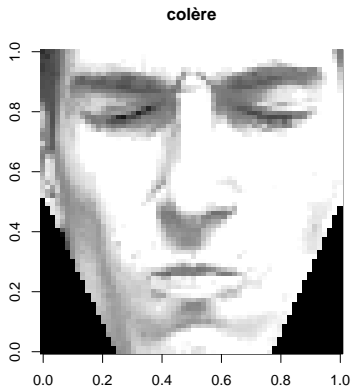
Résultats



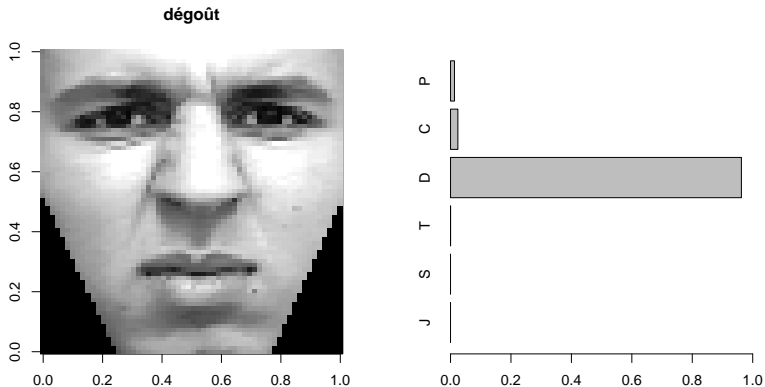
Résultats



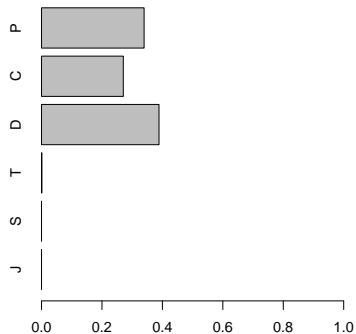
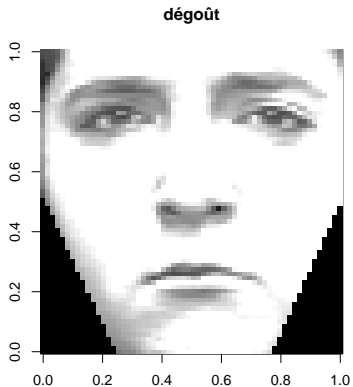
Résultats



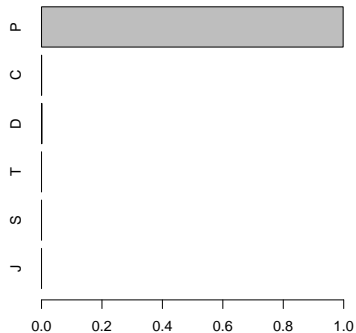
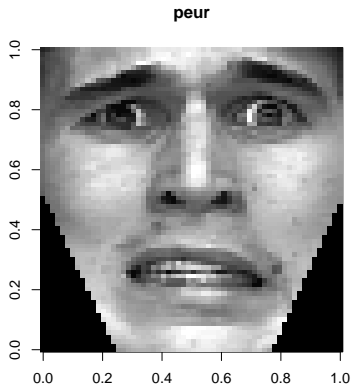
Résultats



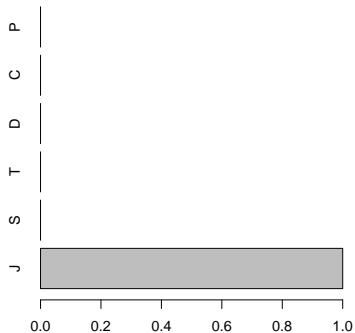
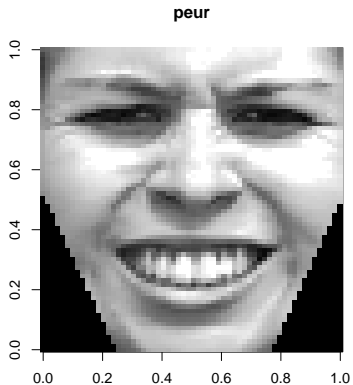
Résultats



Résultats



Résultats



Plan

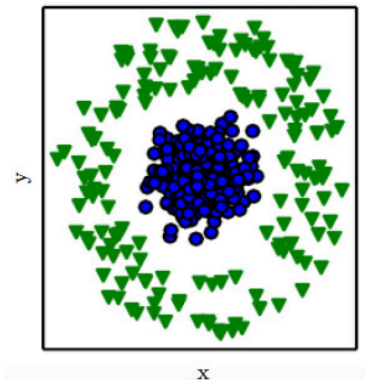
- 1 L'Intelligence Artificielle
 - Les origines
 - IA forte vs. IA faible
- 2 L'apprentissage automatique
 - Une méthode simple : la régression logistique
 - Du linéaire au non linéaire
- 3 Les réseaux de neurones profonds
 - Perceptron
 - Perceptrons multicouches
 - Réseaux de convolution

Problème difficile



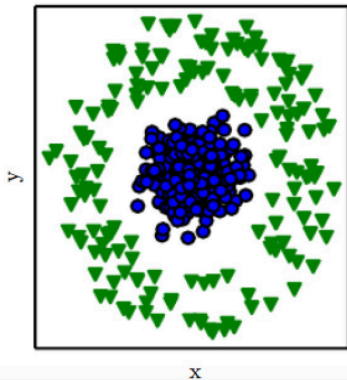
Changement de représentation

Cartesian coordinates

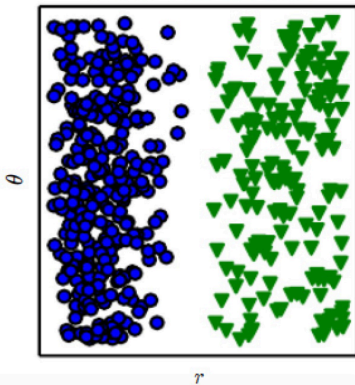


Changement de représentation

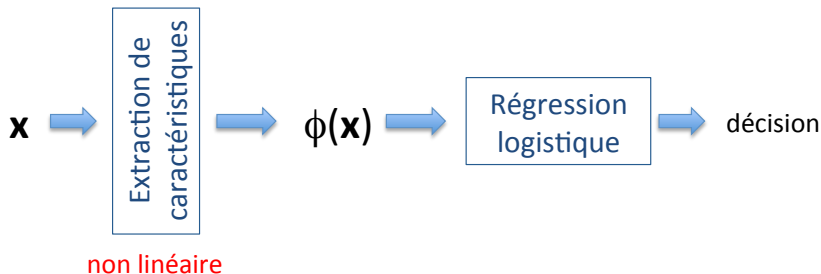
Cartesian coordinates



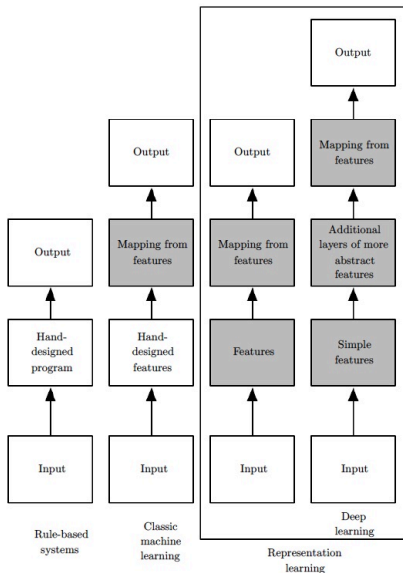
Polar coordinates



Classifieur linéaire généralisé



Apprentissage de représentation



Plan

- 1 L'Intelligence Artificielle
 - Les origines
 - IA forte vs. IA faible
- 2 L'apprentissage automatique
 - Une méthode simple : la régression logistique
 - Du linéaire au non linéaire
- 3 Les réseaux de neurones profonds
 - Perceptron
 - Perceptrons multicouches
 - Réseaux de convolution

Grandes étapes de l'IA connexionniste

Trois étapes :

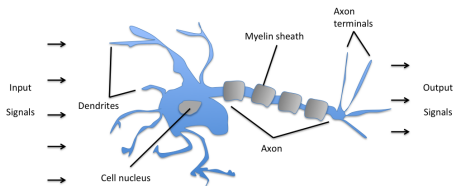
- 1 Perceptron (1955-1965)
- 2 Perceptrons multi-couches (1985-1995)
- 3 Réseaux profonds (2010-)

Plan

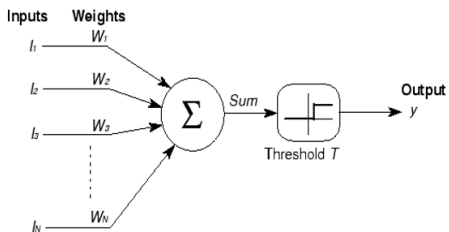
- 1 L'Intelligence Artificielle
 - Les origines
 - IA forte vs. IA faible
- 2 L'apprentissage automatique
 - Une méthode simple : la régression logistique
 - Du linéaire au non linéaire
- 3 Les réseaux de neurones profonds
 - Perceptron
 - Perceptrons multicouches
 - Réseaux de convolution

Le modèle de McCulloch et Pitts

- W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4) :115-133, 1943.
- Idée : neurones biologiques vus comme des **portes logiques** effectuant des opérations de la logique booléenne.

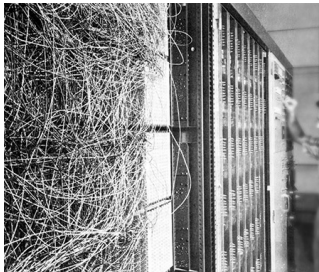


Schematic of a biological neuron.

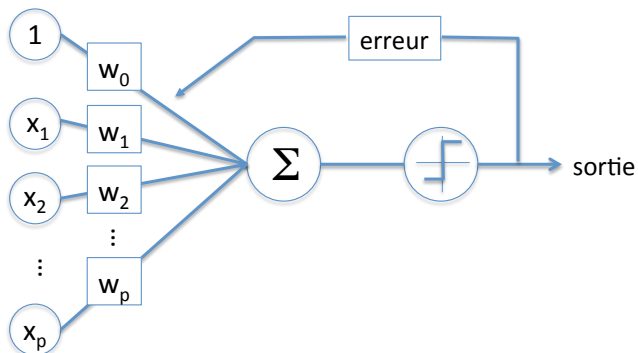


Perceptron

- F. Rosenblatt. *The perceptron, a perceiving and recognizing automaton (Project PARA)*. Cornell Aeronautical Laboratory, 1957.
- Idée : une algorithmme qui **apprend les poids** pour résoudre des problèmes de classification binaire.



Principe du perceptron

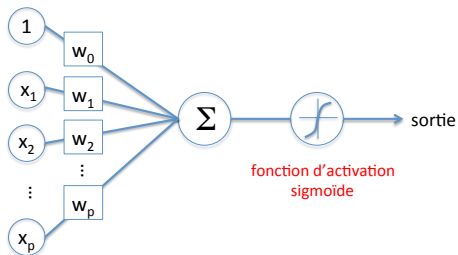


Limites :

- L'algorithme ne converge que si les deux classes sont bien séparées

Difficilement généralisable à plus de deux classes

Version moderne du perceptron



- Sortie :

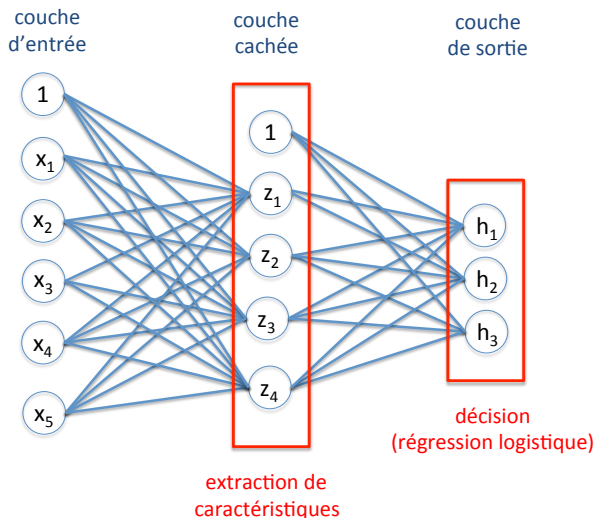
$$g(x) = \frac{1}{1 + \exp[-(w_0 + w_1x_1 + \dots + w_px_p)]}$$

- Apprentissage des poids par minimisation de l'entropie croisée
- C'est exactement le modèle de la **régression logistique** !

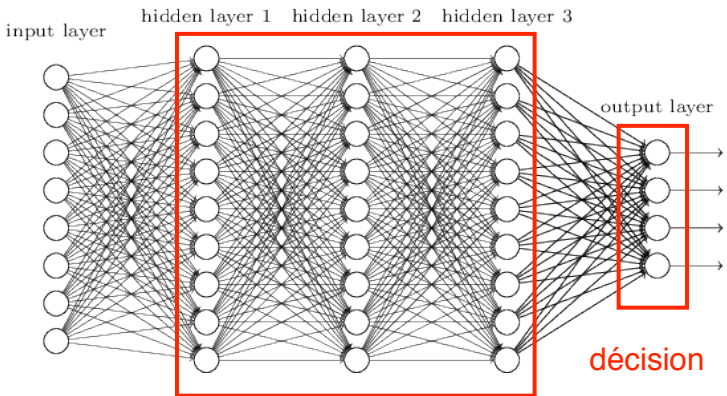
Plan

- 1 L'Intelligence Artificielle
 - Les origines
 - IA forte vs. IA faible
- 2 L'apprentissage automatique
 - Une méthode simple : la régression logistique
 - Du linéaire au non linéaire
- 3 Les réseaux de neurones profonds
 - Perceptron
 - **Perceptrons multicouches**
 - Réseaux de convolution

Architecture (une couche cachée)



Architecture (plusieurs couches cachées)

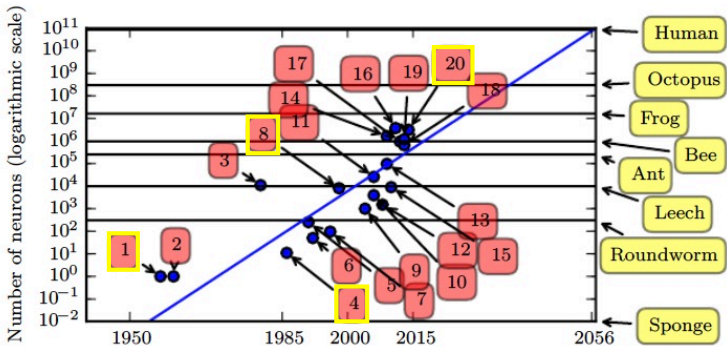


extraction de
caractéristiques

décision

Croissance de la taille des réseaux de neurones

Depuis leur introduction dans les années 1950, la taille des réseaux de neurones a doublé en moyenne toutes les 2.4 années.



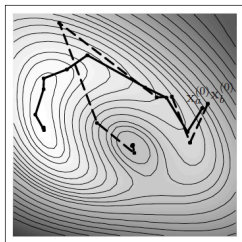
1 : Perceptron ; 4 : Premiers perceptrons multi-couches ; 8 : LeNet-5 (LeCun et al., 1998b) ; 20 : GoogLeNet

Difficultés

- 1 Apprentissage des poids
- 2 Détermination de l'architecture du réseau

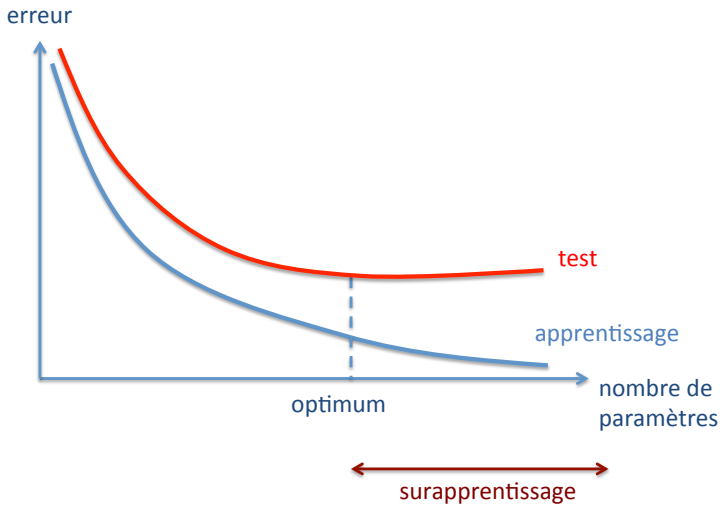
Apprentissage des poids

- Comme pour la régression logistique, l'apprentissage des poids se fait en **minimisant une fonction d'erreur** telle que l'entropie-croisée.

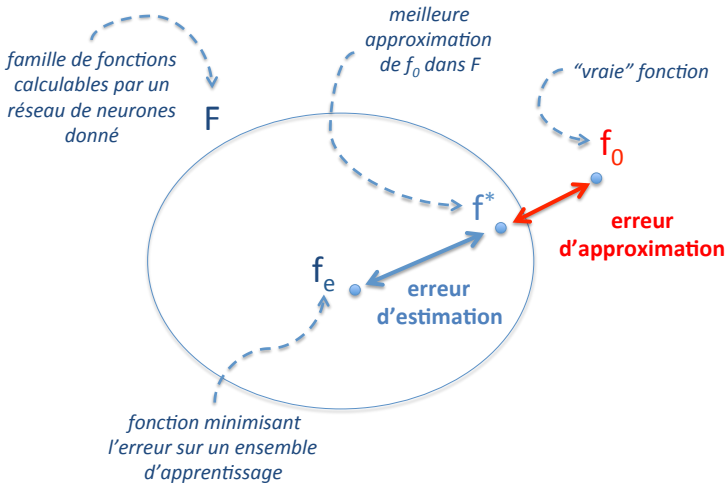


- Problèmes :
 - Fonction d'erreur **non-convexe** : beaucoup de minima locaux
 - Très grand nombre de variables (poids du réseau)
- La résolution de problèmes d'apprentissage complexes nécessite de grosses capacités de calcul.

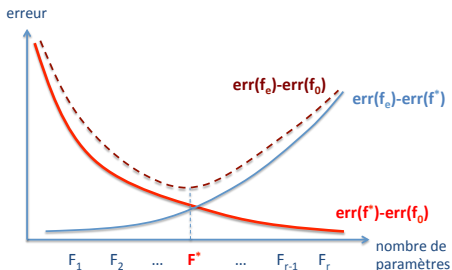
Détermination de l'architecture



Explication

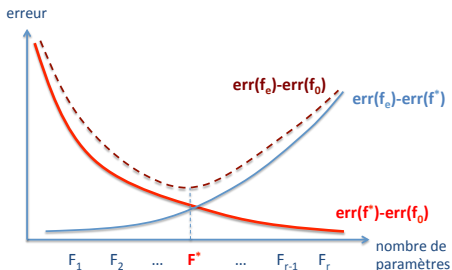


Méthode



- 1 Définir une collection de familles de fonctions emboîtées $F_1 \subset \dots \subset F_r$ (différentes architectures ou différentes contraintes sur les poids)
- 2 Pour chaque architecture F_i , chercher le réseau f_{ei} minimisant l'erreur d'apprentissage, et estimer son erreur de généralisation $err(f_{ei})$
- 3 Sélectionner le réseau ayant la plus petite erreur de généralisation

Méthode

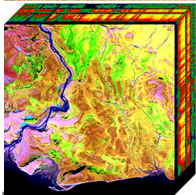
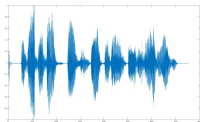


- 1 Définir une collection de familles de fonctions emboîtées $F_1 \subset \dots \subset F_r$ (différentes architectures ou différentes contraintes sur les poids)
- 2 Pour chaque architecture F_i , chercher le réseau f_{ei} minimisant l'erreur d'apprentissage, et estimer son erreur de généralisation $\text{err}(f_{ei})$
- 3 Sélectionner le réseau ayant la plus petite erreur de généralisation

Plan

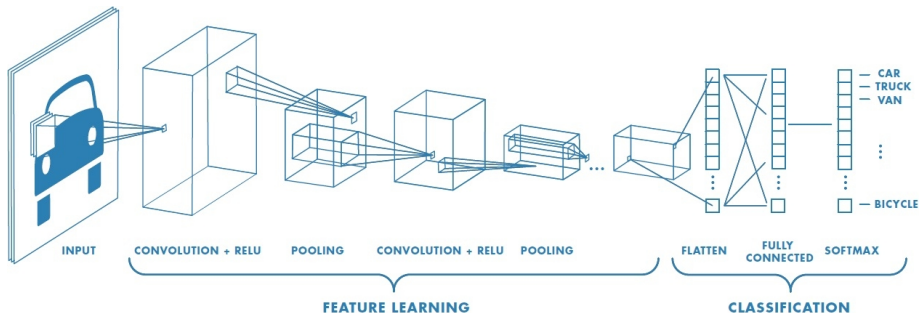
- 1 L'Intelligence Artificielle
 - Les origines
 - IA forte vs. IA faible
- 2 L'apprentissage automatique
 - Une méthode simple : la régression logistique
 - Du linéaire au non linéaire
- 3 Les réseaux de neurones profonds
 - Perceptron
 - Perceptrons multicouches
 - Réseaux de convolution

Apprentissage à partir de données 1D, 2D, 3D, ...



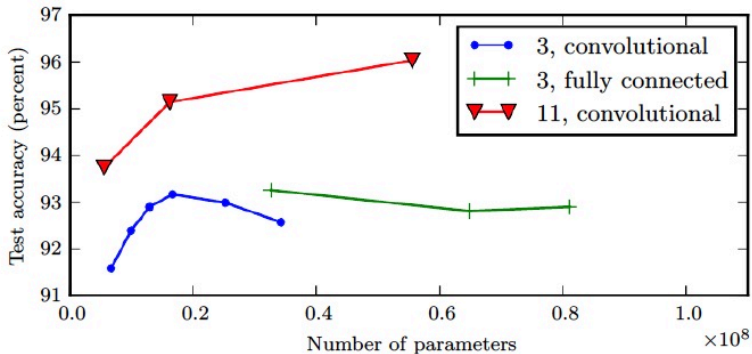
- Depuis quelques années, des progrès importants ont été réalisés dans le traitement de données tensorielles :
 - 1D : textes, musique, parole, séries temporelles
 - 2D : images
 - 3D : séquences vidéo, imagerie médicale, images multi-spectrales, etc.
- Propriétés :
 - Corrélation locale
 - Caractéristiques pouvant apparaître n'importe où dans le signal
 - Objets invariants par translation, homothétie et distorsion

Réseaux de convolution

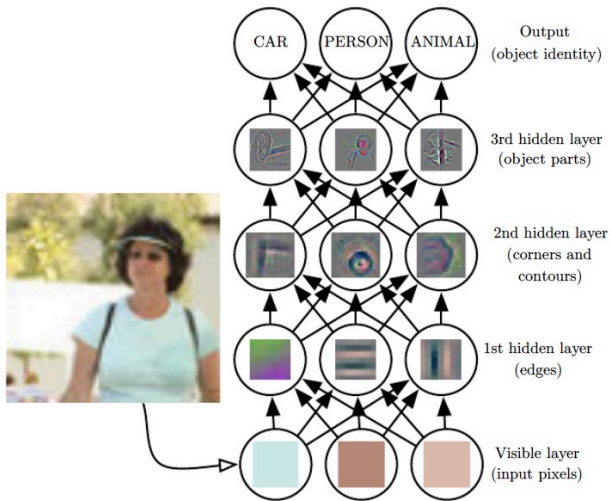


- **Caractéristiques :**
 - Connexions locales (champs réceptifs)
 - Partage des poids
- **Avantages :**
 - Beaucoup moins de paramètres
 - Invariance par translation

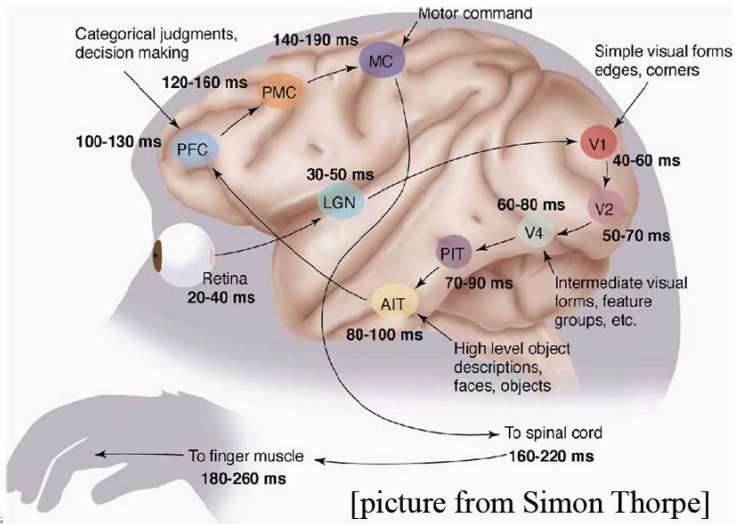
Performances



Extraction de caractéristiques



Base biologique



Défis de l'IA

- L'IA : un des **grands défis scientifiques et technologiques** de notre temps
- Questions **scientifiques** :
 - Intégration de techniques symboliques (raisonnement, gestion des connaissances) et connexionnistes
 - Explication des raisonnements et des décisions
 - Quantification des incertitudes
 - Connaissances a priori en apprentissage (apprendre à partir de peu d'exemples)
 - Etc.
- Questions **éthiques** (protection de la vie privée, robots autonomes, etc.)
- Questions **économiques** (automatisation : impacts négatifs et positifs sur l'emploi, transformation de certains secteurs économiques – transports, services, etc.)

L'IA à l'UTC

- Branche :
 - GI, filières **ICSI** “Ingénierie des Connaissances et des Supports d’Information”, **FDD** “Fouille de Données et Décisionnel (Data Mining)”
 - UVs : **SY02** (statistique), **SY09** (analyse de données), **SY19** (apprentissage automatique), **SY27** (machines intelligentes), **RO04** (optimisation et recherche opérationnelle), **IA01** (IA : représentation), **IA02** (résolution de problèmes et programmation logique)
- Master :
 - Mention **Ingénierie des Systèmes Complexes**, Parcours “Apprentissage et optimisation des systèmes complexes”
 - UVs : Avancées en apprentissage statistique, Apprentissage profond, Représentation des incertitudes
- Recherche :
 - Stage TN09, TN10/Master en laboratoire
 - Doctorat

Heudiasyc

Informatique
Robotique

Automatique
Intelligence artificielle

**Laboratoire
Heudiasyc
UMR 7253**

**UMR HEUDIASYC
7253**

Heudiasyc opère dans le domaine
des sciences de l'information et
du numérique, notamment
l'informatique, l'automatique, la
robotique et l'intelligence
artificielle

[EN SAVOIR PLUS](#)

Equipe CID



CID
**"CONNAISSANCES,
 INCERTITUDES,
 DONNÉES"**

Les activités de recherche de l'équipe relèvent de l'intelligence artificielle : apprentissage statistique, gestion des incertitudes et ingénierie des connaissances

EN SAVOIR PLUS

Equipe SyRI



SYRI "SYSTÈMES ROBOTIQUES EN INTERACTION"

L'objectif des recherches menées dans l'équipe est l'étude et le développement de systèmes embarqués permettant de doter les robots mobiles de capacité d'autonomie

EN SAVOIR PLUS



Merci pour votre attention !