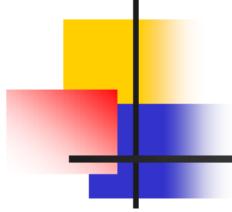


INF5081

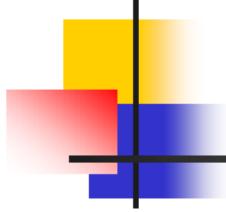
Concepts de l'apprentissage automatique

Mohamed Bouguessa



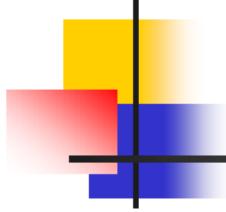
Apprentissage naturel

- **Apprentissage naturel** : La faculté d'apprendre de ses expériences passées et de s'adapter est une caractéristique essentielle des formes de vies. Elle est essentielle à l'être humain dans les premières étapes de la vie pour apprendre des choses aussi fondamentales que reconnaître une voix, un visage familier, apprendre à comprendre ce qui est dit, à marcher et à parler.



Apprentissage artificiel - Généralités

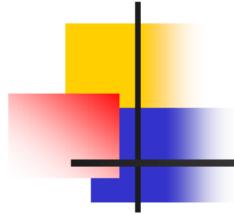
- **Apprentissage machine (Machine Learning)** : est une tentative de comprendre et reproduire cette faculté d'apprentissage dans des systèmes artificiels.
- Il s'agit, très schématiquement, de concevoir des algorithmes capables, à partir d'un nombre important d'exemples (les données correspondant à « l'expérience passée »), d'en assimiler la nature afin de pouvoir appliquer ce qu'ils ont ainsi appris aux cas futurs.



Apprentissage artificiel - Généralités

En résumé :

- **Apprentissage artificiel** : Construction de machines (programmes) capable d'*apprendre* à partir d'un ensemble d'exemples ou d'observations.
- Un **exemple** (on emploiera aussi le terme « donnée ») = consiste en un objet ou sa représentation (un document, une image, ...) accompagné de la catégorie à laquelle il appartient.
- **Objectif de l'apprentissage** : déterminer la relation entre les objets et leurs catégorie pour la prédiction et la découverte des connaissances.

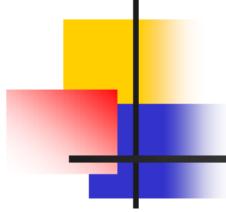


Apprentissage artificiel - Définitions

□ Apprentissage?

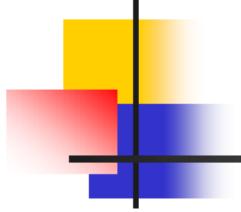
- Capacité d'un programme à **améliorer ses performances** via des interactions avec son **environnement**.

- Un programme possède des capacités d'apprentissage si ses potentialités de comportement sur les données se modifient en fonction de ses performances au fur et à mesure qu'il traite les données → **un programme qui s'améliore avec l'expérience**.



Apprentissage artificiel - Définitions

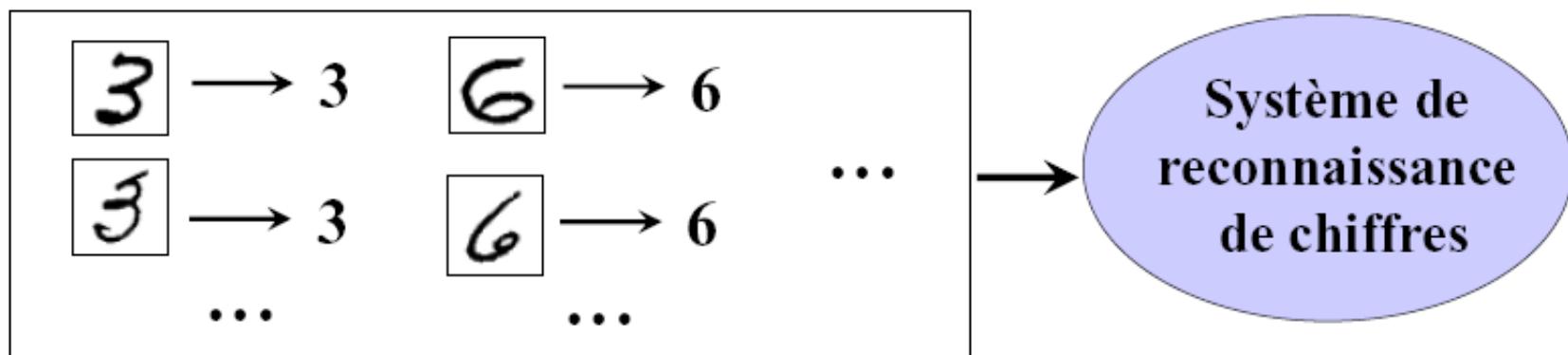
- Un programme possède des capacités d'apprentissage, si au cours du traitement d'exemples représentatifs de données, il est capable de construire et d'utiliser une représentation de ce traitement en vue de son exploitation → **élaboration d'un modèle pour la prédition et la découverte des connaissances.**
- Modèle = Description formelle des relations qui existent entre l'ensemble des attributs qui décrivent les données à traiter.

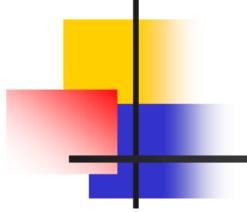


Apprentissage artificiel - Motivations

Construction de logiciels que l'on ne sait pas programmer «à la main» :

Exemple : lecture de codes postaux et reconnaissance de caractère

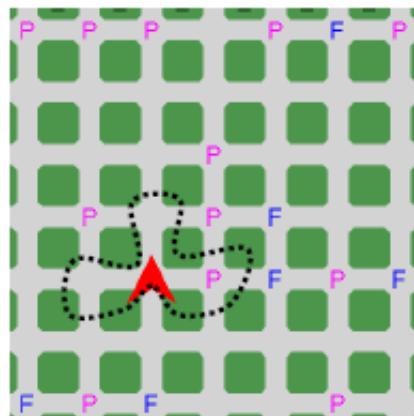




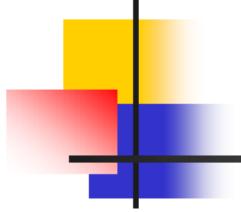
Apprentissage artificiel - Motivations

Construction de programmes qui s'améliorent automatiquement à partir de leur expérience :

- **Programme de jeu** : Il est plus facile d'écrire un programme qui à pour but d'apprendre comment jouer au jeu d'échecs plutôt que de convertir l'expertise d'un joueur professionnel en jeux d'échecs en programme informatique.
- **Robotique** : robot autonome qui apprend à se déplacer dans son environnement



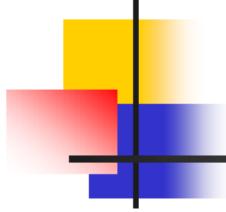
**Navigation « à vue » optimisant
accomplissement d'une tâche
(e.g., collecter de la « nourriture »)**



Algorithme d'apprentissage

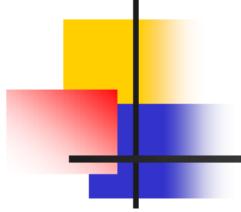


- **Performance visée** : minimiser l'erreur de prédiction.
- **Moyen mis en œuvre** : utiliser des données historiques pour trouver un modèle prédiction le plus correct possible.



Typologie de l'apprentissage artificiel

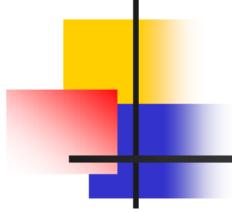
- Capacité d'un programme à **améliorer ses performances** via des interactions avec son **environnement**.
 - Quel « programme » ?
types de modèle (arbre de décision, réseau de neurones, méthodes bayésienne, etc.)
 - Quelles « interactions avec l'environnement » ?
apprentissage « supervisé » ou « non supervisé ».
 - Quelles « performances » ?
fonction de coût, objectif, critère implicite, ...
 - Comment améliorer ?
type d'algorithme (décente de gradient, optimisation sous contrainte, ...)



Quelques notions voisines

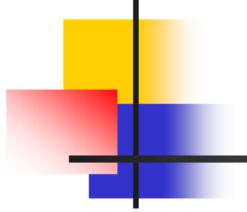
- Extraction d'un concept
- Catégorisation, classification
- Acquisition de connaissances
- Prédiction
- Généralisation
- Compréhension

Forage de données/sciences de données
Data Mining



Apprentissage artificiel

Quelques exemples d'applications



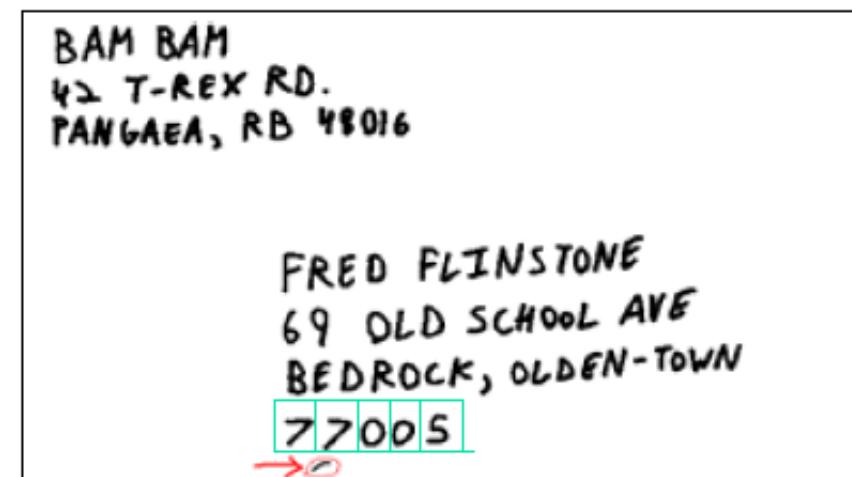
Lecture des codes postaux

Tâche : Classification du courrier à partir des codes postaux



Défis :

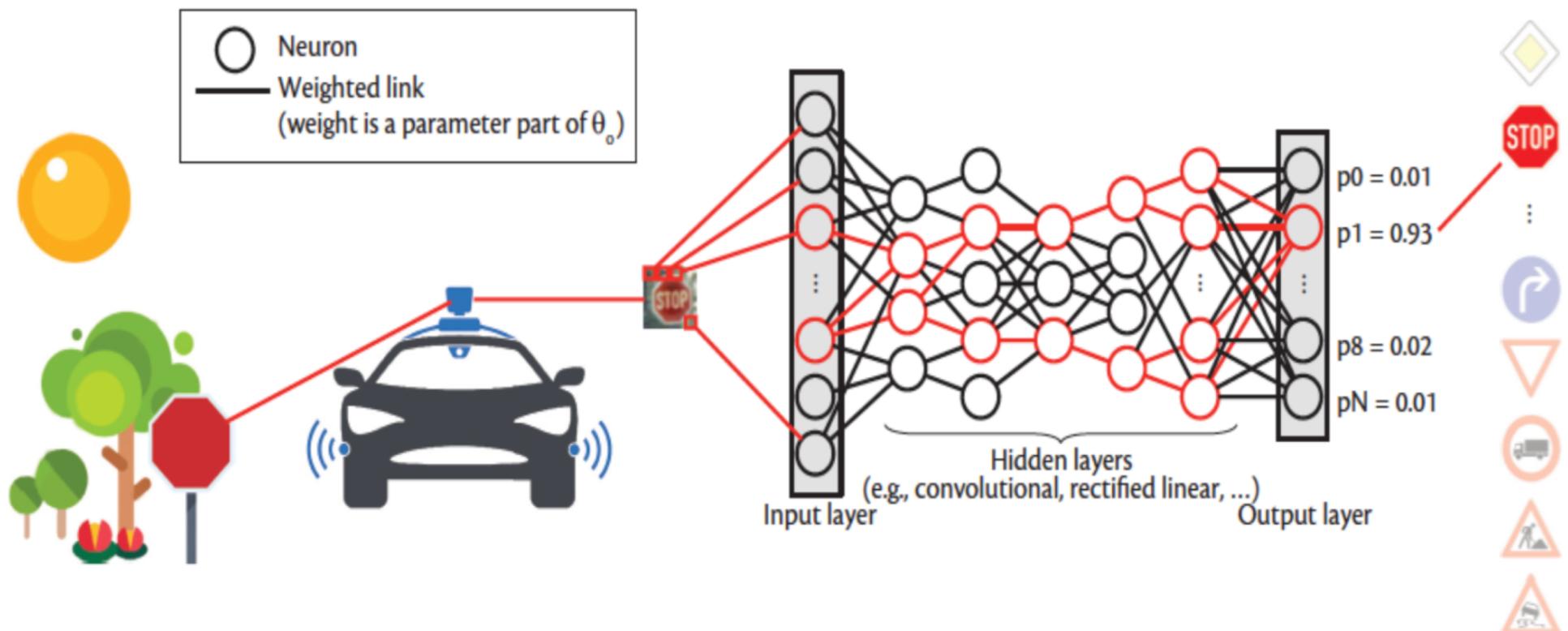
- Variabilité dans les écritures manuelles ;
- Difficulté d'établir des règles fixes pour la reconnaissance des caractères.



Exemple d'application

Traffic-Sign Recognition

[Patrick McDaniel et al., IEEE Security & Privacy, 2016]

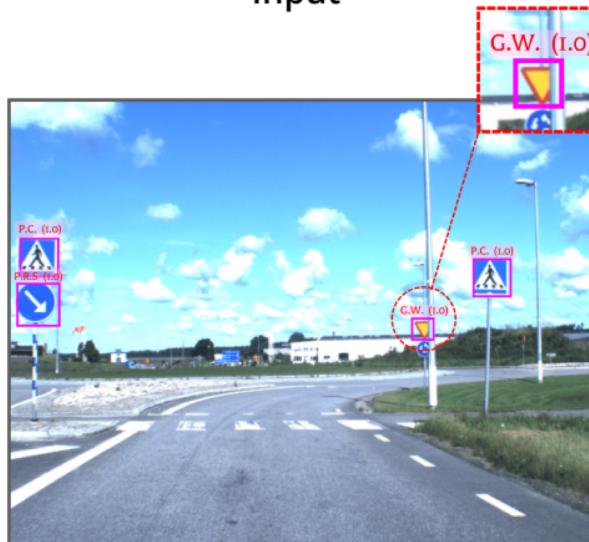


Exemple d'application

Traffic-Sign Recognition



Input



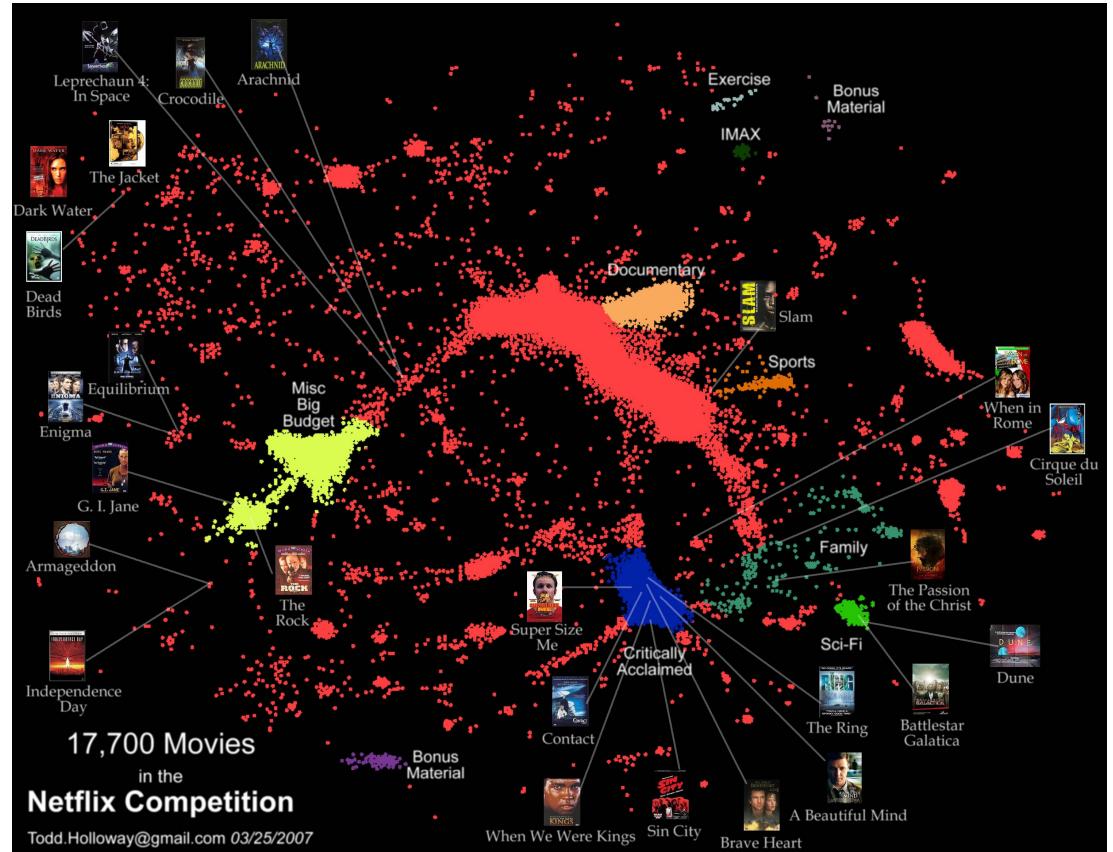
Output

Source : Y. Yang et al. “Efficient Traffic-Sign Recognition with Scale-aware CNN”, BMVC 2017.

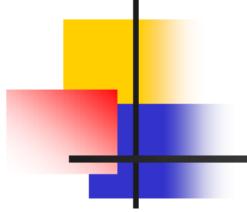
Système de recommandation

Netflix prize : Prédire si un client va aimer ou non un film en se basant sur ces préférences.

Données : plus de 100 million d'évaluations de 18 000 films effectuée par 480 000 utilisateurs.



<http://www.netflixprize.com>



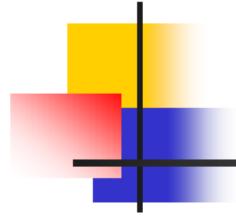
Analyse de crédit

Données :

<i>Customer103: (time=t0)</i>	<i>Customer103: (time=t1)</i>	...	<i>Customer103: (time=tn)</i>
Years of credit: 9	Years of credit: 9		Years of credit: 9
Loan balance: \$2,400	Loan balance: \$3,250		Loan balance: \$4,500
Income: \$52k	Income: ?		Income: ?
Own House: Yes	Own House: Yes		Own House: Yes
Other delinquent accts: 2	Other delinquent accts: 2		Other delinquent accts: 3
Max billing cycles late: 3	Max billing cycles late: 4		Max billing cycles late: 6
Profitable customer?: ?	Profitable customer?: ?		Profitable customer?: No
...

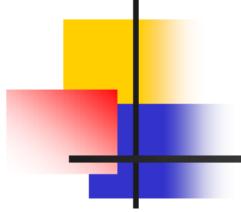
Règles apprises à partir des données :

```
If      Other-Delinquent-Accounts > 2, and  
        Number-Delinquent-Billing-Cycles > 1  
Then   Profitable-Customer? = No  
          [Deny Credit Card application]  
If      Other-Delinquent-Accounts = 0, and  
        (Income > $30k) OR (Years-of-Credit > 3)  
Then   Profitable-Customer? = Yes  
          [Accept Credit Card application]
```



Apprentissage artificiel

Format de données utilisées dans
l'analyse prédictive

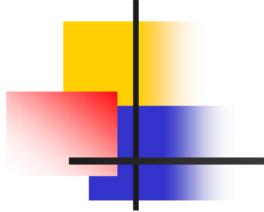


Données en entrée

Enregistrements /Objets	Attributs				Catégorie/ classe
	A1	A2	A3	A4	
Enregistrements /Objets	-0.69	-0.72	Y	0.47	Healthy
	-2.3	-1.2	N	0.15	Disease
	0.32	-0.9	N	-0.76	Healthy
	0.37	-1	Y	-0.59	Disease
	-0.67	-0.53	N	0.33	Healthy
	0.51	-0.09	Y	-0.05	Disease

Données d'apprentissage

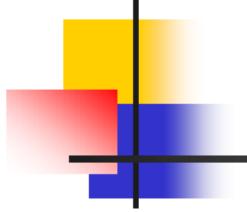
- Tâche : trouver un modèle qui « *approxime et généralise au mieux* » la relation entre les objets et leur catégorie;



But

- La prédition
classification de nouvelle données.

A1	A2	A3	A4	Y
0.83	-0.54	T	0.68	Healthy
-2.3	-1.2	F	-0.83	Disease
0.08	0.63	F	0.76	Healthy
0.06	-0.29	T	-0.57	Disease
-0.98	-0.18	F	-0.38	Healthy
-0.68	0.82	T	-0.95	Disease
0.92	-0.33	F	-0.48	?

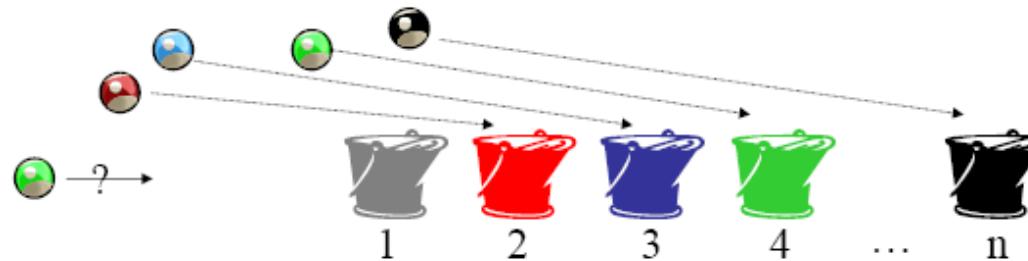


Classification

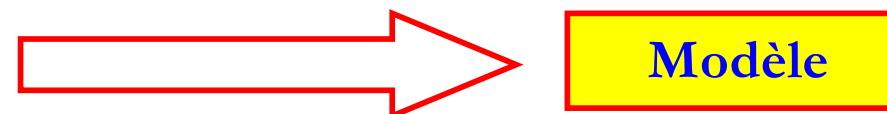
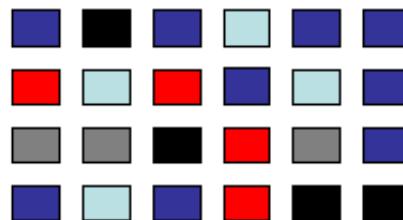
- En quelques termes, la classification:
 - Prédit la catégorie (la classe) d'un objet
 - Construit un modèle basé sur un jeu d'apprentissage et des labels (nom des catégories/classes) et l'utilise pour classer des données nouvelles.

- Applications :
 - ✓ Accord pour un crédit.
 - ✓ Marketing ciblé.
 - ✓ Diagnostique médical.
 - ✓ Prédiction de la faillite.

Processus de classification

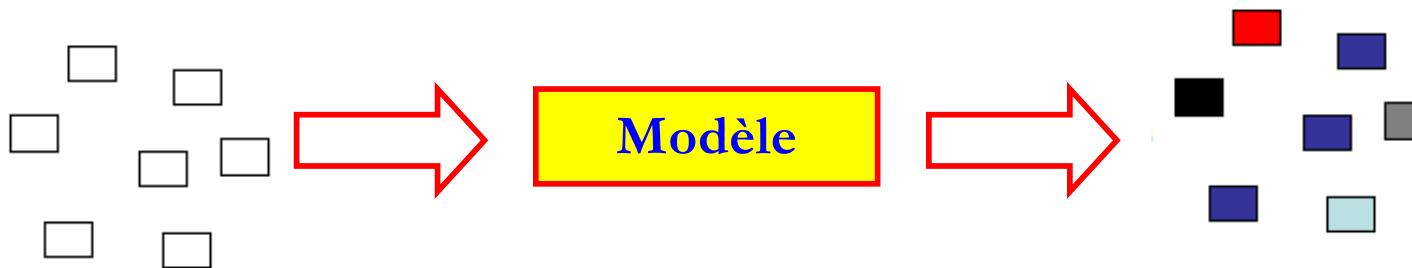


Classification = définition d'un modèle



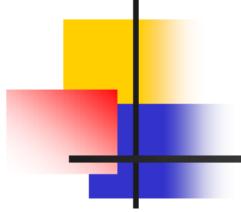
Modèle

Données d'apprentissage – étiquetées
Training Set - labels



Nouvelles
données

Classification



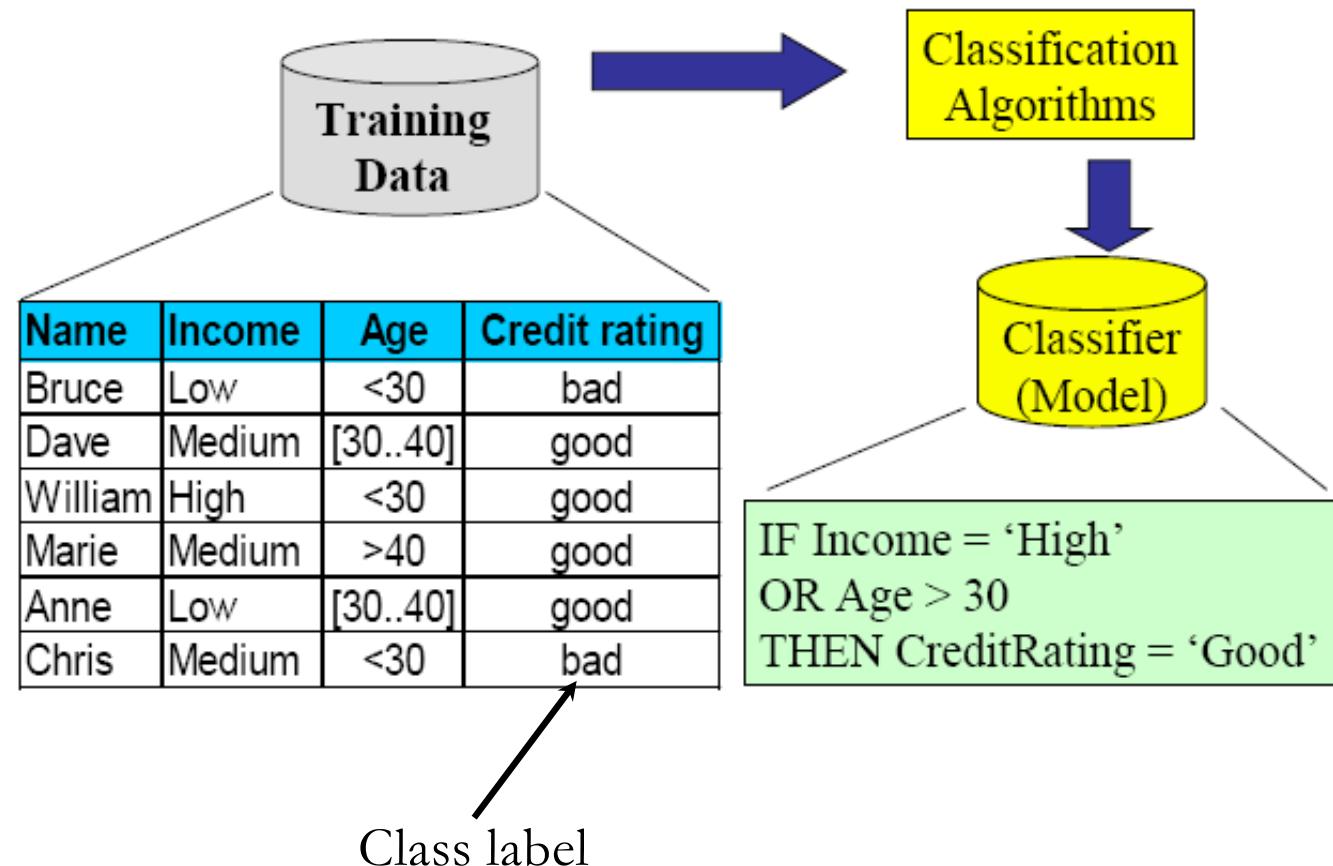
Processus de classification – 3 étapes

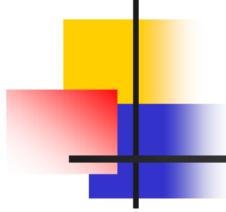
1 – Construction du modèle - Apprentissage

- **Jeu de données d'apprentissage (Training data)** : ensemble des objets utilisés pour la construction du modèle. Dans cet ensemble, chaque objet est supposé appartenir à une classe prédéfinie selon l'attribut « **class label** »
- Le modèle est représentée selon l'une des formes suivantes:
 - ✓ Règles de classification (SI condition(s) ALORS action);
 - ✓ Arbres de décisions;
 - ✓ Réseaux de neurones;
 - ✓ Machines à vecteurs de support (*support vector machines SVM*);
 - ✓ Classificateur Bayesian.
 - ✓ Formule mathématiques.
 - ✓ ...

Processus de classification

1 – Construction du modèle : Schéma





Processus de classification

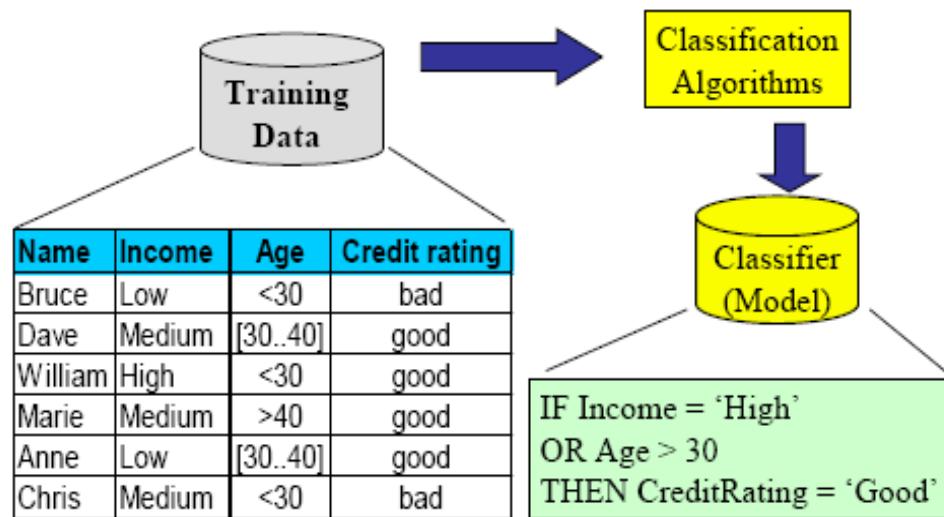
2 – Évaluation du modèle

- Estimation de la précision de la classification du modèle à l'aide des **données de test**.
- On doit définir un ensemble de données de test (les données de test doivent être différentes des données d'apprentissage);
- ✓ Avec les données de tests, on connaît l'appartenance d'un objet à une classe spécifique;
- On va utiliser les données de test pour évaluer la performance du classificateur. Contrairement aux données d'apprentissage, on doit fournir au classificateur les données de test sans les labels des classes. Le classificateur va ainsi procéder à la classification.
- Pour valider les résultats, on doit maintenant utiliser les labels qui sont disponibles pour les données de tests pour calculer la précision de classification.

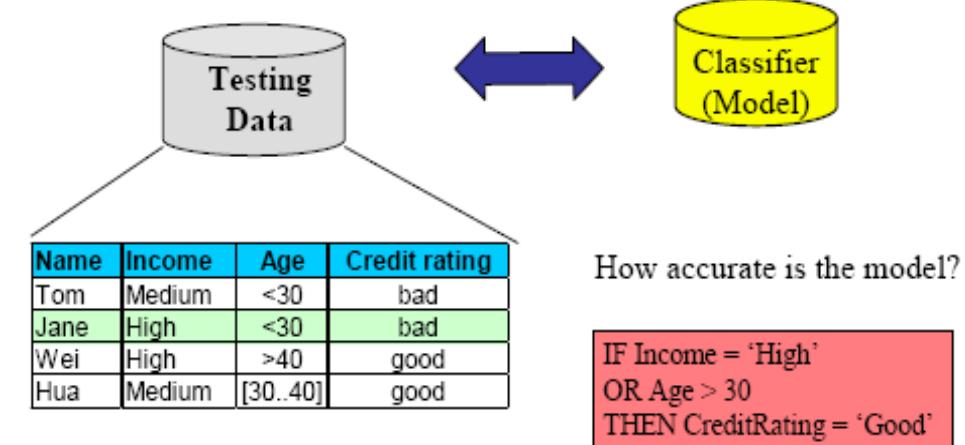
3 – Utilisation du modèle pour classer les objets nouveaux - Classification

Processus de classification

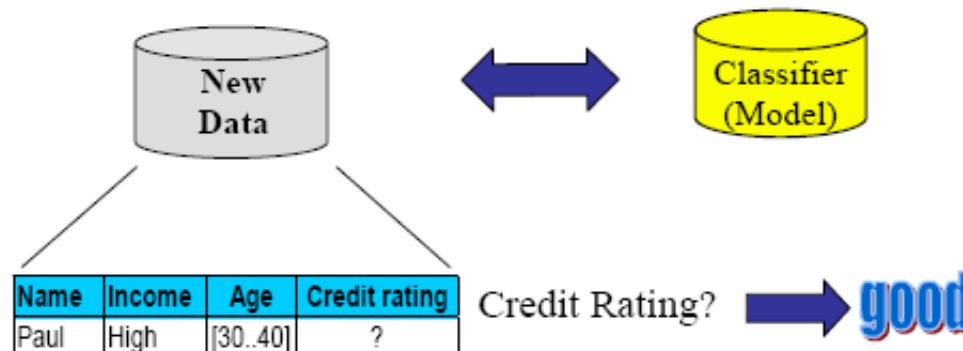
1- Apprentissage



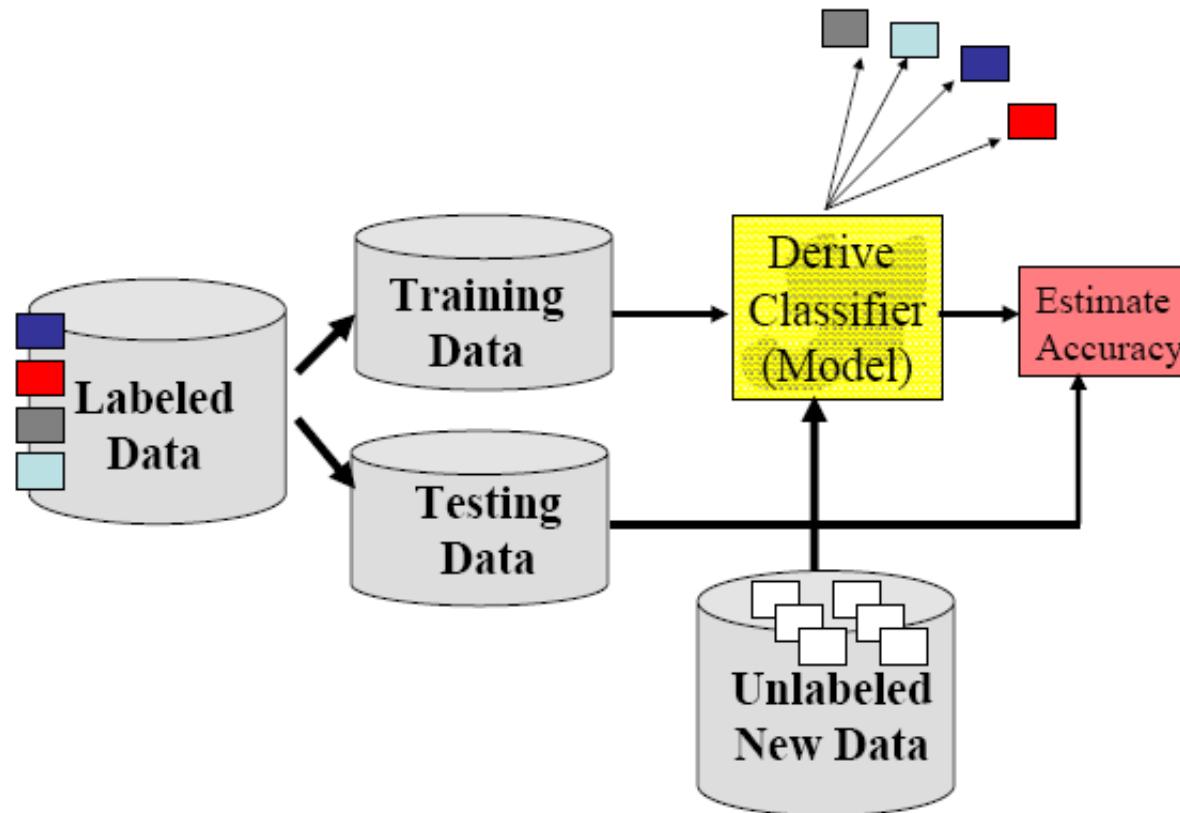
2- Évaluation



3- Classification

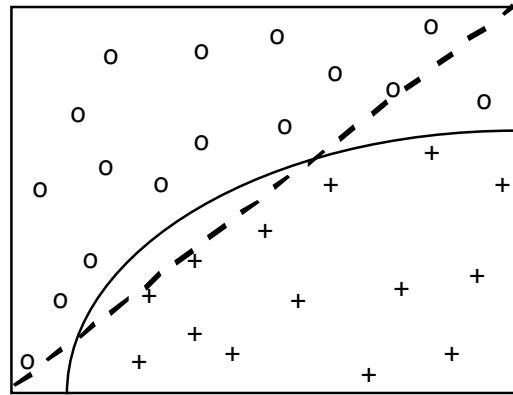


Processus de classification



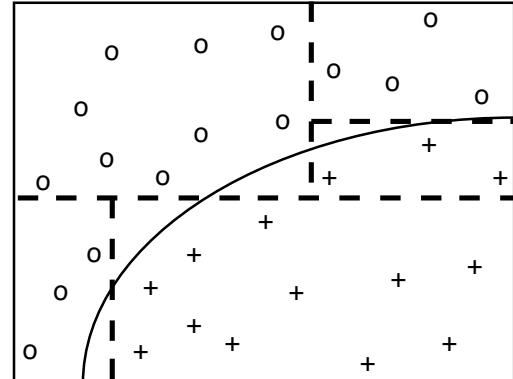
Panel de classificateurs

- Régression



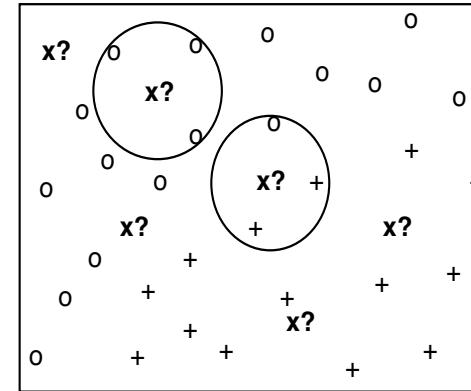
modèle linéaire

- Arbre de décision



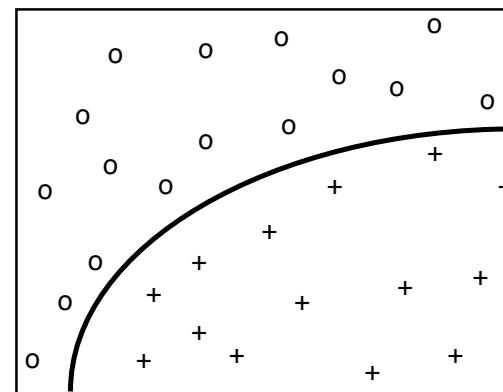
partition récursive

- Plus proches voisins



modèle local

- Réseaux de neurones multicouches



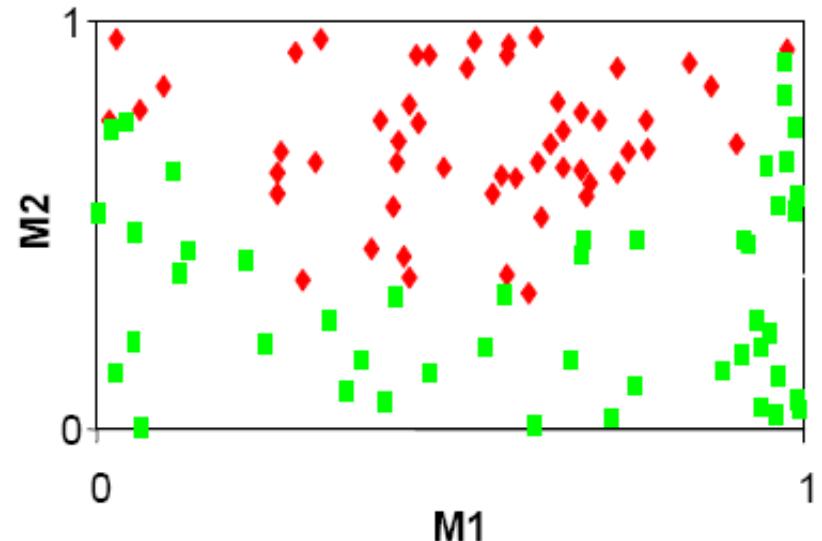
modèle non-linéaire

Sélection d'une hypothèse satisfaisante

□ Un exemple illustratif

- Données d'apprentissage : diagnostique médical à partir de deux mesures (le poids et la température)

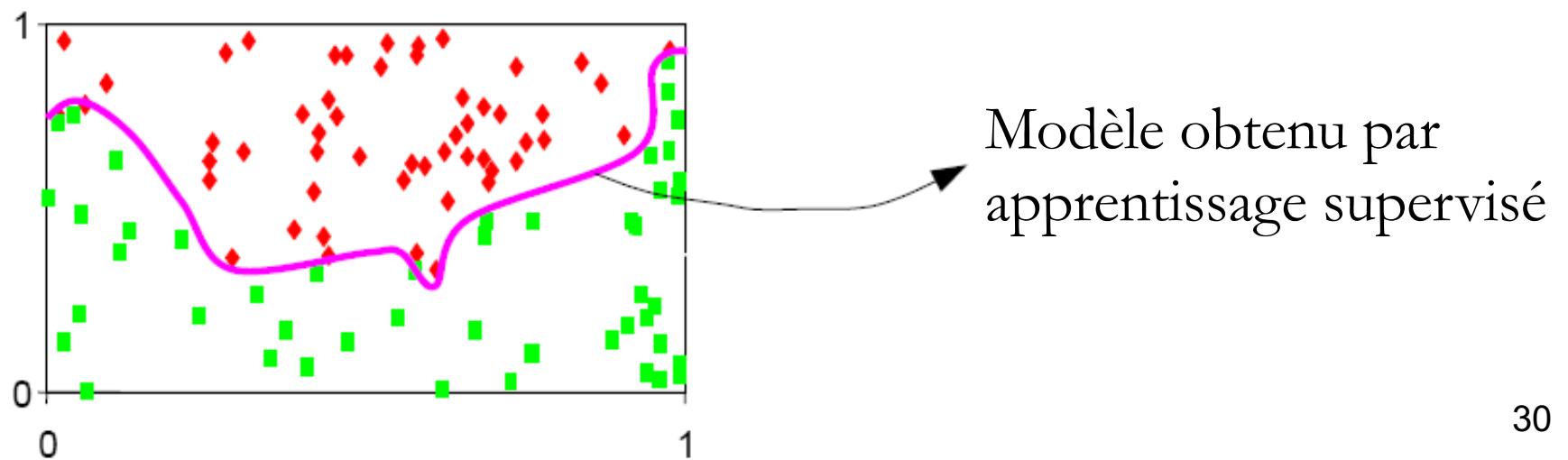
M1	M2	Y
0.52	0.18	Healthy
0.44	0.29	Disease
0.89	0.88	Healthy
0.99	0.37	Disease
...
0.95	0.47	Disease
0.29	0.09	Healthy



- But : trouver un modèle qui classifie au mieux les nouvelles données

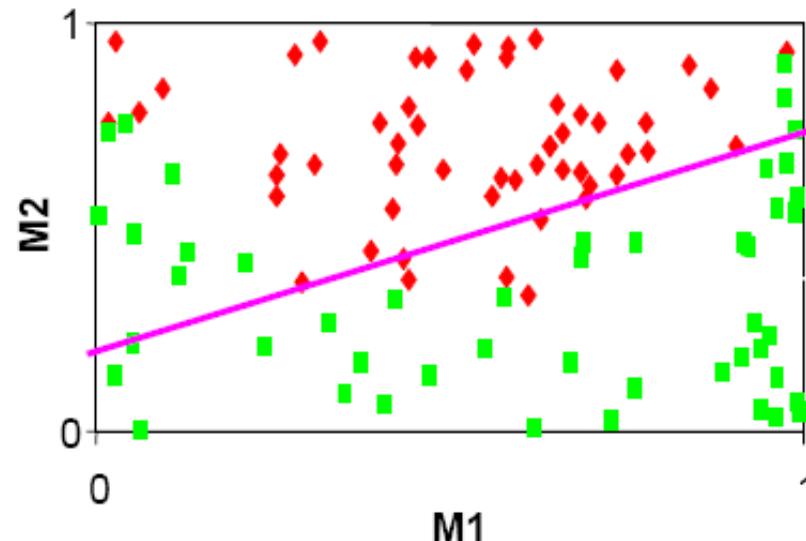
Algorithme d'apprentissage

- Un algorithme d'apprentissage est définie par :
 - Une famille des modèles candidat (l'espace des hypothèses);
 - Une mesure de qualité de la classification;
 - Une stratégie d'optimisation.
- Le résultat d'un algorithme d'apprentissage est une fonction h de H qui assure une « bonne » qualité de classification,



Modèle linéaire

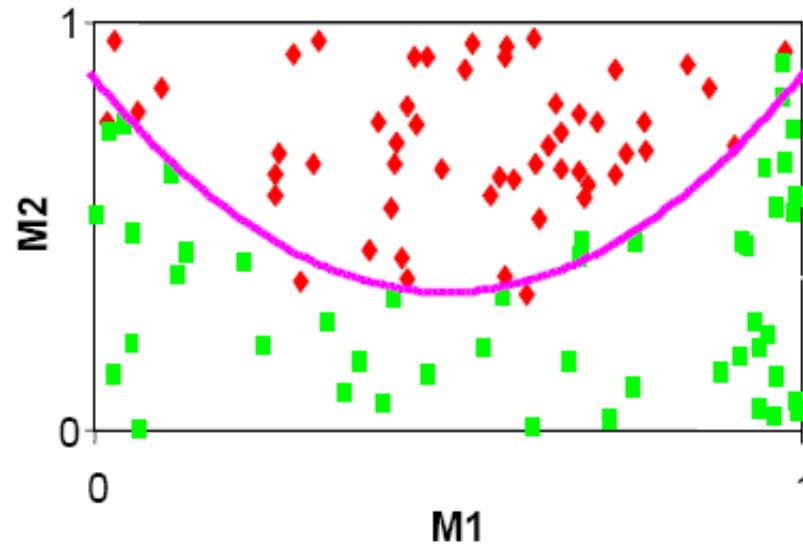
$$h(M1, M2) = \begin{cases} \text{Disease} & \text{if } w_0 + w_1 * M1 + w_2 * M2 > 0 \\ \text{Normal} & \text{otherwise} \end{cases}$$



- Phase d'apprentissage : à partir des données d'apprentissage, trouver les meilleures valeurs pour w_0 , w_1 et w_2 .
- Plusieurs alternatives pour ce modèle simple (analyse discriminante linéaire, SVM, ...)

Modèle quadratique

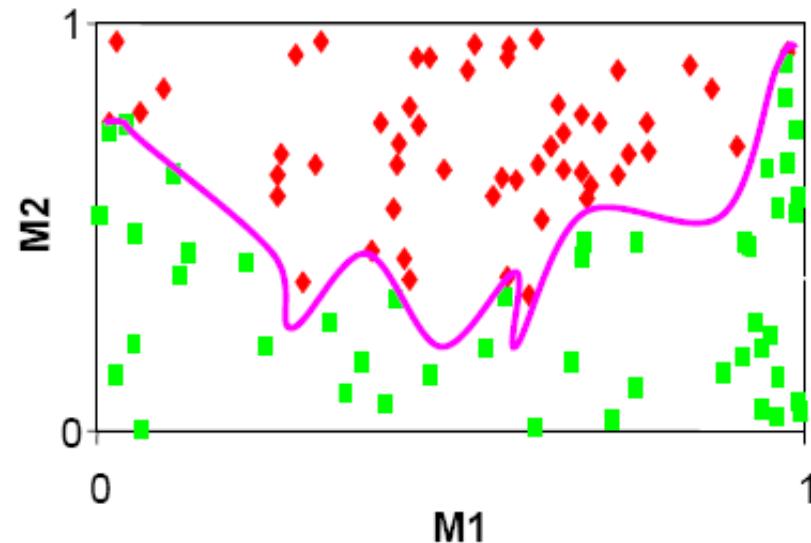
$$h(M1, M2) = \begin{cases} \text{Disease if } w_0 + w_1 * M1 + w_2 * M2 + w_3 * M1^2 + w_4 * M2^2 > 0 \\ \text{Normal otherwise} \end{cases}$$



- Phase d'apprentissage : à partir des données d'apprentissage, trouver les meilleures valeurs pour w_0 , w_1 , w_2 , w_3 et w_4 .
- Plusieurs alternatives pour ce modèle simple (Perceptron, SVM, ...)

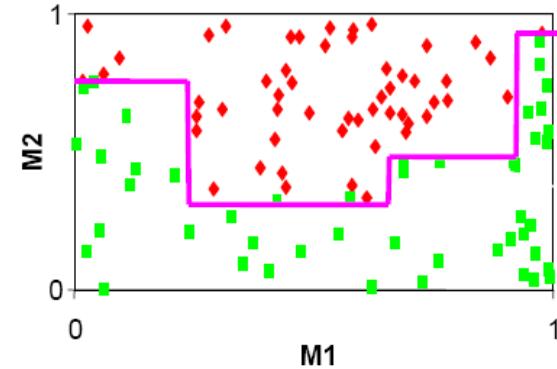
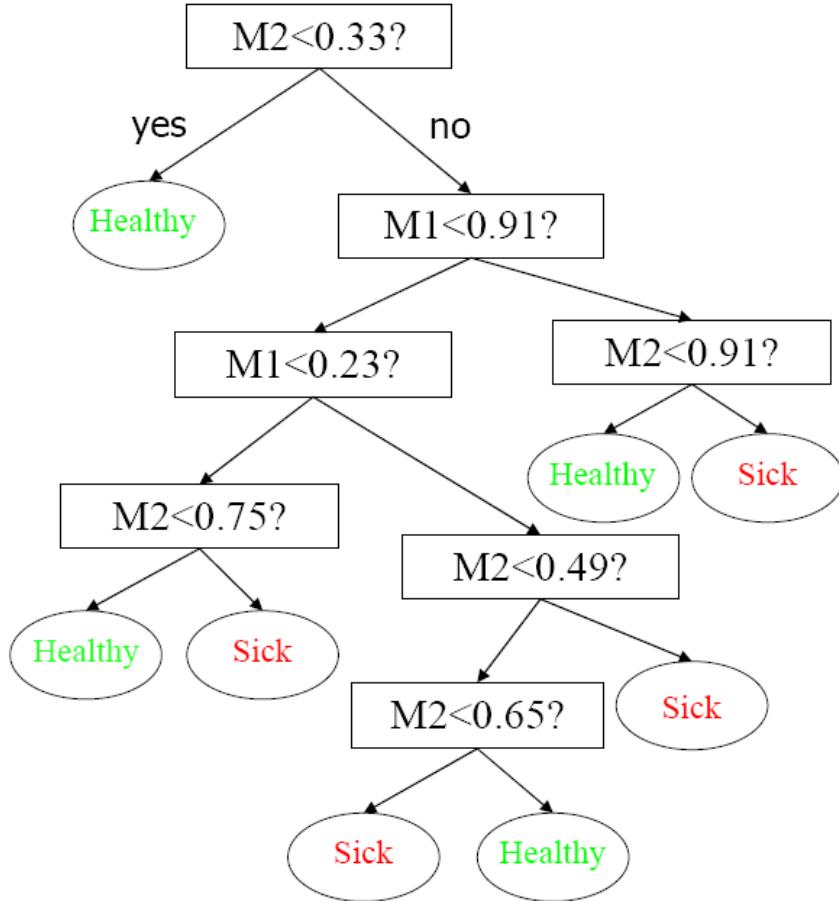
Réseaux de neurones

$h(M1, M2) = \begin{cases} \text{Disease} & \text{if } \text{à partir de quelques fonctions complexes} \\ \text{Normal} & \text{otherwise} \end{cases}$



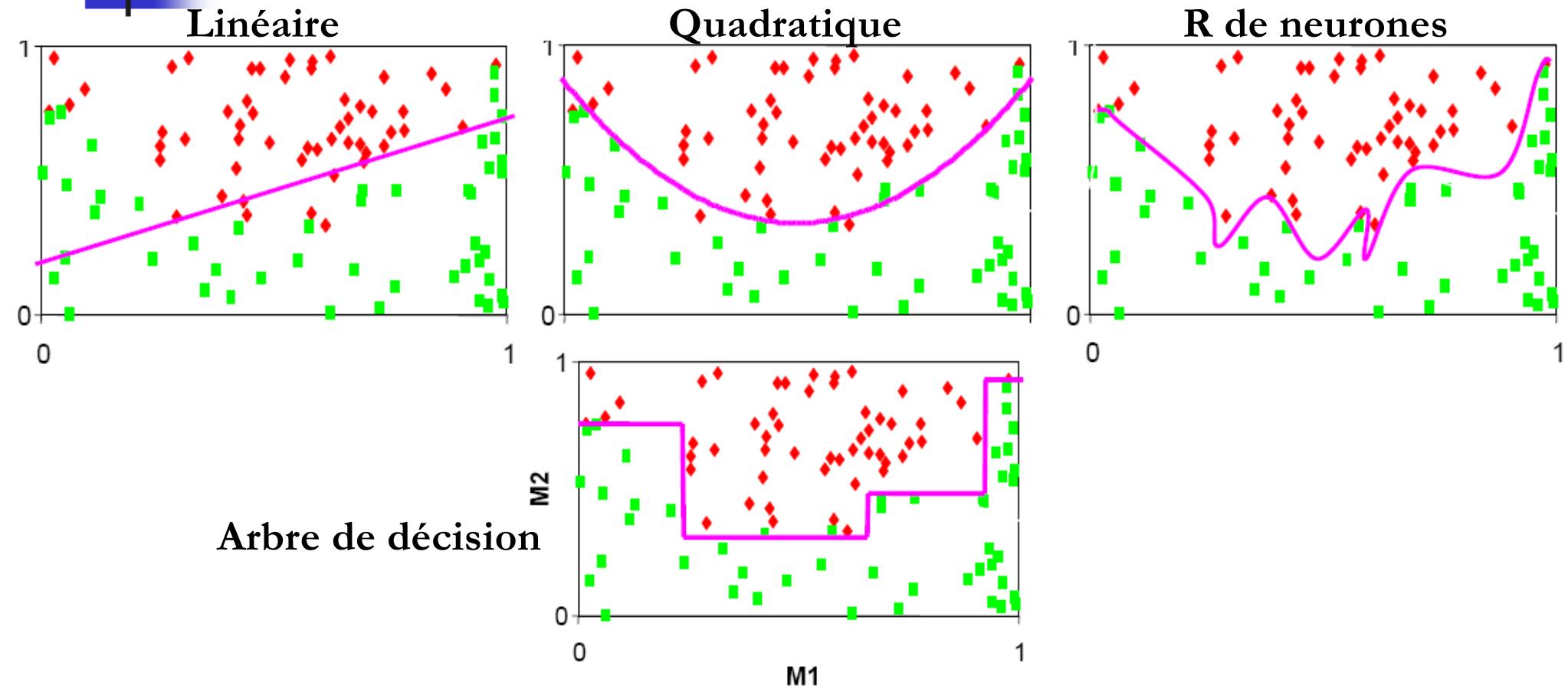
- Phase d'apprentissage : à partir des données d'apprentissage, trouver plusieurs paramètres d'une fonction complexe.

Arbre de décision



- Phase d'apprentissage : à partir des données d'apprentissage, identifier des règles de décisions qui permettent la séparation des deux classes.

Quel est le meilleur modèle (hypothèse)



- Identifier une fonction b qui à un pouvoir de généralisation et ce en minimisant une fonction d'erreur

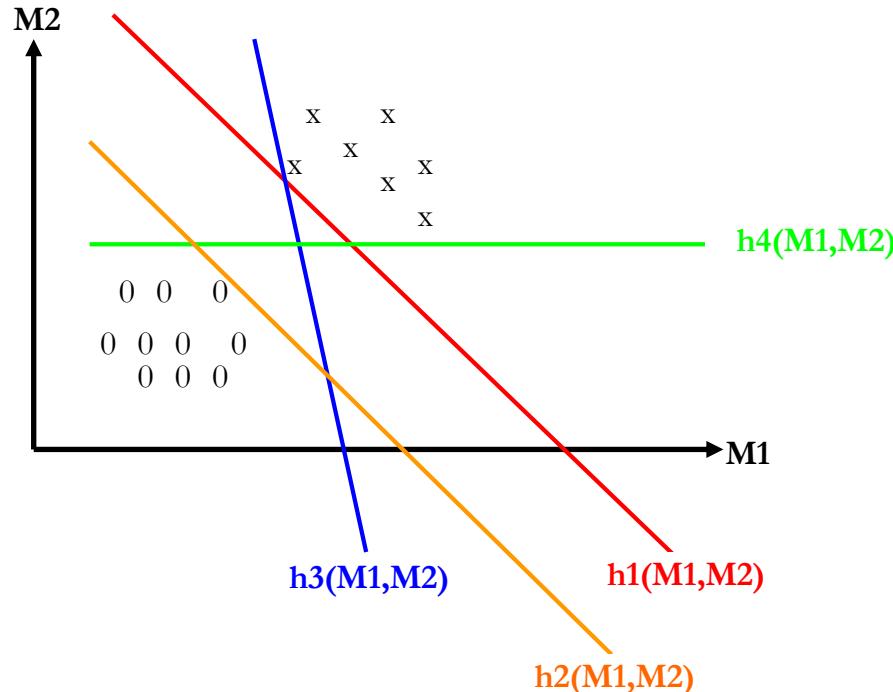


Minimiser l'erreur de généralisation

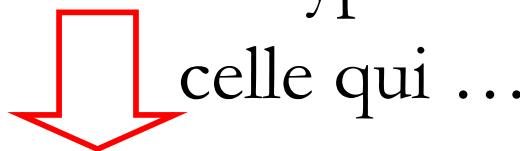
Sélection d'une hypothèse satisfaisante

□ Un autre exemple

- Pour un même problème, plusieurs d'hypothèse possibles

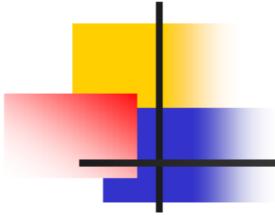


Quelle est la meilleure hypothèse



celle qui ...

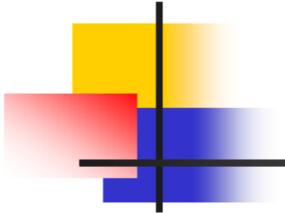
Minimise l'erreur de généralisation



Les types d'erreurs en apprentissage

- Les données d'apprentissage peuvent être bruitées, fausses, mal étiquetées;
- La taille des données d'apprentissage est limitée (échantillon non représentatif);
- L'espace H où l'on cherche une hypothèse est trop restreint.

On ne peut pas être un bon apprenant à tous les coups, mais seulement dans les situations raisonnables (échantillons d'apprentissage représentatifs) qui sont les plus probables.

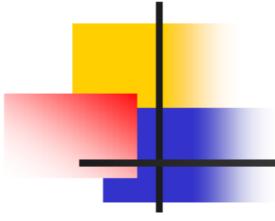


Apprentissage non supervisé

Dans l'apprentissage non supervisé il n'y a pas de notion de sortie désirée, on dispose seulement d'un nombre fini de données d'apprentissage, constituées « d'entrées », sans qu'aucun label n'y soit rattaché.

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19
-0.27	-0.15	-0.14	0.91	-0.17	0.26	-0.48	-0.1	-0.53	-0.65	0.23	0.22	0.98	0.57	0.02	-0.55	-0.32	0.28	-0.33
-2.3	-1.2	-4.5	-0.01	-0.83	0.66	0.55	0.27	-0.65	0.39	-1.3	-0.2	-3.5	0.4	0.21	-0.87	0.64	0.6	-0.29
0.41	0.77	-0.44	0	0.03	-0.82	0.17	0.54	-0.04	0.6	0.41	0.66	-0.27	-0.86	-0.92	0	0.48	0.74	0.49
0.28	-0.71	-0.82	0.27	-0.21	-0.9	0.61	-0.57	0.44	0.21	0.97	-0.27	0.74	0.2	-0.16	0.7	0.79	0.59	-0.33
-0.28	0.48	0.79	-0.14	0.8	0.28	0.75	0.26	0.3	-0.78	-0.72	0.94	-0.78	0.48	0.26	0.83	-0.88	-0.59	0.71
0.01	0.36	0.03	0.03	0.59	-0.5	0.4	-0.88	-0.53	0.95	0.15	0.31	0.06	0.37	0.66	-0.34	0.79	-0.12	0.49
-0.53	-0.8	-0.64	-0.93	-0.51	0.28	0.25	0.01	-0.94	0.96	0.25	-0.12	0.27	-0.72	-0.77	-0.31	0.44	0.58	-0.86
0.04	0.94	-0.92	-0.38	-0.07	0.98	0.1	0.19	-0.57	-0.69	-0.23	0.05	0.13	-0.28	0.98	-0.08	-0.3	-0.84	0.47
-0.88	-0.73	-0.4	0.58	0.24	0.08	-0.2	0.42	-0.61	-0.13	-0.47	-0.36	-0.37	0.95	-0.31	0.25	0.55	0.52	-0.66
-0.56	0.97	-0.93	0.91	0.36	-0.14	-0.9	0.65	0.41	-0.12	0.35	0.21	0.22	0.73	0.68	-0.65	-0.4	0.91	-0.64

Est-ce qu'il y a des structures (regroupements) intéressantes dans les données ? Est-ce qu'il ya certaine corrélation entre les attributs ? Est-ce qu'il y a des attributs qui ne sont pertinent ?



Apprentissage non supervisé

- Apprentissage non supervisé

De l'ensemble de données $S = \{x_1, x_2, \dots, x_n\}$

on cherche des régularités sous-jacente

Sous forme d'une fonction : régression

Sous forme de nuages de points (*e.g.* mixture de gaussiennes)

Sous forme d'un modèle complexe (*e.g.* réseau bayésien)

afin de résumer, découvrir des corrélations, des associations, des tendances, ...