

ENSIIE - Intelligence Artificielle (RIIA) - 1er cours

Benjamin PIWOWARSKI

28 septembre 2015

Plan

- 1 Introduction
- 2 Définitions
- 3 Les apprentissages
- 4 Les grandes questions

Introduction



video droles



video droles

video droles **d'animaux**

video droles **chats**

video droles **facebook**

[En savoir plus](#)

Vidéos Drôles, Insolites, Humour et Buzz - Koreus

<https://www.koreus.com/videos/nouveau/> ▼

Regardez les meilleures vidéos du Net sélectionnées par Koreus. Plus de 6500 vidéos drôles, insolites, vidéos humour et buzz à visionner gratuitement.

[Animaux](#) - [Images](#) - [Véhicules](#) - [Compilation & Gag](#)

la video la plus drole du monde - YouTube

www.youtube.com/watch?v=UjZDVcu6PbA ▼

21 juin 2011 - Ok bravo vous trouvez sa toute le video poche mais pourquoi vous l'avez écouter ?!!!! Pis avant de dire que vous aimer pas sa ÉCRIVEZ DONT ...

Voir toutes les vidéos - | Topito

www.topito.com/video ▼

On appelle ça des vidéos. Nous, on passe du temps pour vous à trouver les plus drôles sur le web. Et aussi à en créer des 100% Topito. Plein même.

Humour - Vidéo drole - Buzz - Insolite

www.lesdebiles.com/ ▼

Le meilleur site d'humour français. Chaque jours, de nouvelles vidéos drôles, du buzz, des fails et des vidéos insolites pour te faire passer un bon moment.

Vidéos - Vidéos virales, drôles et insolites - Ayoye - Les ...

www.ayoye.com/videos/ ▼

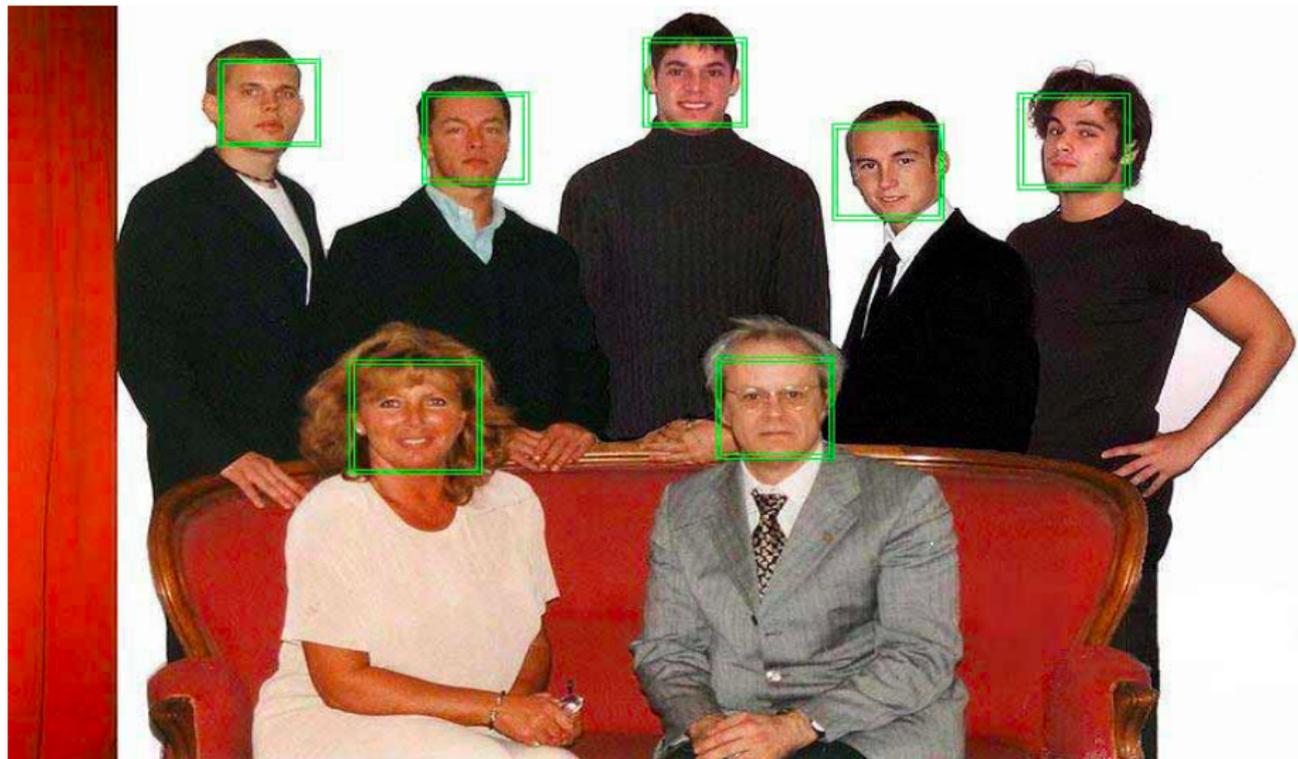
Ce bébé qui semble pourtant ordinaire a surpris tous les docteurs en ouvrant la bouche!
Vidéos - Vidéos virales, drôles et insolites - 26 septembre 2015 ...

Vidéos buzz, insolites et drôles sur Last-Video.com

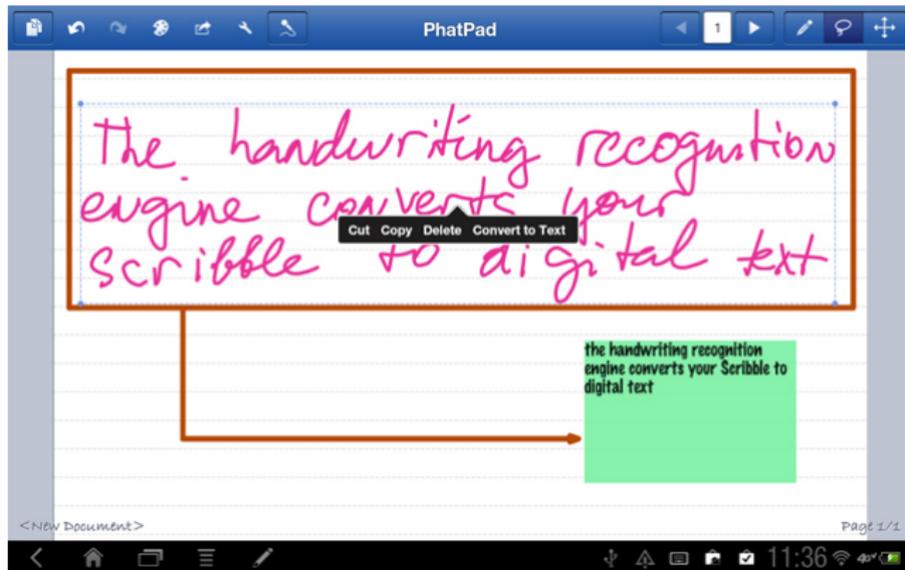
www.last-video.com/ ▼

Toutes les vidéos buzz, insolites, drôles, chocs, football choisies par Last Video. Plus de 10 000 vidéos buzz mais aussi des images humoristiques et jeux flash.

Introduction



Introduction



IT UNDERWORLD

by Phil Johnson

Another awkward FACEBOOK friend suggestion

facebook Search Home Profile Find Fri

Easter Bunny
Egg Deliverer
Lives in a Warren
Married to Mrs Easter Bunny

People You May Know

Rabbi Rosen
1 mutual friend
Add as friend

Wall
Info
Photos (25)
Notes

Share: Status Quest

What's on your mind?

www.ITworld.com PJ

Plan

- 1 Introduction
- 2 Définitions**
- 3 Les apprentissages
- 4 Les grandes questions

Intelligence Artificielle et apprentissage automatique

- Représentation des connaissances
- Résolution de problèmes
- Optimisation
- **Apprentissage automatique (*Machine Learning*)**
 - La reconnaissance des formes
 - Statistiques
 - Optimisation
 - Exploration de données

Plusieurs Définitions

- Learning is making useful changes in mind [Marvin Minsky, 1985]
- Learning is any change in a system that allows it to perform better the second time on repetition of the same task or another task drawn from the same population [Herbert Simon, 1983]
- Learning is the organization of experience [Scott, 1983]
- Learning is constructing or modifying representations of what is being experienced [Ryszard Michalski, 1986]

L'Apprentissage

- Plusieurs définition (écoles) de l'apprentissage :
 - Psychologie : comportementalisme, constructivistes, ...
 - Théories sur le déroulement/fonctionnement de l'apprentissage
 - Cognitivistes :
 - Postulat : Assimilation de l'organisme à un ordinateur, à une machine à traiter des informations
 - Apprentissage : Compréhension et développements de modèles
 - Théories écologiques :
 - Postulat : Le comportement dynamique traduit l'adaptation du système aux contraintes qui pèsent sur lui
 - Apprentissage : Intégrer de nouveaux états stables

Ordinateurs intelligents (par apprentissage)

Problématique :

- Nous souhaitons avoir des ordinateurs
 - intelligents
 - adaptatifs
 - avec un comportement robuste
- Programmer de tels comportement est souvent impossible
 - Par exemple : Intelligence artificielle dans les jeux (scripts)

Solution :

- Faire un ordinateur capable de se programmer lui-même
- à partir d'exemples (apprentissage classique / par imitation)
- à partir de son « expérience » (apprentissage par renforcement)

Quand utiliser l'apprentissage ?

- Analyse de données (capteurs, Web)
- Il n'y a pas d'expert humain
- L'utilisation d'un expert humain est trop coûteuse ou trop peu performante
- La quantité de données est telle qu'une analyse manuelle est impossible
- Les données évoluent rapidement dans le temps
- Les modèles doivent être adaptés à l'utilisateur
- Les systèmes doivent s'adapter facilement à des conditions opérationnelles différentes

Apprentissage automatique (Machine Learning)

Différents contextes d'apprentissage :

- **L'apprentissage supervisé** : un expert est employé pour étiqueter correctement des exemple



L'apprentissage semi-supervisé



- **L'apprentissage non-supervisé** : Aucun expert n'est disponible. L'algorithme doit découvrir par lui-même la structure des données.
- **L'apprentissage par renforcement** : L'algorithme apprend un comportement.

Exemples de modèles :

- Les réseaux de neurones
- Les arbres de décision
- Les modèles probabilistes
- Les machines à vecteur de support

Plan

- 1 Introduction
- 2 Définitions
- 3 Les apprentissages
 - Apprentissage supervisé
 - Apprentissage non-supervisé
 - Apprentissage par renforcement
- 4 Les grandes questions

Les apprentissages

Apprentissage supervisé

Apprentissage supervisé

On considère :

- Un espace de représentation \mathcal{X}
- Un espace de sortie \mathcal{Y}
- Un ensemble d'apprentissage $X = ((x_1, y_1), \dots, (x_n, y_n))$

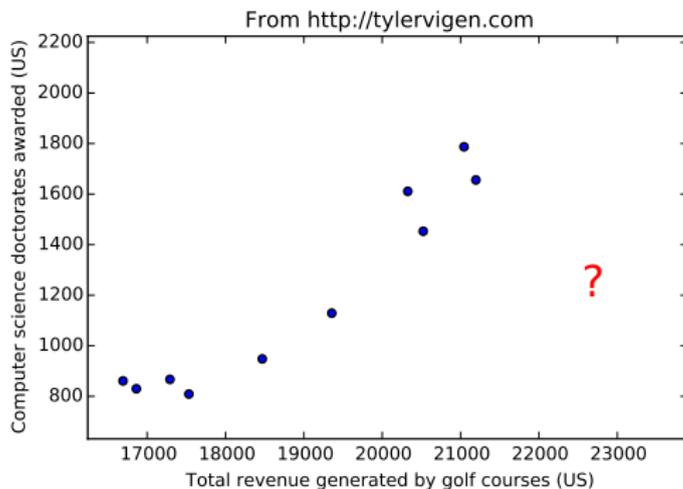
On cherche à apprendre à partir de X une fonction f_θ telle que :

- $f_\theta : \mathcal{X} \rightarrow \mathcal{Y}$
- θ est l'ensemble des paramètres (à apprendre) de la fonction
- $f_\theta(x)$ prédit l'étiquette y associée à x

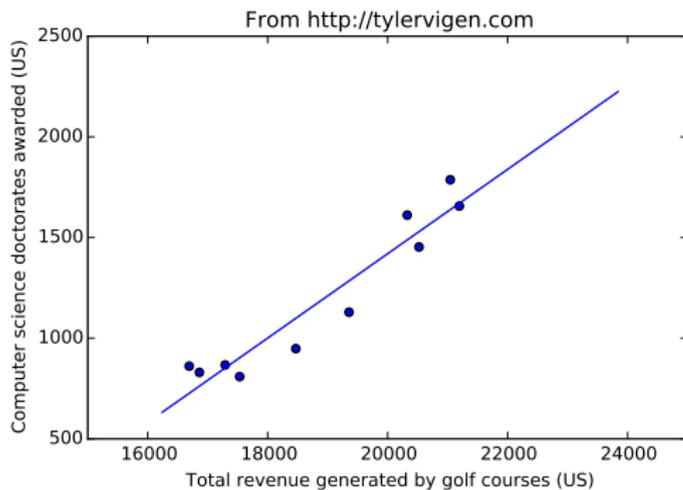
Régression

- x est un vecteur de taille n
- y est un nombre réel
- Exemples :
 - Prédiction de la température
 - Prédiction du cours de la bourse

Régression



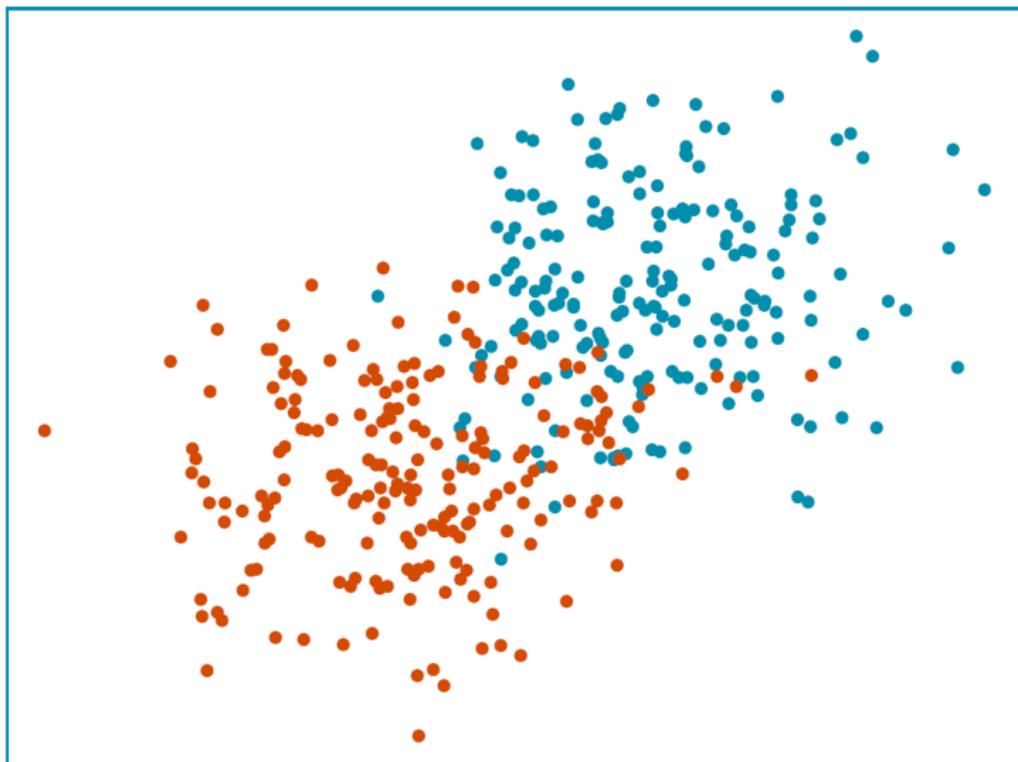
Régression



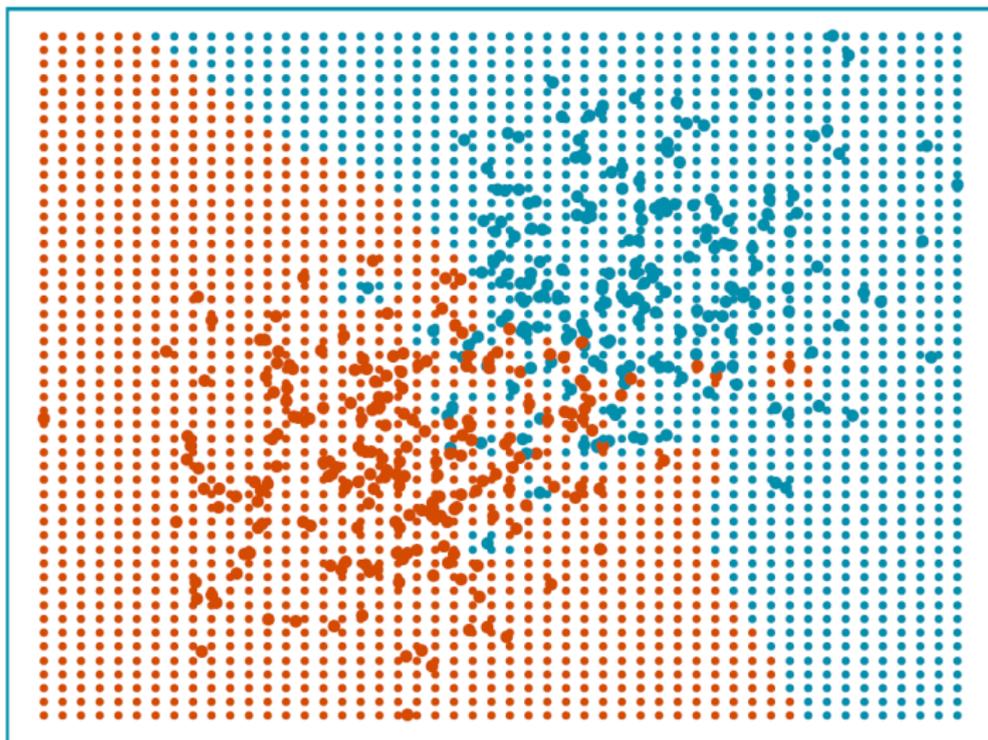
La classification

- x est un vecteur de taille n - taille 2 dans nos illustrations
- y est une catégorie :
 - soit -1 , soit $+1$ (classification binaire)
 - un nombre discret de classes
- Exemples :
 - Binaire : Classification de Spam dans les emails, hausse ou à la baisse de la bourse ; le bien, le mal...
 - Autre : classification de chiffres, ...

La classification : Ensemble d'apprentissage



La classification : Frontière de décision



Les apprentissages

Apprentissage non-supervisé

Apprentissage non-supervisé

- Un espace de représentation \mathcal{X}
- Un espace de sortie \mathcal{Y}
- Un ensemble d'apprentissage $X = (x_1, \dots, x_n)$

On cherche à apprendre à partir de X une fonction f_θ telle que :

- $f_\theta : \mathcal{X} \rightarrow \mathcal{Y}$
- θ est l'ensemble des paramètres (à apprendre) de la fonction

Exemple : Clustering

web news wikipedia jobs more »
advanced preferences

clouds sources sites time
remix

All Results (147)

- Technology (20)
- Cluster Analysis (16)
- Management, Cloud (15)
- Software, Statistical (13)
- Windows (12)
- Definition (10)
- Based (9)
- Patent, News Reporter-Staff News Editor (9)
- Global, Networks (7)
- Connect (4)
- Search, Engine (5)
- Learning, Machine (5)
- Corresponding author (4)
- SQL Server (4)
- Oracle, Analytics (3)
- Blog, Failover (3)
- Brainstorming, Ideas (3)
- Structure, Data Clustering (3)
- High Availability (3)
- History (3)
- Prewriting (2)
- Masters, Computing (2)
- Nature (2)
- County (2)

Top 147 results of at least **6,290,000** retrieved for the query **clustering** (definition) (details)

[Cluster analysis - Wikipedia, the free encyclopedia](#)

Cluster analysis or **clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some ...
[en.wikipedia.org/wiki/Data_clustering](#) - [cache] - Yippy Index IV

[Clustering | Define Clustering at Dictionary.com](#)

noun 1. a number of things of the same kind, growing or held together; a bunch: a **cluster** of grapes. 2. a group of things or persons close together: There was a ...
[dictionary.reference.com/browse/clustering](#) - [cache] - Yippy Index IV

[All of Connect | Symantec Connect](#)

... 20 o... Posted: 14 Apr 2015 Storage and **Clustering** Community Explore: Storage and **Clustering** Community Stats Forums: 5,463 Articles: 176 Blogs: ...
[www.symantec.com/connect](#) - [cache] - Yippy Index

[Computer cluster - Wikipedia, the free encyclopedia](#)

A computer **cluster** consists of a set of loosely or tightly connected computers that work together so that, in many respects, they can be viewed as a single system.
[en.wikipedia.org/wiki/Computer_cluster](#) - [cache] - Yippy Index IV

[SIOS Technology Honored as Gold Winner in the 10th Annual 2015 Hot Companies and Best Products Awards](#)

Jul 7, 2015 - SAN MATEO, CA--(Marketwired - Jul 7, 2015) - SIOS Technology Corp. (www.us.sios.com), maker of SAN and #SANless **clustering** software products, announced today that Network Products Guide, the industry's leading technology research and advisory guide, has named SIOS DataKeeper **Cluster** Edition a Gold winner of the 10th Annual 2015 Hot Companies...
[www.marketwired.com/...-10th-annual-2015-hot-companies-best-products-2036058.htm](#) - [cache] - Yippy News

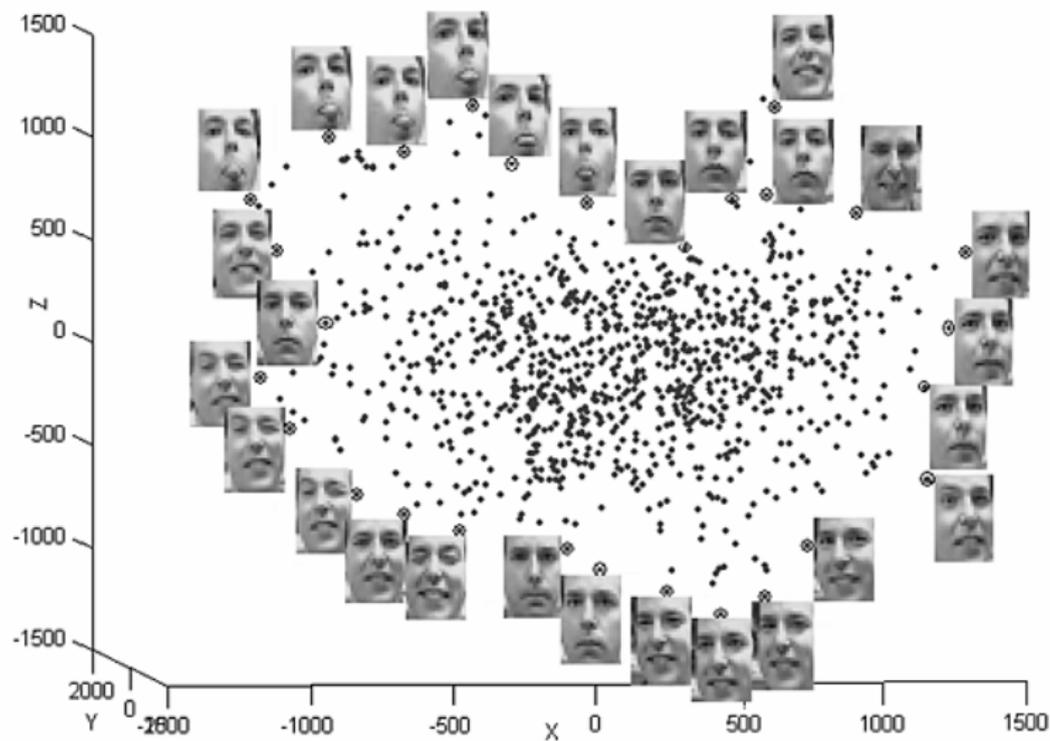
[Avago Technologies | Storage by LSI Home Page](#)

... internal connectivity Learn More 12 Minutes to Server **Clustering** AIC, Open-E and Avago deliver a complete ...
[www.lsi.com/pages/default.aspx](#) - [cache] - Yippy Index

[Writer's Web: Prewriting: Clustering](#)

Prewriting: **Clustering** Melanie Dawson & Joe Essid (printable version here) **Clustering** is a type of prewriting that allows you to explore many ideas as soon as they ...
[writing2.richmond.edu/writing/wweb/cluster.html](#) - [cache] - Yippy Index IV

Exemple : Apprentissage de représentation



Les apprentissages

Apprentissage par renforcement

Apprentissage

- 1 Apprentissage supervisé = on connaît la bonne réponse
- 2 Apprentissage non supervisé = on groupe les objets
- 3 Apprentissage semi-supervisé = ...
- 4 Apprentissage par renforcement = la supervision (« récompense ») arrive après une séquence de décision
 - 1 Un cadre d'apprentissage *naturel*
 - 2 Plus dur que l'apprentissage supervisé

Contexte

Ce contexte décrit des situations très générales est diverses :

- Un robot qui atterrit sur la lune
- Un logiciel qui joue eux échecs (au go, au backgammon, au poker,...)
- Un hélicoptère qui apprend à voler
- Un joueur de foot
- ...

Cadre théorique : Markov Decision Process

Nous définissons un MDP comme un quadruplet $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}\}$:

States \mathcal{S} est l'ensemble d'états

Actions \mathcal{A} est l'ensemble des actions. On note $\mathcal{A}(s)$ l'ensemble des actions dans l'état s tel que $\mathcal{A}(s) \in \mathcal{A}$

Probability \mathcal{P} est la fonction de transition : $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0; 1]$

$$\mathcal{P}(s, a, s') = P(s' = s_{t+1} | s = s_t, a_t = a)$$

Reward \mathcal{R} est la fonction de récompense (*reward*).

$\mathcal{R} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow R$ telle que

$$R(s, a, s') = E[r_t | s_t = s, s_{t+1} = s', a_t = a]$$

Plan

1 Introduction

2 Définitions

3 Les apprentissages

4 Les grandes questions

- Question 1 : Quelle fonction utiliser ?
- Question 2 : Comment trouver la meilleure fonction f ?
- Comment évaluer la qualité de f_θ ?

Apprentissage

Question 1

Quelle fonction f_θ utiliser - quelle hypothèse ?

Question 2

Comment trouver les valeurs de θ les meilleures ?

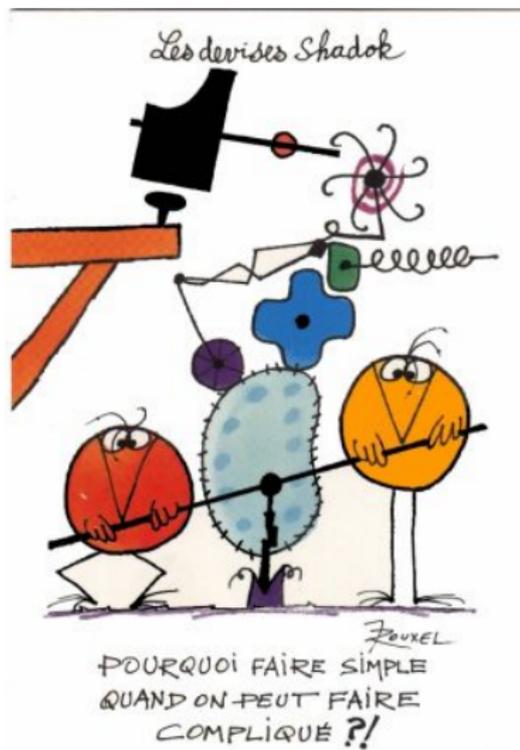
Question 3

Quelle est la qualité du modèle trouvé ?

Les grandes questions

Question 1 : Quelle fonction utiliser ?

Introduction



Apprentissage

Plein de modèles existent

- **KNN**
- **Arbres de décision**
- **Réseaux de neurones**
- **Machines à vecteurs de support**
- **Réseaux Bayésiens**

Apprentissage

Plein de modèles existent

Les différences entre les modèles existants sont :

- Complexité (mémoire / CPU)
- Type d'entrée/sortie (vecteur, séquences, structures, ...)
- Méthode d'optimisation sous-jacente (optimisation convexe, optimisation non-convexe)
- Performance

Fonctions paramétriques

- Dans la majorité des cas, \mathcal{F} est une famille paramétrique
 $\Rightarrow \theta \in \mathbb{R}^n$ définit totalement une fonction dans \mathcal{F}

Polynomes de degré n

$$\theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_n x^n$$

Les grandes questions

Question 2 : Comment trouver la meilleure fonction f ?

Critère d'apprentissage

Critère d'apprentissage

Principe : Trouver les paramètres qui, sur l'ensemble d'apprentissage minimisent (ou maximisent) un critère donné (problème d'optimisation).

$$\theta^* = \operatorname{argmin}_{f \in \mathcal{F}} R(f, X)$$

- Quel critère ?
 - Maximum de vraisemblance
 - Risque (fonction de coût)
 - ...

Note : Risque empirique et théorique

- Risque théorique : c'est ce qu'on voudrait connaître...

$$R_0(f) = \int_{x,y} \Delta(f(x), y) p(x, y) dx$$
$$f_0 = \operatorname{argmin}_{f \in \mathcal{F}} R_0(f)$$

- Risque empirique : c'est ce qu'on observe !

$$R(f, X) = \frac{1}{N} \sum_{i=1}^N \Delta(f(x_i), y_i)$$
$$f_* = \operatorname{argmin}_{f \in \mathcal{F}} R_{\text{emp}}(f)$$

- Conséquence : il faut contrôler l'apprentissage...

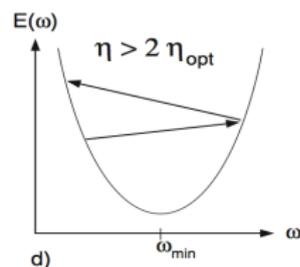
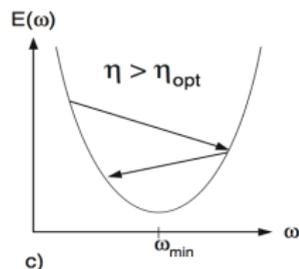
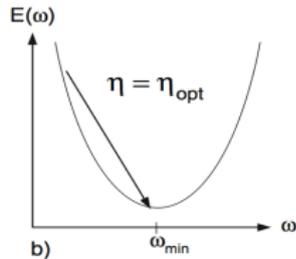
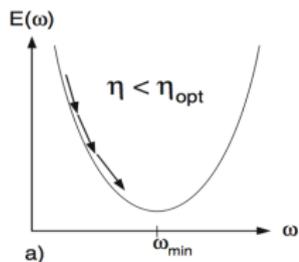
Cas des fonctions paramétriques

- Si \mathcal{F} est une famille paramétrique... le coût s'écrit $R(\theta, X)$, ou plus simplement $R(\theta)$
on peut trouver les paramètres θ par descente de gradient

$$\nabla_{\theta} R(\theta^*) = 0 \Leftrightarrow \theta^* \text{ est un extremum (ou un col)}$$

- Principe itératif :
 - on part de θ_0
 - à chaque itération, $\theta_k = \theta_{k-1} - \epsilon \nabla_{\theta} R(\theta_{k-1})$

Trouver les meilleurs θ (descente de gradient)



Les grandes questions

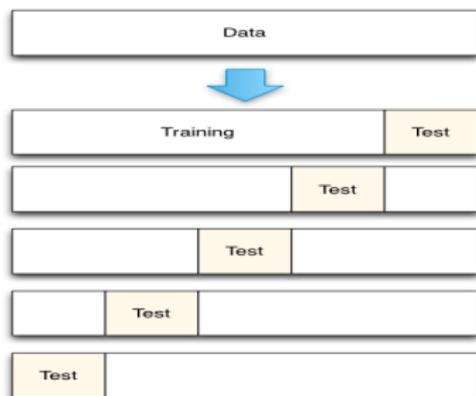
Comment évaluer la qualité de f_θ ?

Evaluation

Problème non trivial

- Evaluation sur les données d'apprentissage ? **NON**
- Evaluation sur des nouvelles données étiquetées :
 - Prendre la base d'apprentissage
 - La couper en deux
 - Entraîner sur une partie (train)
 - Tester sur une autre (test)
- Problème : réduction du nombre d'exemples d'apprentissage

Validation croisée



Validation croisée

La validation croisée ("cross-validation") est une méthode d'estimation de fiabilité d'un modèle fondé sur une technique d'échantillonnage

- Valeurs typiques : $K=3, 10$
- Cas extrême : leave-one-out

Hyper-paramètres

C'est quoi ?

- Degré maximum d'un polynôme
 - Pas de gradient
 - ...
-
- On ne peut pas utiliser le jeu de test !
 - Découpage du jeu d'entraînement en deux :
 - 1 Jeu d'entraînement du modèle
 - 2 Jeu de validation

Sous- et sur-apprentissage

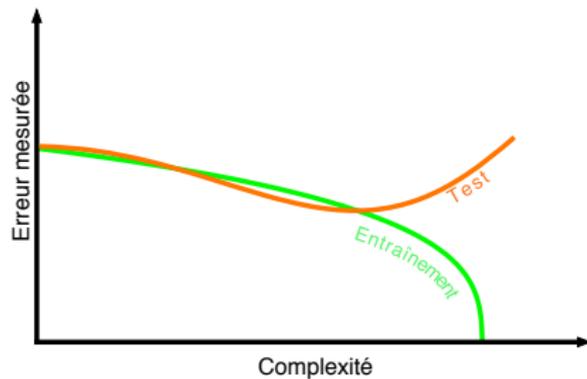
Sous-apprentissage

Le sous-apprentissage est le fait qu'un modèle n'a pas la capacité à bien modéliser les données d'apprentissage

Sur-apprentissage

Le **surapprentissage** ou overfitting (*apprentissage par coeur*) provient du fait que de par sa trop grande capacité à stocker des informations, une structure dans une situation de surapprentissage aura de la peine à généraliser les caractéristiques des données.

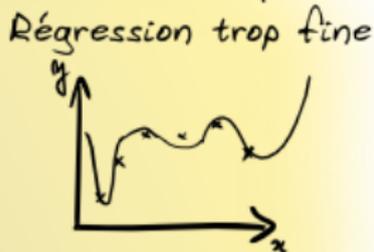
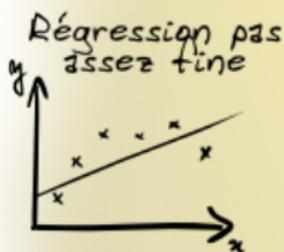
Sous- et sur-apprentissage



Différents types de fonction

RÉGRESSION

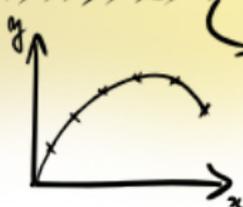
Estimation d'un comportement global à l'aide de données locales incomplètes.



Régression du mec qui s'est planté de formule

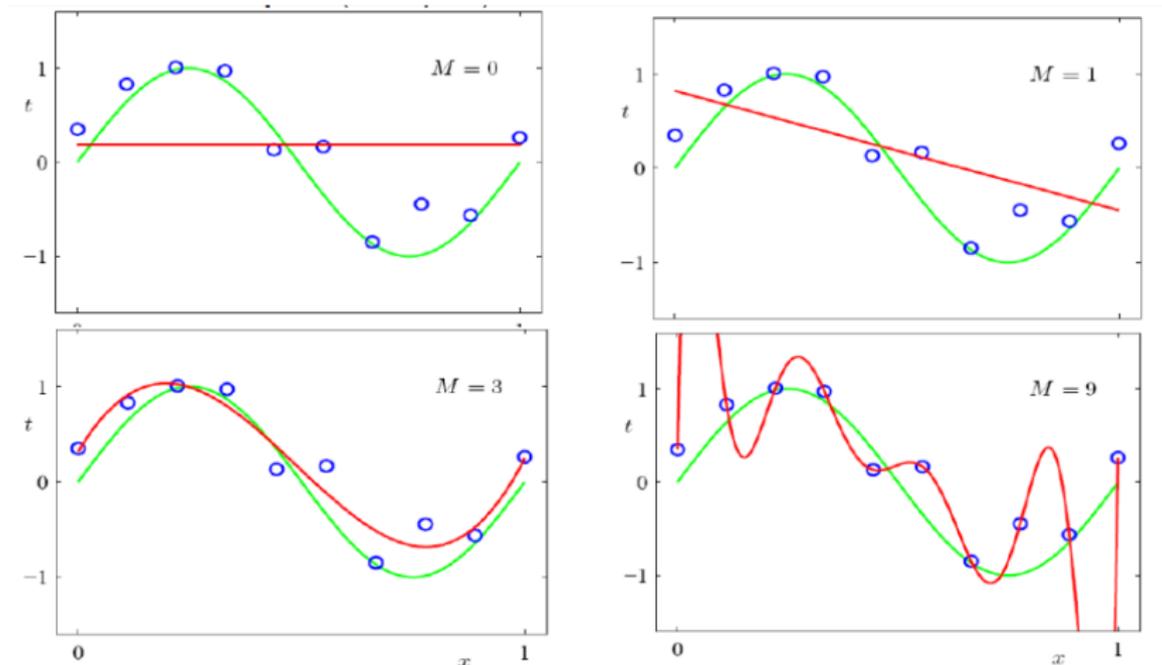


~~Bonne régression~~

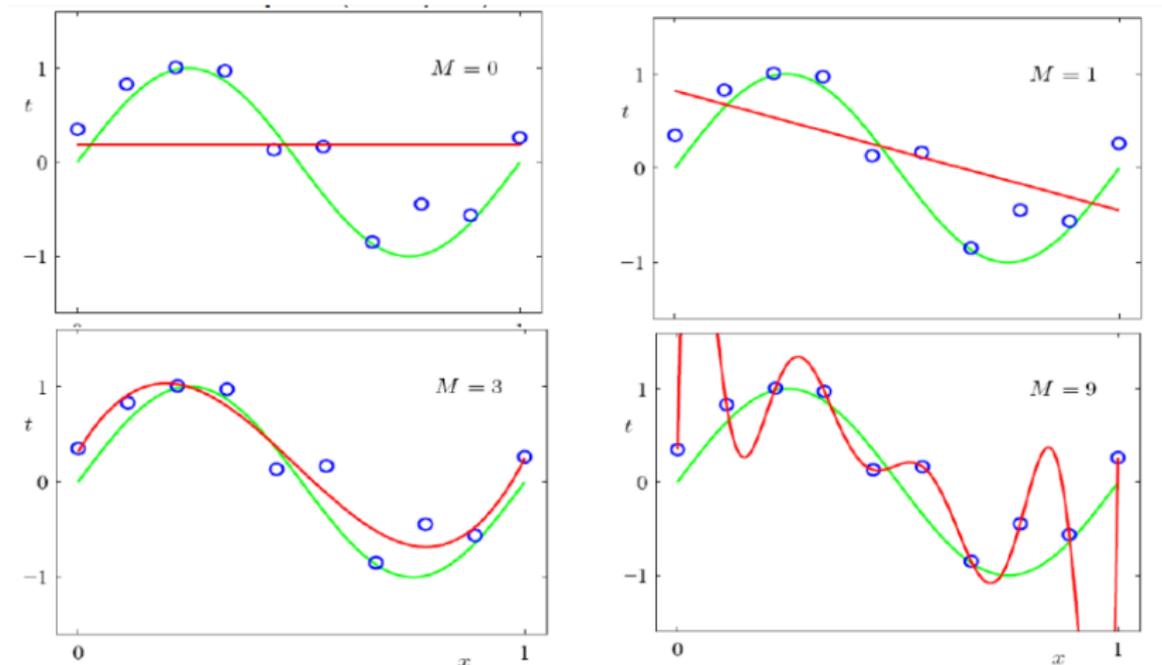


Régression du mec qui bidonne ses résultats pour faire tomber les points pile sur la courbe

Sur/Sous apprentissage



Sur/Sous apprentissage

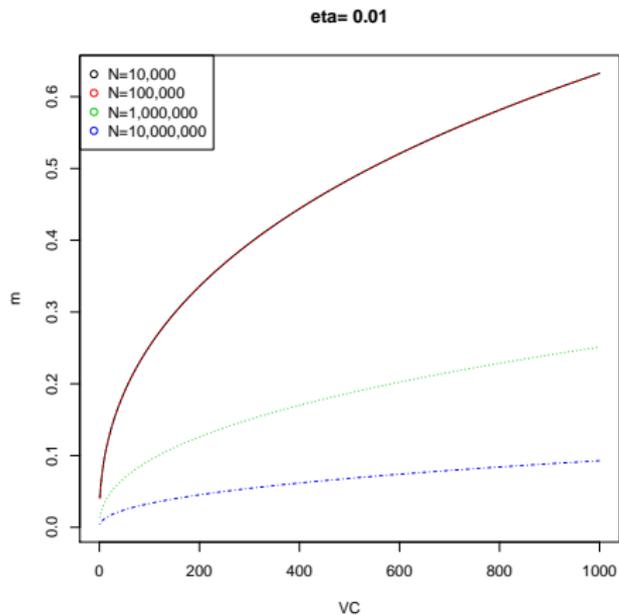


Théorie de Vapnik-Chervonenkis (VC)

- Théorie mathématique et informatique (1960-1990, Vladimir Vapnik et Alexey Chervonenkis) : contrôle de la capacité d'apprentissage automatique
- Principal concept = **la dimension VC**, le plus grand nombre de points que le modèle peut apprendre à classifier correctement (quelque soit leurs labels)
- Principal résultat théorique, la **borne sur l'erreur de test** (N nombre d'exemples $\gg h$ la dimension VC, η quelconque)

$$P \left(\text{test error} \leq \text{training error} + \sqrt{\frac{h(\log(2N/h) + 1) - \log(\eta/4)}{N}} \right) = 1 - \eta$$

Théorie de Vapnik-Chervonenkis (VC)



Plan

5 Pour finir...

Qui suis-je ?

- Benjamin Piwowarski
- Chercheur CNRS au LIP6 (Université Paris 6)
- Spécialisé en accès à l'information
 - Applications de l'apprentissage statistique
 - Méthodes d'apprentissage de représentation pour le texte & les graphes

But du cours

- Pour vous :
 - Connaissance que ça existe
 - Connaissance des champs d'applications
 - Connaissance des problématiques de base
 - Connaissance des modèles de base
- Pour moi :
 - Aller à Evry... ?
 - Trouver des stagiaires
 - Trouver des thésards

Plan du cours

- 3 jours
 - ① Introduction, K-NN, arbres de décision
 - ② Réseaux de neurones
 - ③ Apprentissage de représentation
- Evaluation
 - 1 rendu de TP (jours 1 & 2)
 - Une présentation d'articles (jour 3)