

Webinar



sesstim.univ-amu.fr

openhealth-institute.org

Cécile CAPPONI

Université d'Aix-Marseille, UFR Sciences,
Laboratoire d'Informatique et Systèmes (UMR 7020)

Deep Learning for Health Apprentissage profond pour la santé (1^{ère} Partie)

février 2018



Cliquez ici pour voir l'intégralité des ressources associées à ce document

Deep Learning for Health

Apprentissage profond pour la santé

1^{ère} Partie

Cécile CAPPONI
(avec l'aide de Thierry Artières)

LIS, Aix-Marseille Université, CNRS
Equipe QARMA

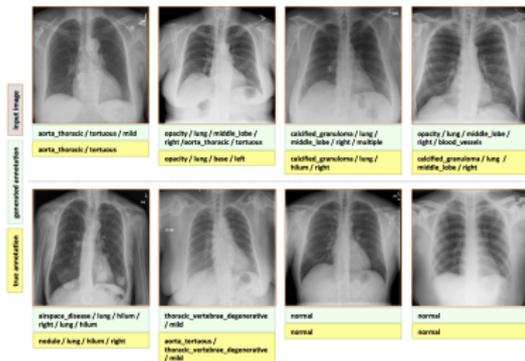


Webinar SESSTIM / OHI – 02 février 2018

I think health care 10 years from now will use a lot more AI and will look very different than it does today. Andrew Ng (Stanford, Baidu, 2017)



(source : iRythm)

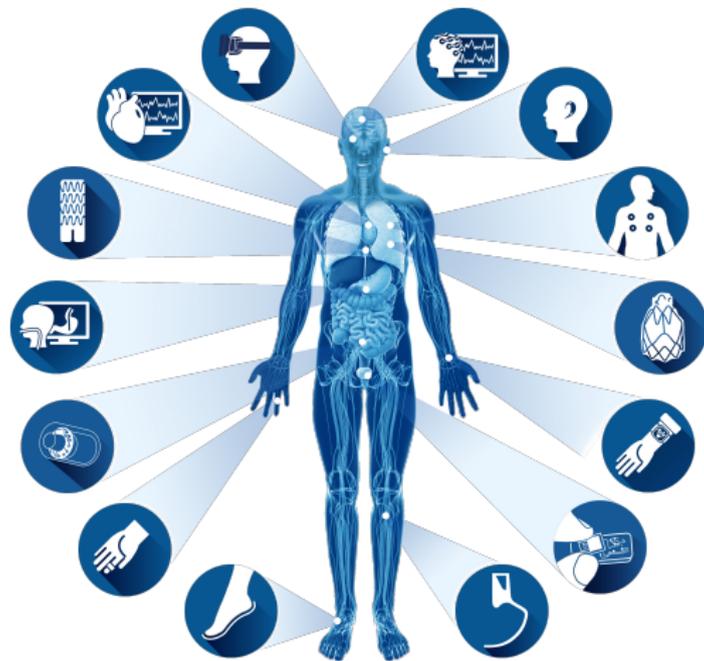


(source : nvidia)

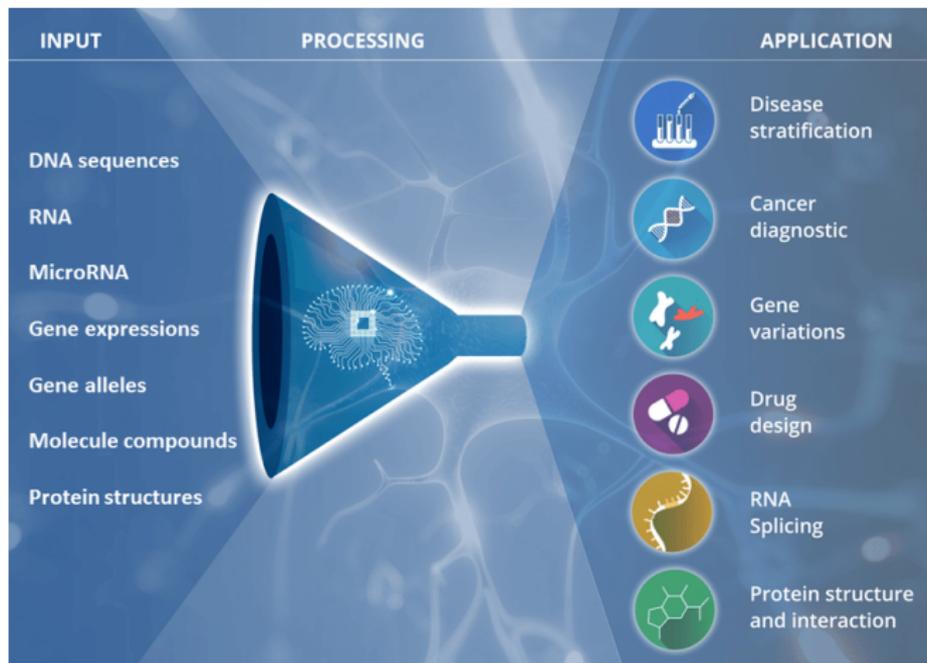
“This machine-learning approach has the potential to improve the accuracy and quality of medical care.”

— Michael Chiang, a professor of ophthalmology and a clinician at Oregon Health & Science University’s Casey Eye Institute

MIT Technology Review



(crédits : Daniele Ravi)



(crédits : Daniele Ravi)

Plan

1 Machine Learning

2 Deep Learning

3 Deep health

4 Conclusion

Informatique, statistique, intelligence artificielle

Développement, analyses formelle et expérimentale, et implémentation de méthodes qui **permettent à une machine d'évoluer dans son environnement** grâce à un processus d'apprentissage

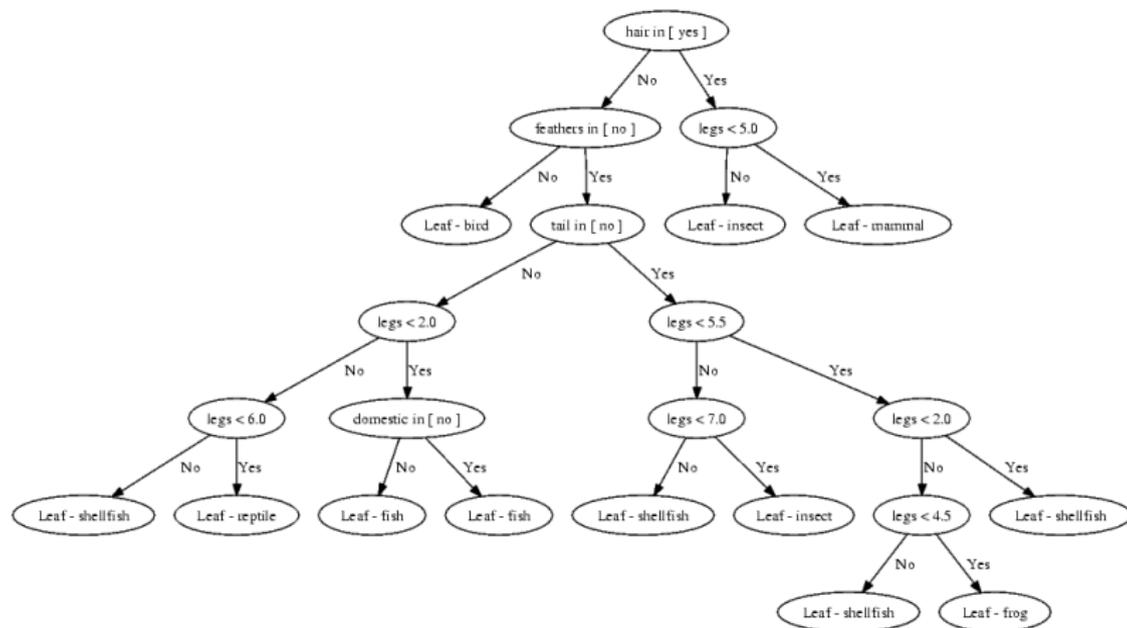
- Traitement automatique des observations de l'environnement
- Acquisition de connaissances
- Minimisation des erreurs

Machines, robots, mobiles et programmes

Aide à la décision à partir d'observations

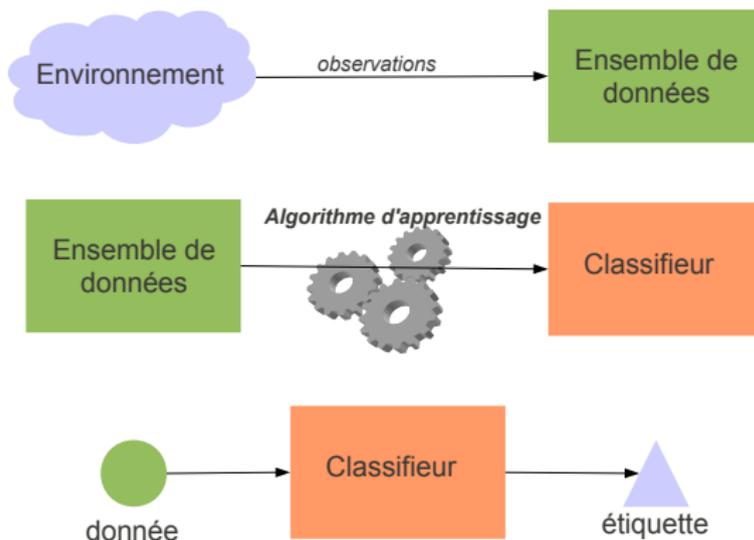
- Vision et reconnaissance d'objets
- Reconnaissance de caractères manuscrits
- Classification automatique d'appels téléphoniques
- **Aide au diagnostic médical**
- Fouille du web, moteurs de recherche
- Paramétrage automatique de systèmes complexes
- Indexation d'images

Exemple : classification d'animaux par arbre de décision



- Arbre construit à partir de données observées
- Proscrire l'apprentissage par cœur, limiter les erreurs
- **Ordre des questions** : gain en information des attributs

Schéma général de la classification



Jargon mathématique classique

Un échantillon fini de données $\mathcal{S} = \{(x_1, y_1), \dots, (x_n, y_n)\}$

où $x_i \in \mathcal{X}$ l'espace des caractéristiques (x_i est souvent un vecteur de caractéristiques)

et $y_i \in \mathcal{Y}$ les étiquettes (en classification binaire, $\mathcal{Y} = \{-1, +1\}$)

Le cas des champignons

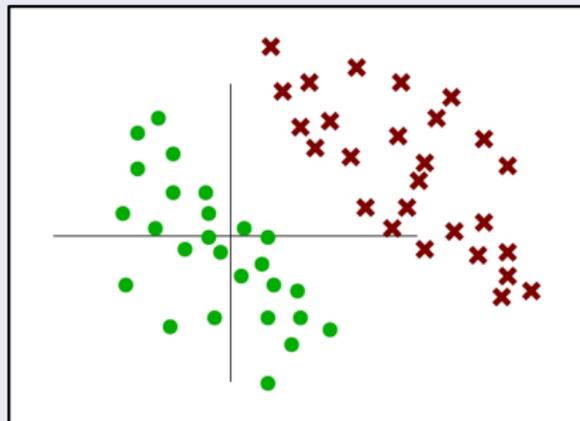
(8124 exemples, 22 attributs, vénéneux ou comestibles ?)



| | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| p | x | s | n | t | ... | k | s | u |
| e | x | s | y | t | ... | n | n | g |
| e | b | s | ? | t | ... | n | n | m |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| p | x | y | w | t | ... | k | s | u |
| e | x | s | g | f | ... | n | ? | g |
| e | x | y | y | t | ... | k | n | g |

Apprendre une fonction de classification (une *hypothèse*)

Exemple de séparation (linéaire)

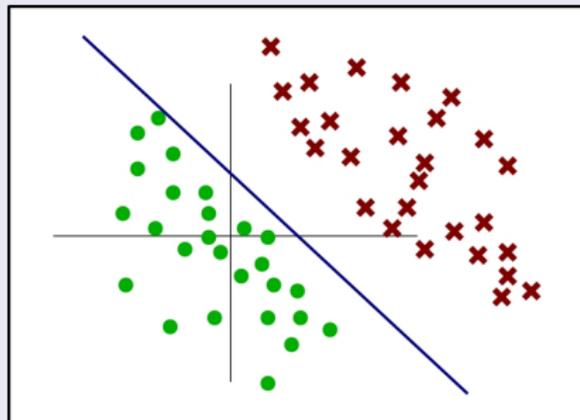


Quelques problèmes

- L'espace d'hypothèses peut être infini
- Bonnes performances en généralisation
- Qualité de l'échantillon

Apprendre une fonction de classification (une *hypothèse*)

Exemple de séparation (linéaire)

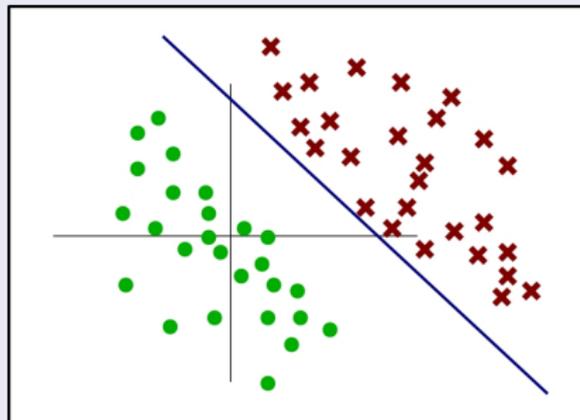


Quelques problèmes

- L'espace d'hypothèses peut être infini
- Bonnes performances en généralisation
- Qualité de l'échantillon

Apprendre une fonction de classification (une *hypothèse*)

Exemple de séparation (linéaire)

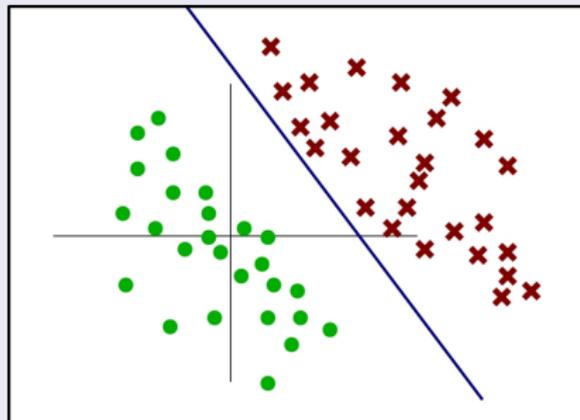


Quelques problèmes

- L'espace d'hypothèses peut être infini
- Bonnes performances en généralisation
- Qualité de l'échantillon

Apprendre une fonction de classification (une *hypothèse*)

Exemple de séparation (linéaire)

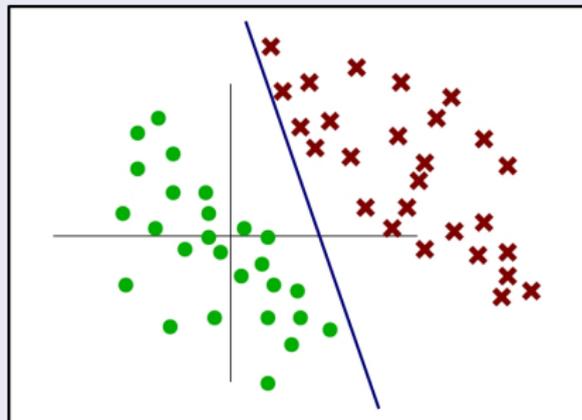


Quelques problèmes

- L'espace d'hypothèses peut être infini
- Bonnes performances en généralisation
- Qualité de l'échantillon

Apprendre une fonction de classification (une *hypothèse*)

Exemple de séparation (linéaire)

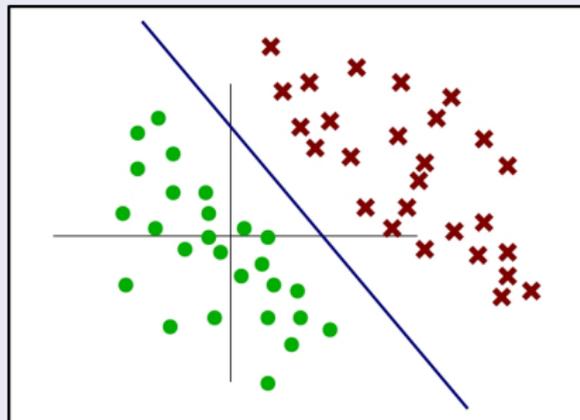


Quelques problèmes

- L'espace d'hypothèses peut être infini
- Bonnes performances en généralisation
- Qualité de l'échantillon

Apprendre une fonction de classification (une *hypothèse*)

Exemple de séparation (linéaire)

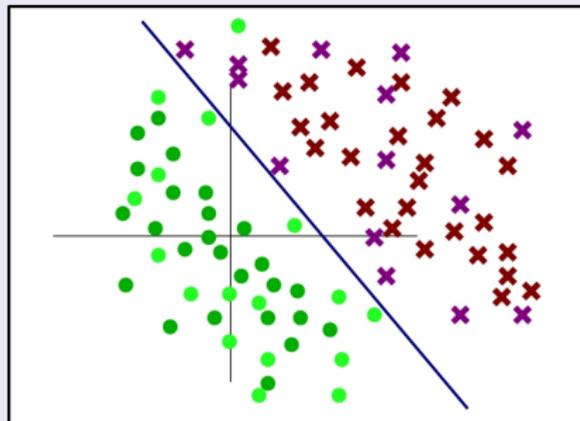


Quelques problèmes

- L'espace d'hypothèses peut être infini
- Bonnes performances en généralisation
- Qualité de l'échantillon

Apprendre une fonction de classification (une *hypothèse*)

Exemple de séparation (linéaire)

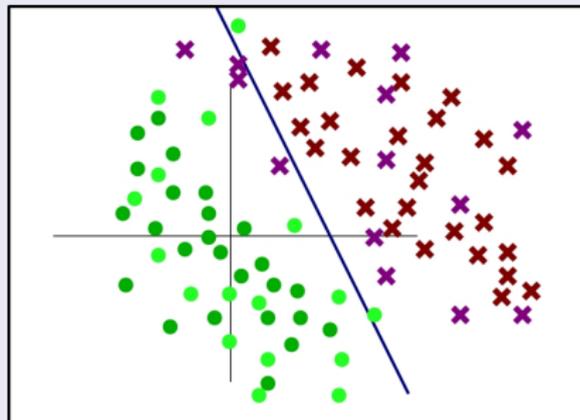


Quelques problèmes

- L'espace d'hypothèses peut être infini
- Bonnes performances en généralisation
- Qualité de l'échantillon

Apprendre une fonction de classification (une *hypothèse*)

Exemple de séparation (linéaire)

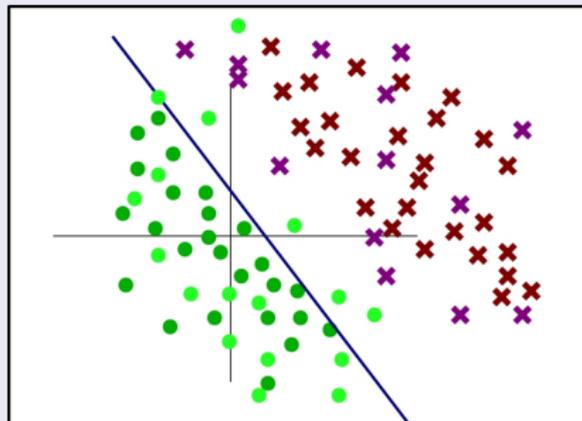


Quelques problèmes

- L'espace d'hypothèses peut être infini
- Bonnes performances en généralisation
- Qualité de l'échantillon

Apprendre une fonction de classification (une *hypothèse*)

Exemple de séparation (linéaire)

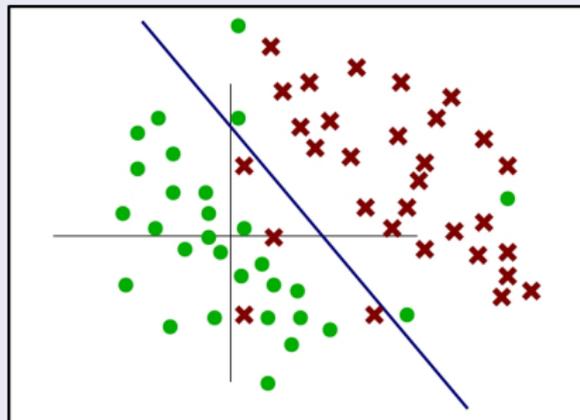


Quelques problèmes

- L'espace d'hypothèses peut être infini
- Bonnes performances en généralisation
- Qualité de l'échantillon

Apprendre une fonction de classification (une *hypothèse*)

Exemple de séparation (linéaire)

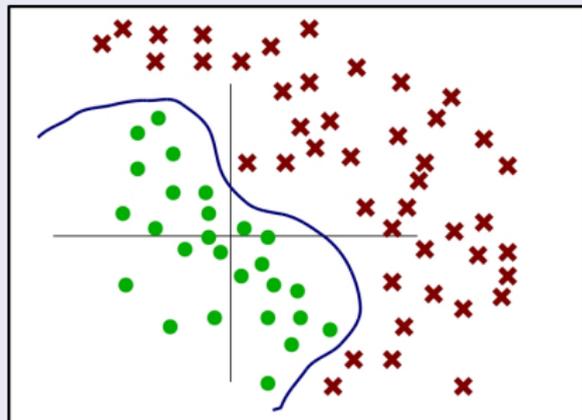


Quelques problèmes

- L'espace d'hypothèses peut être infini
- Bonnes performances en généralisation
- Qualité de l'échantillon

Apprendre une fonction de classification (une *hypothèse*)

Exemple de séparation (linéaire)



Quelques problèmes

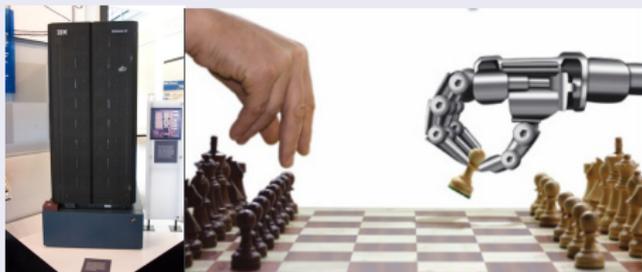
- L'espace d'hypothèses peut être infini
- Bonnes performances en généralisation
- Qualité de l'échantillon

Exemple : apprendre à jouer

Approches exhaustives et apprentissage

Othello, backgammon, échecs, GO

Les échecs



(crédits : IBM+RAG)

- 1997 : Deep Blue bat G. Kasparov
- 2006 : DeepFritz bat V. Kramnik (sur un ordinateur standard)

Exemple : apprendre à jouer

Approches exhaustives et apprentissage

Othello, backgammon, échecs, GO

Le GO : deep et reinforcement learning



(Crédits : unbalance.co.jp et N. Sheldon pour Wired)

- 2012, 2013 : Zen, Crazy Stone (avec handicap)
- 2015 : Alpha Go (sans handicap) bat Fan Hui, puis Lee Sedol en 2016
- 2017 : parties plus rapides, victoire face à Ke Jie.

Mathématiques

- Probabilités, statistiques
- Mathématiques discrètes
- Analyse et optimisation
- Algèbre linéaire

Sciences cognitives

- Sciences du langage
- Neurosciences
- Logiques

Communications of the ACM, 27, 1984. : fondements du cadre PAC (Probably Approximately Correct Learning)

RESEARCH CONTRIBUTIONS

Artificial
Intelligence and
Language Processing

David Waltz
Editor

A Theory of the Learnable

L. G. VALIANT

ABSTRACT: Humans appear to be able to learn new concepts without needing to be programmed explicitly in any conventional sense. In this paper we regard learning as the phenomenon of knowledge acquisition in the absence of explicit programming. We give a precise methodology for studying this phenomenon from a computational viewpoint. It consists of choosing an appropriate information gathering mechanism, the learning protocol, and exploring the class of concepts that can be learned using it in a reasonable (polynomial) number of steps. Although inherent algorithmic complexity appears to set serious limits to the range of concepts that can be learned, we claim that these limits



a genetically preprogrammed element, whereas others consist of executing an explicit sequence of instructions that has been memorized. There remains a large area of skill acquisition where no such explicit programming is identifiable. It is this area that we describe here as learning. The recognition of familiar objects, such as tables, provides such examples. The skills often have the additional property that, although we have learned them, we find it difficult to articulate what algorithm we are really using. In these cases it would be especially significant if machines could be made to acquire them by learning.

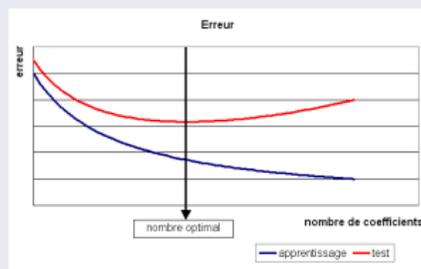
Quelques groupes de méthodes

- Les réseaux de neurones, le **deep learning**
- Les machines à vecteur support (SVM), méthodes à noyaux
- Les arbres de décision
- Les modèles de mélanges gaussiens, EM
- Les ensembles de classifieurs (MKL, boosting)
- L'apprentissage par renforcement
- Les approches stochastiques

Au delà de la classification binaire

- Régression, estimation de densité
- Multi-classes, multi-étiquettes, multi-vues
- Ordonnancement
- Apprentissage actif (Cadres semi-supervisé, non-supervisé)
- **Apprentissage de représentation**

Erreurs



- Erreur apparente (calculée : empirique, d'apprentissage)
- Erreur de **généralisation** (estimée)
- *Associées à fonction de coût/perte*

Mesures supplémentaires

Précision, rappel, F-mesure, Courbe de ROC, AUC, Mean Average Precision, etc.

Le plus important

- Ne pas apprendre par coeur : savoir généraliser
- Forte dépendance à la qualité/taille de l'échantillon (bruit, trous, etc.)
- Sélection de modèles

Apprentissage supervisé – formalisation

- \mathcal{X} espace d'entrée, \mathcal{Y} espace des cibles
- D distribution sur $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$
- $S_{\text{train}} = \{(X_i, Y_i)\}_{i=1}^m$ échantillon de n v.a. **Indépendamment et Identiquement Distribuées (IID)** suivant D
- Fonction de perte $\ell : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$

But : minimisation du (vrai) risque

A partir de S_{train} trouver $f : \mathcal{X} \rightarrow \mathcal{Y}$ telle que $f = \operatorname{argmin}_h R(h)$ avec

$$R(h) = \mathbb{E}_{X,Y} \ell(h(X), Y) = \int_{\mathcal{X} \times \mathcal{Y}} \ell(h(x), y) dD(x, y)$$

Pour la classification/catégorisation, $|\mathcal{Y}| < +\infty$, $\ell(y, y') = \mathbb{I}(y \neq y')$ et

$$R(h) = \mathbb{P}_{X,Y \sim D}(h(X) \neq Y)$$

Plan

- 1 Machine Learning
- 2 Deep Learning
 - Pouvoir d'approximation
 - Quelques architectures répandues
- 3 Deep health
- 4 Conclusion

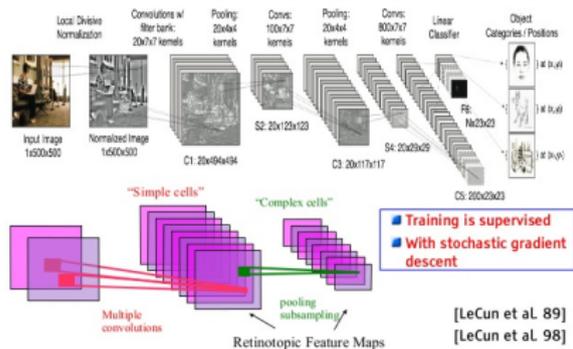
La révolution du deep learning

| | VGGNet | DeepVideo | GNMT |
|------------|--|--|---|
| Used For | Identifying Image Category | Identifying Video Category | Translation |
| Input | Image  | Video  | English Text  |
| Output | 1000 Categories | 47 Categories | French Text |
| Parameters | 140M | ~100M | 380M |
| Data Size | 1.2M Images with assigned Category | 1.1M Videos with assigned Category | 6M Sentence Pairs, 340M Words |
| Dataset | ILSVRC-2012 | Sports-1M | WMT'14 |

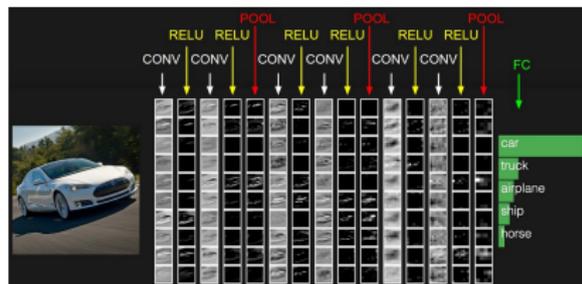
La révolution du deep learning

The Convolutional Net Model (Multistage Hubel-Wiesel system)

Y LeCun



(crédits : Y. Le Cun)



(crédits : Y. Le Cun)

Fin de la 1^{ère} partie

Cécile CAPPONI
(avec l'aide de Thierry Artières)

LIS, Aix-Marseille Université, CNRS
Equipe QARMA

