

# Notes de cours

## OPT6 : Apprentissage avancé

Adrien Pavao

September 2017

### 1 Introduction

Apprentissage avancé. Le domaine de l'apprentissage automatique évolue vite.

#### 1.1 Quelques notions

- Vladimir vapnik Bases statistiques de l'apprentissage...
- Apprentissage PAC : Probably Approximately Correct.
- **No-free-lunch theorem** : Pas de méthode d'apprentissage qui soit meilleure que les autres. Même performance que l'aléatoire. Si on est bon sur un problème on sera mauvais sur un autre (conservation de la goniétance en gros). L'intégrale sur tous les problèmes est toujours la même, quelque soit la méthode. Pareil pour les mesures (de distance par exemple). Les méthodes sont adaptées à des classes de problèmes. **Tous les algorithmes inductifs se valent.**
- adversarial learning
- ordre  $l_0$ ,  $l_1$ , distance manhattan.
- Hofstadter, "Godel, Game, Bach"
- Complexité de kolmogorov
- Effets de séquences
- L'algorithme d'apprentissage du perceptron converge si les données sont linéairement séparables. Cette convergence est en nombre fini d'étapes et ce nombre est indépendant du nombre d'exemples, de la distribution des exemples, et quasiment pas de la dimension de l'espace d'entrée. Il dépend de la marge de séparation des nuages de points et du diamètre de la boule.
- L'espace des version : Toutes les hypothèses correctes (qui ne font pas d'erreurs) d'après les données d'apprentissage.

## 2 Induction

On déduit à partir d'un échantillon.

1. Espace d'hypothèse.
2. Critere inductif  $S \times h \in H \rightarrow \text{Score E R}$ . Par exemple minimiser le taux d'erreur.
3. Methode d'exploration de H.

On ne dispose pas de théorie bien établie de l'induction (exemple fibonacci).

### 2.1 Différents types d'apprentissage

- **Descriptif** : Non supervisé. On cherche des régularités dans les données. On ne souhaite pas extrapoler, on ne s'intéresse qu'à l'échantillon de données dont on dispose. La "matière noire" de l'apprentissage.
- **Prédictif** : Supervisé. L'échantillon de données sert à apprendre une hypothèse sur les données pour prédire ensuite sur de nouvelles données. On cherche des corrélations entre les données.
- **Prescriptif** : On cherche des **causalités**. Il s'agit d'une tâche difficile.

Les frontières entre ces types d'apprentissage peuvent être floues, on peut par exemple commencer par une description des données pour les comprendre avant de faire un modèle prédictif.

### 2.2 Différentes méthodes d'apprentissage

- Méthodes d'ensemble
- Supervisé (?)
- L'induction :
  - Apprentissage semi supervisé
  - Apprentissage de modèles parcimonieux (peu de paramètres).
- Apprentissage en ligne
- **Apprentissage par transfert** : Cela vise à transférer des connaissances d'une ou plusieurs tâches sources vers une ou plusieurs tâches cibles. On peut le voir comme la capacité d'un système à reconnaître et appliquer des connaissances et des compétences, apprises à partir de tâches antérieures, sur de nouvelles tâches ou domaines partageant des similitudes.

## 2.3 Types de biais

L'induction nécessite d'être biaisé, d'avoir une préférence pour certaines hypothèses, de ne pas connaître toutes les fonctions possibles

- **Biais de représentation (déclaratif) :** On ne peut représenter qu'un petit nombre de fonctions possibles. Le langage dans lequel on exprime le problème ne permet pas de tout représenter.
- **Biais de recherche (procédural) :** La procédure de recherche avantage certaines hypothèses. Par exemple on reste proche de notre point initial de recherche dans l'espace des fonctions possibles.

Souvent un peu des deux.

## 2.4 Risques

On distingue deux types de risques :

- **Risque réel** (ce que l'on veut minimiser). La performance à venir si j'utilise l'hypothèse  $h$ . Cela prend la forme d'une espérance de coût d'usage de  $h$ . On ne le risque réel  $R(h)$  = sale formule voir cours (loss function). Meilleure hypothèse  $k^* = \operatorname{Argmin} R(h)$  avec  $h \in H$ .
- **Risque empirique.**  $R(h) = \frac{1}{m} \sum_{i=1}^m R(h, x_i, y_i)$  sale formule voir cours.  $h$  chapeau minimise risque empirique.

On cherche les liens entre toutes ces notions.

## 3 Transduction

On ne s'intéresse qu'à un point. Le nombre de données requises pour tirer une conclusion n'est plus influencé par le nombre de dimension. On parle de **transductif learning**.

- Transduction : cas d'induction limite (une question).
- Analogie : cas de transduction limite (un exemple).