Apprentissage (Machine learning)

Notes par Sarah Vita Paardekooper

24 Septembre 2020

1 Introduction

L'apprentissage consiste à estimer des modèles à partir de données limitées, comme une liste d'exemples ou par des expériences. L'apprentissage contient plusieurs aspects :

- Complexité
- Logique
- Algorithmique
- Statistique

Les méthodes d'apprentissage peuvent être regroupé dans les quatre catégories différentes suivantes :

1.1 Apprentissage supervisé (supervised learning)

L'ensemble $X = \{x_1, \cdots, x_n\}$ est appelé le *domaine*. L'ensemble $Y = \{y_1, \cdots, y_n\}$, s'il existe, est appelé l'*ensemble d'étiquettage* (label set). Souvent on aura $Y = \{0, 1\}$, et on parlera de classification.

Dans un apprentissage supervisé on a en données un domaine X et un ensemble d'étiquettes Y. On cherche une fonction $h : X \to Y$ qui généralise les examples (x, y) données.

INPUT : $(x_1, y_1), \cdots, (x_n, y_n)$

OUTPUT : $h : X \to Y$

1.2 Apprentissage non supervisé (unsupervised learning)

Dans un apprentissage non supervisé on a seulement le domaine X en entrée sans étiquettes, et un entier k. Le but est de faire une partition de X en k classes.

INPUT : $x_1, \cdots, x_n$, $k \in \mathbb{N}$

OUTPUT : partition de X en k classes

1.3 Apprentissage par récompense (Reinforcement Learning ou RL)

Un ou plusieurs agents agissent sur un environnement (inconnu de base) et l'environnement répond aux actions des agents avec une récompense $r \in \mathbb{R}$. Le but est de choisir par essai-erreur les actions qui maximisent les récompenses. Contrairement à l'apprentissage supervisé et non supervisé on n'a pas de domaine et/ou ensembles d'étiquettage en INPUT/OUTPUT.

Exemple : AlphaGo

1.4 Online Learning (apprentissage incrémentiel)

L'apprentissage incrémentiel peut être supervisé ou non supervisé. Dans le cas supervisé, un couple supplémentaire (x_i, y_i) est donné à chaque incrémentation i et on en déduit une nouvelle hypothèse h_i.

On se concentrera par la suite sur l'apprentissage supervisé.

2 Consistency model

Soient X un domaine, Y un ensemble d'étiquettes et $\mathcal{H} \subset \{f : X \to Y\}$ un ensemble d'hypothèses.

Etant donné un échantillon (sample) $S = \{(x_i, y_i) : i \in [1, |S|]\}$, on cherche une hypothèse $h \in \mathcal{H}$ tel que $h \models S$.

INPUT : un échantillon (sample) $S = \{(x_i, y_i) : i \in [1, |S|]\}$

OUTPUT : une hypothèse $h \in \mathcal{H}$ tel que $h \models S$ (c'est à dire $h(x_i) = y_i$ pour tout $i \in [1, |S|]$)

Définition 2.1. Un algorithme **apprend** $\mathcal{H}$ si pour tout échantillon S, l'algorithme retourne $h \models S$ avec $h \in \mathcal{H}$, ou "il n'existe pas de $h \in \mathcal{H}$ tel que $h \models S$".

De plus il est **efficace** s'il est en temps $poly(dim(x_i), |S|)$.

Voici quelques exemples de consistency models (modèles de cohérence) :

2.1 Formules conjonctives

$X = \{0,1\}^n$, $n \in \mathbb{N}$,
$Y = \{0,1\}$
$\mathcal{H}_{FC} = \{\phi = \bigwedge_{p \in X_1} p \wedge \bigwedge_{p \in X_2} \neg p : X_1, X_2 \subseteq [1, n]\}$

Par exemple pour une dimension $n = 4$ on a $\mathcal{H} = \{p_1, p_2, p_1 \wedge \neg p_2, p_1 \wedge p_2 \wedge p_3 \wedge p_4, p_4 \wedge \neg p_3, \cdots\}$. Soit un échantillon $S = \{(x_1, y_1), \cdots, (x_5, y_5)\}$ définit comme suit :

	p_1	p_2	p_3	p_4	p_5		
$x_1 =$	0	1	0	1	1	$y_1 =$	1
$x_2 =$	0	1	1	1	0	$y_2 =$	1
$x_3 =$	0	1	1	0	1	$y_3 =$	0
$x_4 =$	1	0	1	1	0	$y_4 =$	0
$x_5 =$	0	0	0	1	1	$y_5 =$	0

Soit $\phi = p_2 \wedge p_4$, on a $x_1 \models \phi$, $x_2 \models \phi$, $x_3 \not\models \phi$, $x_4 \not\models \phi$, et $x_5 \not\models \phi$, donc $S \models \phi$.

Voici une méthode pour trouver $\phi \in \mathcal{H}$ tel que $S \models \phi$:

1. On part de la plus longue formule $\phi = \bigwedge_{p \in X} p \wedge \bigwedge_{p \in X} \neg p$,
2. puis, pour chaque x_i tels que $y_i = 1$, on enlève les p_j tels que $x_{i,j} = 0$ et les $\neg p_j$ tels que $x_{i,j} = 1$,
3. si $S \models \phi$, retourner ϕ,
4. sinon, retourner "il n'y a pas de ϕ".

Cet algorithme est de complexité $O(n|S|)$.

Dans le cas de l'exemple, on obtient :

$\phi = p_1 \wedge \neg p_1 \wedge p_2 \wedge \neg p_2 \wedge p_3 \wedge \neg p_3 \wedge p_4 \wedge \neg p_4 \wedge p_5 \wedge \neg p_5$

$\phi = \quad \wedge \neg p_1 \wedge p_2 \wedge \quad \wedge \quad \wedge \neg p_3 \wedge p_4 \wedge \quad \wedge p_5 \wedge$

$\phi = \neg p_1 \wedge p_2 \wedge p_4$

$\mathcal{H}$ est donc apprenable efficacement.

2.2 Dysjunctive Normal Form (DNF)

$X = \{0,1\}^n$, $n \in \mathbb{N}$,

$Y = \{0,1\}$

$\mathcal{H}_{DNF} = \{\phi = \bigvee_{i=1}^{k} C_i : C_i = \bigwedge_{p \in X_i} \pm p, X_i \subseteq X\}$

Par exemple pour une dimension $n = 3$ on a $(p_1 \wedge \neg p_2) \vee (p_3 \wedge p_2) \in \mathcal{H}$. Remarquons que $\mathcal{H}_{FC} \subset \mathcal{H}_{DNF}$. Par ailleurs, pour $S = \{(x_i, y_i) : i \in [1, |S|]\}$ on obtient facilement une formule $\phi = \bigvee_{\{i: y_i = 1\}} (\bigwedge_{\{j: x_{i,j}=1\}} p \wedge \bigwedge_{\{j: x_{i,j}=0\}} \neg p)$ tel que $S \models \phi$.

Avec l'échantillon S décrit précédemment on obtient :

$$\phi = (\neg p_1 \wedge p_2 \wedge \neg p_3 \wedge p_4 \wedge p_5) \\ \vee \\ (\neg p_1 \wedge \neg p_2 \wedge p_3 \wedge p_4 \wedge \neg p_5)$$

$\mathcal{H}$ est trivialement apprenable.

2.3 3-Dysjunctive Normal Form (3-DNF)

$X = \{0,1\}^n$, $n \in \mathbb{N}$,

$Y = \{0,1\}$

$\mathcal{H}_3 = \{\phi = C_1 \vee C_2 \vee C_3 : C_i = \bigwedge_{p \in X_i} \pm p, X_i \subseteq X\}$

On va montrer que $\mathcal{H}_3$ n'est pas apprenable efficacement en montrant l'hypothèse cohérente suivante :

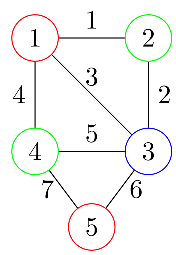
Étant donné un graphe G, on peut construire un échantillon $S = \{(x_i, y_i)\}$ tel que G est 3-coloriable si et seulement si il existe $\phi \in \mathcal{H}_3$, tel que $S \models \phi$.

Le problème de 3-colorabilité est défini comme suit :

INPUT : un graphe G

OUTPUT : Est-ce qu'il existe une coloration $c : V \to \{R, V, B\}$, tel que $\forall e = (u, v) \in E$, $c(u) \neq c(v)$?

Par exemple, le graphe à $n = 5$ et $m = 7$ arêtes sommets suivant est 3-coloriable :



Et on peut construire un échantillon
$S = \{(v_1, y_1), \cdots, (v_n, y_n), (e_1, y_{n+1}), \cdots, (e_m, y_{n+m})\}$ de dimension n définit par :
- $y_i = 1$ pour $i \leq n$, et $y_i = 0$ sinon,
- $v_{i,i} = 1$ et $v_{i,j} = 0$ pour tout $i \neq j$,
- $e_{i,j} = 1$ si et seulement si l'arête i est incidente au sommet j.

	p_1	p_2	p_3	p_4	p_5	y
$v_1 =$	1	0	0	0	0	1
$v_2 =$	0	1	0	0	0	1
$v_3 =$	0	0	1	0	0	1
$v_4 =$	0	0	0	1	0	1
$v_5 =$	0	0	0	0	1	1
$e_1 =$	1	1	0	0	0	0
$e_2 =$	0	1	1	0	0	0
$e_3 =$	1	0	1	0	0	0
$e_4 =$	1	0	0	1	0	0
$e_5 =$	0	0	1	1	0	0
$e_6 =$	0	0	1	0	1	0
$e_7 =$	0	0	0	1	1	0

Preuve. 1. $\Rightarrow$: Supposons que G est 3-coloriable et soit $c : V \to \{R, V, B\}$ une coloration de G. Construisons $\phi = C_R \vee C_V \vee C_B$ tel que $S \models \phi$.
C_R doit être satisfaite par v_1 et v_5 : $C_R = \neg p_2 \wedge \neg p_3 \wedge \neg p_4$. Comme les sommets 1 et 5 dans G sont de même couleur, ils n'ont pas d'arête commune, donc pour tout i on a $e_i \not\models \phi$. De même, on a $C_V = \neg p_1 \wedge \neg p_3 \wedge \neg p_5$ et $C_B = \neg p_1 \wedge \neg p_2 \wedge \neg p_4 \wedge \neg p_5$.

2. $\Leftarrow$: Supposons qu'il existe une formule $\phi = C_V \vee C_R \vee C_B$ tel que $S \models \phi$. Alors G est bien 3-coloriable. $\square$

Notons que dans le modèle de cohérence on assure que tout les x_i tel que $y_i = 1$ sont vrais et tout les x_i tel que $y_i = 0$ sont faux, hors on a plus de chance de généraliser avec une marge d'erreur, c'est à dire si on autorise quelques couples de l'échantillons à ne pas satisfaire l'hypothèse.

3 PAC model