Jeux, Synthèses et Contrôles

Cours 3 : Agnostic setting

CARN, Zahra

Thursday 8 October 2020

Enseignant :
Nathanaël FIJALKOW

Table des matières

1 Inégalités utiles à savoir

$$e^x \geq 1 + x, \quad x \in \mathbb{R} \tag{1}$$

$$(1+x)^a \geq 1 + xa, \quad a, x \geq 0 \tag{2}$$

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B) \tag{Union Bound}$$

2 Agnostic Setting

En français : modèle agnostique

modèle	$X, Y \quad \mathcal{H}$ ensemble de fonctions $X \to Y$
consistant	On suppose $\exists c \in \mathcal{H}$ et le sample $S = \{(x_i, c(x_i))\}_i$ Objectif : Trouver $h \in \mathcal{H}$ cohérent avec S
PAC	On suppose $\mathcal{D}$ sur $X, \exists c \in \mathcal{H}$ et $S \sim \mathcal{D}^m, S = \{(x_i, c(x_i))\}_i$ Objectif : $\mathbb{P}_{S\sim\mathcal{D}^m}(\text{err}_c(h) \leq \epsilon) \geq 1 - \delta$ où $\text{err}_c(h) = \mathbb{P}_{x\sim\mathcal{D}}(h(x) \neq c(x))$
agnostic PAC	On suppose $\mathcal{D}$ sur $X \times Y, S \sim \mathcal{D}^m, S = \{(x_i, y_i)\}_i$ $\text{err}(h) = \mathbb{P}_{(x,y)\sim\mathcal{D}}(h(x) \neq y)$ Objectif : $\mathbb{P}_{S\sim\mathcal{D}^m}(\text{err}(h) - \inf_{h'\in\mathcal{H}}\text{err}(h') \leq \epsilon) \geq 1 - \delta$

On note que $\inf_{h'\in\mathcal{H}}\text{err}(h')$ le modèle agnostic PAC prend en quelque sorte le rôle de $c \in \mathcal{H}$ dans le modèle PAC.

Définition 2.1 (Erreur empirique)**.**

$$h \in \mathcal{H}, \widehat{\text{err}}(h) = \widehat{\text{err}}(h) = \frac{1}{m}\Sigma_{i=1}^{m}\mathbb{1}_{h(x_i)\neq y_i}$$

Note. $\mathbb{1}$ est la fonction indice défini par $\mathbb{1}_A = \begin{cases} 1 & \text{si } A \\ 0 & \text{sinon} \end{cases}$

Note. $\widehat{\ }$ est utilisé en probabilités et statistiques pour représenter une valeur **empirique**. C'est à dire une valeur qui est calculée à partir de données.

Note. $\arg\min_{x\in X} f(x)$ retourne une valeur $x_{min} \in X$ tel que $f(x_{min}) = \min_{x\in X} f(x)$.

Définition 2.2 (ERM)**.** h est une hypothèse ERM (*empirical risk minimisation*) si $h = \arg\min_{h'\in\mathcal{H}}\widehat{\text{err}}(h')$.

Un algorithme ERM retourne h ERM.
$\mathcal{H}$ est apprenable au sens agnostic PAC s'il existe un algo :

$$\forall\epsilon, \forall S, \forall\mathcal{D} \text{ sur } X \times Y, \exists M \in \mathbb{N}, \forall m \geq M,$$
$$\mathcal{P}_{S\sim\mathcal{D}^m}(\text{err}(h) - \inf_{h'\in\mathcal{H}}\text{err}(h') \leq \epsilon) \geq 1 - \delta \tag{3}$$

Théorème 2.1. *$\mathcal{H}$ est apprenable au sens agnostic PAC si et seulement si $\mathcal{H}$ a VC dimension ($VCdim$) fini.*

Cette théorème sera prouvé au cours 4.

Définition 2.3 (UCP)**.** Propriété de convergence uniforme (*uniform convergence property*) avec un ϵ fixé
$\forall h \in \mathcal{H}, |\text{err}(h) - \widehat{\text{err}}(h)| \leq \frac{\epsilon}{2}$ avec $\text{err}(h) = \mathbb{P}_{(x,y)\in\mathcal{D}}(h(x) \neq y)$

Cette définition permet de simplifier l'objectif de agnostic PAC.

Lemme 2.2. *On note h_{ERM} une hypothèse ERM tel que $\widehat{err}(h_{ERM}) = min_{h\in\mathcal{H}}\widehat{err}(h)$*
Si UCP est vrai (pour ϵ fixé), alors l'algorithme ERM satisfait

$$err(h_{ERM}) - inf_{h\in\mathcal{H}} err(h) \leq \epsilon$$

Démonstration. Notons $\text{err}(h_{OPT}) = \inf_{h\in\mathcal{H}}\text{err}(h) \Leftrightarrow h_{OPT} = \arg\inf_{h\in\mathcal{H}}\text{err}(h)$
Objectif : Montrer que $\text{err}(h_{ERM}) - \text{err}(h_{OPT}) \leq \epsilon$.
On a

$$\begin{aligned}
\text{err}(h_{ERM}) - \text{err}(h_{OPT}) &= \underbrace{\text{err}(h_{ERM}) - \widehat{\text{err}}(h_{ERM})}_{\leq \frac{\epsilon}{2} \text{ UCP}} + \underbrace{\widehat{\text{err}}(h_{ERM})}_{\leq \widehat{\text{err}}(h_{OPT}) \text{ par définition}} - \text{err}(h_{OPT}) \\
&\leq \frac{\epsilon}{2} + \underbrace{\widehat{\text{err}}(h_{OPT}) - \text{err}(h_{OPT})}_{\leq \frac{\epsilon}{2} \text{ UCP}} \\
&\leq \epsilon
\end{aligned}$$

□

$$X, Y = 0, 1, S \subseteq X$$

$\mathcal{H}$ ensemble de fonctions $X \to 0, 1$ correspond à un ensemble de sous-ensembles de X.

Définition 2.4. S est *shattered* ("explosé" en français) par $\mathcal{H}$ si

$$\forall P \subseteq S, \exists h \in \mathcal{H}, \forall x \in S, h(x) = 1 \Leftrightarrow x \in P$$

Exemple. $X = \mathbb{R}^n, \mathcal{H}_n = \{f|\mathbb{R}^n \to \{0,1\} : f(\vec{x}) = \begin{cases} 1 & \text{si } \vec{x}.\vec{a} \geq 0 \\ 0 & \text{sinon} \end{cases}\}$
avec $\vec{a} \in \mathbb{R}^n, b \in \mathbb{R}$
$\mathcal{H}_2$ explose $S := \{(x, y), (x', y')\}$, avec $(x, y), (x', y') \in \mathbb{R}^2$. Une visualisation d'un exemple est donnée dans la figure 1.
$\mathcal{H}_2$ explose $S :=$ 3 points non-alignées du plan.
$\mathcal{H}_2$ n'explose pas $S :=$ 4 points quelconques du plan.

Définition 2.5. $\text{VCdim}(\mathcal{H}) = \max\{|S| : S \text{ explosé par } \mathcal{H}\}$

Exemple ($\mathcal{H}_2$)*.* $\text{VCdim}(\mathcal{H}_2) = 3$ car $\exists S, |S| = 3$ explosé et $\forall S, |S| \geq 4$ non explosé

Exemple ($\text{VCdim}(\mathcal{H}) = \infty$)*.* $X = \mathbb{N}, \mathcal{H}$ toutes les fonctions

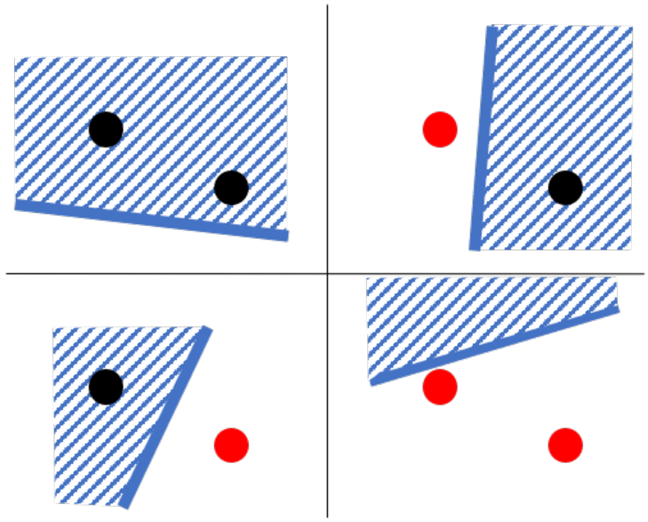

Figure 1 – Visualisation d'une explosion par $\mathcal{H}_2$

Exemple ($\mathcal{H}_n$). VCdim($\mathcal{H}_n$) $= n + 1$

Une présentation au sujet aura lieu dans quelques semaines.

Définition 2.6 (Fonction de croissance)**.**

$$S \subseteq X, h \in \mathcal{H}, h : X \to \{0,1\}, h|_s : S \to \{0,1\}$$

$$\Pi_{\mathcal{H}} : \mathbb{N} \to \mathbb{N} \qquad m \to \max_{|S|=m, S \subseteq X} |\{h|_S : h \in \mathcal{H}\}| \tag{4}$$

Exemple. Si VCdim($\mathcal{H}$) $= d$ que vaut $\Pi_{\mathcal{H}}(d)$? 2^d

Que vaut $\Pi_{\mathcal{H}}(d-1)$? 2^m, pour $m \leq d$

Que vaut $\Pi_{\mathcal{H}}(d+1)$? $< 2^{d+1}$

Remarque. S explosé par $\mathcal{H} \Leftrightarrow |\{h|S : h \in \mathcal{H}\}| = 2^d$

Lemme 2.3 (Sauer)**.** *Si VCdim($\mathcal{H}$) $= d$ alors*

$$\Pi_{\mathcal{H}}(m) \leq \binom{m}{0} + \binom{m}{1} + ... + \binom{m}{d} = \sum_{i=0}^{d} \binom{m}{i} \leq (\frac{em}{d})^d = O(m^d)$$

Les instances de la forme X infini, $\mathcal{H}$ toutes les fonctions $X \to \{0,1\}$ sont impossibles à apprendre.

Détails à venir