Apprentissage: cours 6 (2ème partie): preuves

Sylvain Arlot

22 mars 2012

Démonstration du Théorème 1. Pour tout $f \in \mathcal{F}$, l'inégalité de Hoeffding (Théorème 1 de la boîte à outils de probabilités) appliquée aux variables $\xi_i = -n^{-1}\ell(f;(X_i,Y_i)) \in [-n^{-1};0]$, qui sont indépendantes et d'espérance $-n^{-1}\mathcal{R}(f)$, montre que

$$\mathbb{P}\left(\mathcal{R}(f) - \widehat{\mathcal{R}}_n(f) \geqslant \varepsilon\right) \leqslant \exp\left(-2n\varepsilon^2\right) \quad .$$

Comme S est fini, on en déduit (par une borne d'union) que

$$\begin{aligned}
\mathbb{P}\left(\sup_{f\in S}\left\{\mathcal{R}(f) - \widehat{\mathcal{R}}_n(f)\right\} \geqslant \varepsilon\right) &= \mathbb{P}\left(\exists f \in S\,,\, \mathcal{R}(f) - \widehat{\mathcal{R}}_n(f) \geqslant \varepsilon\right) \\
&\leqslant \min\left\{1, \mathrm{card}(S)\exp\left(-2n\varepsilon^2\right)\right\} \\
&= \exp\left(-\left(2n\varepsilon^2 - \ln(\mathrm{card}(S))\right)_+\right) \quad .
\end{aligned}$$

Pour conclure, remarquons que

$$\begin{aligned}
&\mathbb{E}\left[\sup_{f\in S}\left\{\mathcal{R}(f) - \widehat{\mathcal{R}}_n(f)\right\}\right] \qquad \text{(Section 5.2 de la boîte à outils de probabilités)} \\
&\leqslant \int_0^\infty \mathbb{P}\left(\sup_{f\in S}\left\{\mathcal{R}(f) - \widehat{\mathcal{R}}_n(f)\right\} \geqslant \varepsilon\right) \mathrm{d}\varepsilon \\
&\leqslant \int_0^{\sqrt{\ln(\mathrm{card}(S))/(2n)}} \mathrm{d}\varepsilon + \int_{\sqrt{\ln(\mathrm{card}(S))/(2n)}}^\infty \exp\left(-\left(2n\varepsilon^2 - \ln(\mathrm{card}(S))\right)_+\right) \mathrm{d}\varepsilon \\
&\leqslant \sqrt{\frac{\ln(\mathrm{card}(S)}{2n}} + \int_{\sqrt{\ln(\mathrm{card}(S))/(2n)}}^\infty \exp\left(-2n\left(\varepsilon - \frac{\sqrt{\ln(\mathrm{card}(S))}}{\sqrt{2n}}\right)^2\right) \mathrm{d}\varepsilon \\
&= \sqrt{\frac{\ln(\mathrm{card}(S)}{2n}} + \frac{1}{2\sqrt{n}}\int_0^\infty \exp(-u^2/2)\mathrm{d}u \qquad \text{en posant } u = 2\sqrt{n}\varepsilon - \sqrt{2\ln(\mathrm{card}(S)} \\
&= \sqrt{\frac{\ln(\mathrm{card}(S)}{2n}} + \frac{1}{2\sqrt{n}} \times \frac{\sqrt{\pi}}{\sqrt{2}} \ ,
\end{aligned}$$

où l'on a utilisé que si $a \geqslant b \geqslant 0$, alors $a^2 - b^2 \geqslant (a-b)^2$. □

Démonstration de la remarque suivant le Théorème 1. D'après la Proposition 3 de la boîte à outils de probabilités, pour tout f, $\mathcal{R}(f) - \widehat{\mathcal{R}}_n(f)$ est une variable sous-gaussienne de paramètre $1/(4n)$. On peut donc appliquer le Théorème 4 de la boîte à outils de probabilités avec $v = 1/(4n)$ et l'on obtient que

$$\mathbb{E}\left[\sup_{f \in S}\left\{\mathcal{R}(f) - \widehat{\mathcal{R}}_n(f)\right\}\right] \leqslant \frac{\sqrt{\ln(\operatorname{card}(S))}}{\sqrt{2n}} \quad .$$

□

Démonstration du Corollaire 2. L'algorithme de majorité $\widehat{f}^{\mathrm{maj}}$ est l'un des algorithmes minimisant le risque empirique sur $\mathcal{F}$. On a juste précisé l'étiquette à attribuer à un point x où l'on a observé autant de fois $Y = 1$ et $Y = 0$. De plus, $\mathcal{F}$ est fini (de cardinal $2^{\operatorname{card}(\mathcal{X})}$) car $\mathcal{X}$ est fini. Enfin, l'erreur d'approximation de $\mathcal{F}$ est nulle (car $f^\star \in \mathcal{F}$). Ainsi,

$$\begin{aligned}
\mathbb{E}\left[\mathcal{R}\left(\widehat{f}^{\mathrm{maj}}(D_n)\right) - \mathcal{R}(f^\star)\right] &= \mathbb{E}\left[\mathcal{R}\left(\widehat{f}\right) - \inf_{f \in \mathcal{F}} \mathcal{R}(f)\right] \\
&\leqslant \mathbb{E}\left[\sup_{f \in \mathcal{F}}\left\{\mathcal{R}(f) - \widehat{\mathcal{R}}_n(f)\right\}\right] && \text{(Exercice 1)} \\
&\leqslant \frac{1}{\sqrt{2n}}\left[\sqrt{\ln(\operatorname{card}(\mathcal{F}))} + \frac{\sqrt{\pi}}{2}\right] && \text{(Théorème 1 avec } S = \mathcal{F}\text{)} \\
&= \frac{1}{\sqrt{2n}}\left[\sqrt{\ln(2)\operatorname{card}(\mathcal{X})} + \frac{\sqrt{\pi}}{2}\right] \quad .
\end{aligned}$$

□