On utilise le lemme suivant :

Si X va réelle positive :

$$\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}(X \geq t)\,dt \quad (\text{Fubini-Tonelli})$$

On a donc :

$$\mathbb{E}\left[\max_{g \in \mathcal{G}} |\hat{R}(g) - R(g)|\right]$$

$$= \int_0^{+\infty} \mathbb{P}\left(\max_{g \in \mathcal{G}} |\hat{R}(g) - R(g)| \geq t\right) dt$$

On va contrôler : $\mathbb{P}\left(\max_{g \in \mathcal{G}} |\hat{R}(g) - R(g)| \geq t\right)$

$\forall t \geq 0$.

C'est à dire, que l'on cherche à contrôler les déviations du processus empirique

$\max_{g \in \mathcal{G}} |\hat{R}(g) - R(g)|$

On a : $\left\{\max_{g \in \mathcal{G}} |\hat{R}(g) - R(g)| \geq t\right\}$

$$= \left\{\exists g \in \mathcal{G},\ |\hat{R}(g) - R(g)| \geq t\right\}$$

$$= \bigcup_{g \in \mathcal{G}} \left\{|\hat{R}(g) - R(g)| \geq t\right\}$$

$$\Rightarrow \mathbb{P}\left(\max_{g \in \mathcal{G}} |\hat{R}(g) - R(g)| \geq t \right)$$

$$= \mathbb{P}\left(\bigcup_{g \in \mathcal{G}} |\hat{R}(g) - R(g)| \geq t \right)$$

$$\leq \sum_{g \in \mathcal{G}} \mathbb{P}\left(|\hat{R}(g) - R(g)| \geq t \right)$$

$$\Rightarrow \mathbb{E}\left[\mathcal{E}\left(\hat{g}(D_n) \right) \right] \leq 2 \sum_{g \in \mathcal{G}} \int_0^{+\infty} \mathbb{P}\left(|\hat{R}(g) - R(g)| \geq t \right) dt$$

Etape 3 Application de l'inégalité de Hoeffding

pour $t \geq 0$, on va borner à l'aide de l'inégalité de Hoeffding :

$$\mathbb{P}\left(|\hat{R}(g) - R(g)| \geq t \right) \quad \text{pour } g \in \mathcal{G}.$$

l'inégalité de Hoeffding peut être vue comme un raffinement de l'inégalité de Markov (ou Bienaymé-Tchebychev)

D'après Bienaymé-Tchebychev, on

$$\mathbb{P}\left(|\hat{R}(g) - R(g)| \geq t \right) \leq \frac{\mathrm{Var}(\hat{R}(g))}{t^2}$$

$$\mathbb{E}[\hat{R}(g)] = R(g)$$

On : $\mathrm{Var}(\hat{R}(g))$

(x_i, y_i) iid

$$= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{g(X_i) \neq Y_i\}}\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left(\mathbb{1}_{\{g(X_i) \neq Y_i\}}\right)$$

$$= \frac{\mathrm{Var}\left(\mathbb{1}_{\{g(X) \neq Y\}}\right)}{n}$$

On : $\mathbb{1}_{\{g(X) \neq Y\}} \sim \mathcal{B}e\left(R(g)\right)$

Car : $\mathbb{P}\left(\mathbb{1}_{\{g(X) \neq Y\}} = 1\right) = \mathbb{P}\left(g(X) \neq Y\right) = R(g)$

$$\Rightarrow \mathrm{Var}(\hat{R}(g)) = \frac{R(g)(1-R(g))}{n}$$

$$\Rightarrow \forall t > 0 \quad \mathbb{P}\left(|\hat{R}(g) - R(g)| \geq t\right) \leq \frac{R(g)(1-R(g))}{n t^2}$$

$$\leq \frac{1}{4 n t^2}$$

Inégalité de Hoeffding : Soit $(Z_1, \ldots, Z_n)$ n variables indépendantes et centrées tq $\forall i \; a_i \leq Z_i \leq b_i$ alors $\forall t > 0$

$$\mathbb{P}\left(\sum_{i=1}^{n} Z_i \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

et $\mathbb{P}\left(\left|\sum_{i=1}^{n} Z_i\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) \leftarrow$

8

On va appliquer Hoeffding à

$|\hat{R}(g) - R(g)|$

$$= \Big| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{g(x_i) \neq y_i\}} - \overbrace{\mathbb{E}\left[\mathbb{1}_{\{g(x) \neq y\}}\right]}^{\text{„} R(g)}\Big|$$

$$= \Big| \frac{1}{n} \sum_{i=1}^{n} \Big(\underbrace{\mathbb{1}_{\{g(x_i) \neq y_i\}} - \mathbb{E}\left[\mathbb{1}_{\{g(x) \neq y\}}\right]}_{Z_i} \Big)\Big|$$

$$= \frac{1}{n} \Big| \sum_{i=1}^{n} Z_i \Big| \qquad \left[R(g) = \frac{1}{n} \sum_{i=1}^{n} R(g) \right]$$

On Z_i centrée $\forall i$ et indépendantes.

De plus : $-R(g) \leq Z_i \leq 1 - R(g)$

$$\left(0 \leq \mathbb{1}_{\{g(x_i) \neq y_i\}} \leq 1 \right)$$

$\Rightarrow \forall i \quad a_i = -R(g)$ et $b_i = 1 - R(g)$

et $b_i - a_i = 1$

$$\Rightarrow \sum_{i=1}^{n} (b_i - a_i)^2 = n.$$

par Hoeffding :

$$\mathbb{P}\left(\frac{1}{n} \Big| \sum_{i=1}^{n} Z_i \Big| \geq t \right) = \mathbb{P}\left(\Big| \sum_{i=1}^{n} Z_i \Big| \geq nt \right)$$

$$\leq 2 \exp\left(-\frac{2n^{\cancel{2}} t^2}{\cancel{n}} \right)$$

Finalement : On a $\forall t > 0$

$$\mathbb{P}\left(|\hat{R}(g) - R(g)| \geq t\right) \leq 2\exp(-2nt^2)$$

$$\Rightarrow \mathbb{P}\left(\max_{g \in \mathcal{G}} |\hat{R}(g) - R(g)| \geq t\right)$$

$$\leq \sum_{g \in \mathcal{G}} 2\exp(-2nt^2) = 2\,\mathrm{Card}(\mathcal{G})\exp(-2nt^2)$$

$$\Rightarrow \mathbb{P}\left(\max_{g \in \mathcal{G}} |\hat{R}(g) - R(g)| \geq t\right)$$

$$\leq \min\left(1, 2\,\mathrm{Card}(\mathcal{G})\exp(-2nt^2)\right)$$

$$\leq \min\left(1, \exp\left(-\left(2nt^2 - \log(2\,\mathrm{Card}(\mathcal{G}))\right)\right)\right)$$

$$= \exp\left(-\left(2nt^2 - \log(2\,\mathrm{Card}(\mathcal{G}))\right)_+\right)$$

$$= \exp\left(-\left(2nt^2 - (K+1)\log(2)\right)_+\right)$$

$$\Rightarrow \mathbb{E}\left[\mathcal{E}(\hat{g}(D_n))\right] \leq 2\int_0^{+\infty} \exp\left(-\left(2nt^2 - (K+1)\log(2)\right)_+\right) dt$$

Etape 4 : Conclusion.

Majoration de $\int_0^{+\infty} \exp\left(-\left(2nt^2 - (K+1)\log(2)\right)_+\right) dt$

On note : $f(t) = \exp\left(-\left(2mt^2 - (K+1)\log(2)\right)_+\right)$ 10

On a $f(t) = 1$

Si $2mt^2 - (K+1)\log(2) \leq 0$

Rappel : $(x)_+ = x$ si $x > 0$
$= 0$ sinon.

$\Rightarrow \quad t \leq \sqrt{\frac{(K+1)\log(2)}{2m}}$

$$\int_0^{+\infty} f(t)\,dt = \int_0^{\sqrt{\frac{(K+1)\log(2)}{2m}}} dt + \int_{\sqrt{\frac{(K+1)\log(2)}{2m}}}^{+\infty} \exp\left(-\left(2mt^2 - (K+1)\log(2)\right)\right) dt$$

$$= \sqrt{\frac{(K+1)\log(2)}{2m}} + \int_{\sqrt{\frac{(K+1)\log(2)}{2m}}}^{+\infty} \exp\left(-\left(2mt^2 - (K+1)\log(2)\right)\right) dt$$

On a : $\left(2mt^2 - (K+1)\log(2)\right)$

$$= 2m\left(t^2 - \frac{(K+1)\log(2)}{2m}\right)$$

Or : $a^2 - b^2 \geq (a-b)^2$ pour $a \geq b \geq 0$

$\Rightarrow \left(2mt^2 - (K+1)\log(2)\right)$

$$\geq 2m\left(t - \sqrt{\frac{(K+1)\log(2)}{2m}}\right)^2$$

(11)

$$\int_{\sqrt{\frac{(K+1)\log(2)}{2n}}}^{+\infty} \exp\left(-\left(2nt^2 - (K+1)\log(2)\right)\right) dt$$

$$\leq \int_{\sqrt{\frac{(K+1)\log(2)}{2n}}}^{+\infty} \exp\left(-2n\left(t - \sqrt{\frac{(K+1)\log(2)}{2n}}\right)^2\right) dt$$

$$\underset{\uparrow}{=} \int_0^{+\infty} \exp(-2nu^2)\, du$$

chgt de variable

$$\underset{\uparrow}{=} \frac{1}{\sqrt{2n}} \int_0^{+\infty} \exp(-x^2)\, dx = \frac{\sqrt{\pi}}{2\sqrt{2n}}$$

change de variable

$$\Rightarrow \int_0^{+\infty} f(t)\, dt \leq \sqrt{\frac{(K+1)\log(2)}{2n}} + \frac{\sqrt{\pi}}{2\sqrt{2n}}$$

$$\Rightarrow \mathbb{E}\left[R(\hat{g}(D_n)\right] - R(g^*)$$

$$\leq 2\left(\sqrt{\frac{(K+1)\log(2)}{2n}} + \frac{\sqrt{\pi}}{2\sqrt{2n}}\right)$$

$$\leq C\sqrt{\frac{K}{n}}$$ où C est une constante Dès que $K \geq 4$

$$\leq C\sqrt{\frac{\mathrm{Card}(\mathcal{Z})}{n}}$$

5.3 Conclusion et remarques

On a mq dans le cas où $\mathcal{X}$ est de cardinal fini égal à $K \geq 4$ en classification binaire le minimiseur du risque empirique satisfait :

$\forall P$ la distribution de (X,Y)

$$\mathbb{E}\left[R(\hat{g}(D_n))\right] - R(g^*) \leq C\sqrt{\frac{K}{n}}.$$

On a donc montré que le minimiseur du risque empirique est universellement consistant. En fait, on a montré mieux : la vitesse de convergence est uniforme en la loi de (X,Y) :

$$\lim_{n \to +\infty} \sup_{P \sim (X,Y)} \mathbb{E}\left[R(\hat{g}(D_n))\right] - R(g^*) = 0.$$

$\Rightarrow$ On dit que $\hat{g}$ est uniformément universellement consistant.