

Ex 1]

$\Psi: \mathbb{R} \rightarrow \mathbb{R}$, k -Lipschitz

$$F_0 = \left\{ x \mapsto w^\top x \mid \|w\|_1 \leq B \right\}, \quad O \in F_0$$

$$F_n = \left\{ x \mapsto \Psi \left(v + \sum_{i=1}^m w_i f_i(x) \right) \mid |v| \leq V, \|w\|_1 \leq B, f_i \in F_0 \right\}$$

$$F_P = \left\{ x \mapsto \Psi \left(v + \sum_{i=1}^m w_i^* f_i(x) \right) \mid |v| \leq V, \|w\|_1 \leq B, f_i \in F_P \right\}$$

$$\text{1)} \quad \widehat{R}_n(F_n) \leq \frac{V}{n} \mathbb{E} \left[\sup_{\substack{|v| \leq V \\ \|w\|_1 \leq B \\ f_i \in F_0}} \frac{1}{n} \sum_{i=1}^n \varepsilon_i (v + \sum_{j=1}^m w_j^* f_j(x)) \right]$$

$$= k \left(B \widehat{R}_n(\text{absconv}(F_0)) + \mathbb{E} \left[\sup_{|v| \leq V} \frac{1}{n} \sum \varepsilon_i v \right] \right)$$
$$\leq k \widehat{R}_n(\text{conv}(F_0 \cup (-F_0))) + \frac{V}{n} \mathbb{E} \left[|\sum \varepsilon_i| \right]$$

$$= k \left(B \widehat{R}_n(F_0 \cup (-F_0)) + \frac{V}{n} \mathbb{E} \left[(\sum \varepsilon_i)^2 \right] \right)$$

$$\leq k \left(B \widehat{R}_n(F_0 + (-F_0)) + \frac{V}{n} \mathbb{E} \left[(\sum \varepsilon_i)^2 \right] \right)$$

$$= k \left(2B \widehat{R}_n(F_0) + \frac{V}{n} \sqrt{n} \right)$$

($F_0 = -F_0 \Rightarrow \widehat{R}_n(-F_0) = \widehat{R}_n(F_0)$)

$$\mathbb{E}[(\sum \varepsilon_i)^2] = k \left(2B \widehat{R}_n(F_0) + \frac{V}{n} \right) \quad (O \in F_0 \Rightarrow F_0 \cup F_0 \subset F_0 + (-F_0))$$

$$= V(\sum \varepsilon_i) + \mathbb{E}[\sum \varepsilon_i]$$

$$= \sum_{i=1}^n 1 + \mathbb{E}[\varepsilon_i]$$

$$= n$$

$$2) Supposons \quad \mathcal{X} = \{x \in \mathbb{R}^d \mid \|x\|_\infty \leq 1\}$$

$$\begin{aligned} \widehat{\mathcal{R}}_n(F_0) &= \mathbb{E} \left[\sup_{w \in \mathcal{W}} \sum_{i=1}^n \epsilon_i w^\top x_i \right] \\ &= \mathbb{E} \left[\sup_w \|w\|_1 \left\| \sum_{i=1}^n \epsilon_i x_i \right\|_\infty \right] \\ &\leq \mathbb{E} \left[\sup_w \|w\|_1 \left\| \sum_{i=1}^n \epsilon_i x_i \right\|_\infty \right] \quad (\text{Hölder}) \end{aligned}$$

$$\begin{aligned} &\leq B \mathbb{E} \left[\max_{1 \leq i \leq d} \left| \sum_{j=1}^n \epsilon_j x_i^{(j)} \right| \right] \\ &= B \mathbb{E} \left[\max_{a \in A} \left| \sum_{i=1}^n \epsilon_i a_i \right| \right] \end{aligned}$$

$$\text{où } A = \left\{ \{x_1^{(1)}, \dots, x_n^{(1)}\}, \dots, \{x_1^{(d)}, \dots, x_n^{(d)}\} \right\}$$

$$= B \mathbb{E} \left[\max_{a \in A \cup \{0\}} \sum_{i=1}^n \epsilon_i a_i \right]$$

$$\leq Br\sqrt{2\ln 2d}$$

$$\text{avec } r = \max_{a \in A \cup \{0\}} \|a\|_1 = \max_{a \in A} \left(\sum_{i=1}^n |a_i|^r \right)^{\frac{1}{r}} \leq \left(\sum_{i=1}^n 1^r \right)^{\frac{1}{r}} = \sqrt{n}$$

$$\Rightarrow \widehat{\mathcal{R}}_n(F_0) \leq \frac{B\sqrt{2\ln 2d}}{\sqrt{n}}$$

3) Recurrence : ($\Psi(-u) = -\Psi(u)$)

Test F_0 :

$$\widehat{R}_n(F_0) \leq \frac{1}{\sqrt{n}} \cdot \mathcal{D} \sqrt{2 \ln(2d)} = \frac{\mathcal{D} \sqrt{2 \ln 2d}}{\sqrt{n}} \left(\frac{\epsilon}{2}\right)$$

Supposons p . Montrons $p+1$:

$$\begin{aligned} \widehat{R}_n(F_{p+1}) &= \frac{1}{n} \mathbb{E} \left[\sup_{\|w\|_n \leq V} \sum_{i=1}^n \varepsilon_i \Psi(v + w^\top f(x_i)) \right] \\ &\quad \|w\|_n \leq B \\ &\quad f_i \in F_p \\ &= \frac{1}{n} \mathbb{E} \left[\sup_{\|w\|_n \leq V} \sum_{i=1}^n \varepsilon_i (v + w^\top f(x_i)) \right] \text{ (Lipschitz)} \\ &\quad \|w\|_n \leq B \\ &\quad f_i \in F_p \\ &\leq \frac{1}{n} \mathbb{E} \left[\sup_{\|w\|_n \leq B} \sum_{i=1}^n \varepsilon_i w^\top f(x_i) \right] + \frac{V}{\sqrt{n}} \\ &\quad f_i \in F_p \\ &= \mathcal{D} \widehat{R}_n(\text{absconv}(F_p)) + \frac{V}{\sqrt{n}} \\ &= \mathcal{D} \widehat{R}_n(F_p \cup (-F_p)) + \frac{V}{\sqrt{n}} \end{aligned}$$

Or, comme $F_0 = -F_0$ et que Ψ est impaire, on peut facilement montrer par récurrence que, $\forall p \geq 0$, $F_p = -F_{p-1}$. On obtient donc:

$$\begin{aligned}
\widehat{Q}_n(F_{P,n}) &\leq \mathfrak{B} Q_n(F_P) + \frac{V}{\sqrt{n}} \\
&\leq \frac{\mathfrak{B}}{\sqrt{n}} \left(\mathfrak{B}^{p+1} \sqrt{2 \ln 2d} + V \sum_{l=0}^{p-1} \mathfrak{B}^l \right) + \frac{V}{\sqrt{n}} \\
&= \frac{1}{\sqrt{n}} \left(\mathfrak{B}^{p+2} \sqrt{2 \ln 2d} + \mathfrak{B} V \sum_{l=0}^{p-1} \mathfrak{B}^l + V \right) \\
&= \frac{1}{\sqrt{n}} \left(\mathfrak{B}^{p+2} \sqrt{2 \ln 2d} + V \left(\sum_{l=1}^p \mathfrak{B}^l + 1 \right) \right) \\
&= \frac{1}{\sqrt{n}} \left[\mathfrak{B}^{p+2} \sqrt{2 \ln 2d} + V \sum_{l=0}^p \mathfrak{B}^l \right]
\end{aligned}$$

□

Ex 2]

$$F_\lambda = \left\{ f = \sum_{j=1}^N w_j q_j : N \in \mathbb{N}, q_j \in \mathcal{G}_j, w_j \in \mathbb{R}, \sum_{j=1}^N |w_j| \leq \lambda \right\}$$

$$\begin{aligned} 1) R_n(F_\lambda) &= \mathbb{E} \left[\sup_{\substack{N \in \mathbb{N} \\ q_j \in \mathcal{G}_j \\ \|w_j\|_1 \leq 1}} \left\{ \frac{1}{n} \sum_{i=1}^n \varepsilon_i \sum_{j=1}^N w_j q_j(x) \right\} \Big| D_n \right] \\ &= \lambda \mathbb{E} \left[\sup_{\substack{N \in \mathbb{N} \\ q_j \in \mathcal{G}_j \\ \|w_j\|_1 \leq 1}} \left\{ \frac{1}{n} \sum_{i=1}^n \varepsilon_i \sum_{j=1}^N w_j q_j(x) \right\} \Big| D_n \right] \\ &= \lambda R_n(\text{absconv}(g)) \\ &= \lambda R_n(\text{conv}(g \vee -g)) \\ &= \lambda R_n(g) \end{aligned}$$

On suppose que $g_j = -g_j$ (Fait sens: inverser tous les labels ne change pas la classification)

$$\text{Ainsi, } g_j \vee (-g_j) = g_j$$

$$\text{On en déduit } \lambda R_n(g_j \vee -g_j) = \lambda R_n(g_j)$$

$$\leq \lambda \left(\frac{2 \log \delta(q_{\text{err}})}{n} \right)^{1/2} \quad (\text{Corollary 2.1b})$$

$$\leq \lambda \left(\frac{2 \log \left(\frac{en}{\delta} \right)}{n} \right)^{1/2} \quad (\text{Sauer's Lemma})$$

2) $\Psi(n) = \log_2 (1 + e^n)$

$$A(f) = \mathbb{E} [\Psi(-Y f(x))] \quad \left\{ \begin{array}{l} f^*(x) = \ln \left(\frac{2(x)}{1 - 2(x)} \right) \\ \text{ex 7.6 T01} \\ p. 11 \end{array} \right.$$

$$A(f^*) = \mathbb{E} \left[q(x) \underbrace{\log_2 \left(1 + \frac{1-q(x)}{q(x)} \right)}_{q=1} + (1-q(x)) \underbrace{\log_2 \left(1 + \frac{1}{1-q(x)} \right)}_{q=-1} \right]$$

$$= \mathbb{E} \left[q(x) \log_2 \left(\frac{1}{q(x)} \right) + (1-q(x)) \log_2 \left(\frac{1}{1-q(x)} \right) \right]$$

$$= \mathbb{E} \left[-q(x) \log_2 (q(x)) - (1-q(x)) \log_2 (1-q(x)) \right]$$

$$= \mathbb{E} \left[-q(x) \log_2 (q(x)) - \log (1-q(x)) + q(x) \log (1-q(x)) \right]$$

$$= \mathbb{E} \left[q(x) \log_2 \left(\frac{1-q(x)}{q(x)} \right) - \log (1-q(x)) \right]$$

$$\Rightarrow H(t) = t \log_2 \left(\frac{1-t}{t} \right) - \log_2 (1-t)$$

$$3) H(t) = H\left(\frac{1}{2}\right) + H'\left(\frac{1}{2}\right)\left(t - \frac{1}{2}\right) + \frac{H''(v)}{2} \left(t - \frac{1}{2}\right)^2$$

On a :

$$H\left(\frac{1}{2}\right) = 1, \quad H'(t) = \log_2\left(\frac{2t}{t}\right) - \frac{1}{1-t} + \frac{1}{t-1}$$

$$= \log_2\left(\frac{1-t}{t}\right)$$

$$H'\left(\frac{1}{2}\right) = 0$$

$$H''(t) = \frac{1}{\ln 2} \left(\ln\left(\frac{1-t}{t}\right) \right)'$$

$$= \frac{1}{\ln 2} \left(\frac{t}{1-t} \times -\frac{1}{t^2} \right)$$

$$= \frac{-1}{\ln 2 t (1-t)}.$$

Or, $\forall t \in [0,1]$, $t(1-t) \leq \frac{1}{4}$. Ainsi,

$$H''(t) \leq -\frac{4}{\ln 2}$$

On en déduit

$$H(t) \leq 1 - \frac{2}{\ln 2} \left(t - \frac{1}{2}\right)^2$$

$$= 1 - \frac{2}{\ln 2} \left(\frac{1-2t}{2}\right)^2 = 1 - \left(\sqrt{\frac{2}{\ln 2}} \frac{1-2t}{2}\right)^2$$

$$\Rightarrow H(t) \leq 1 - \left(\frac{1-t}{2c} \right)^L, \quad c \triangleq \sqrt{\frac{\ln 2}{2}}.$$

h) φ est une fonction de cost et :

$$\begin{aligned} \forall x, \quad & \inf_{\alpha \in \mathbb{R}} (g(x)\varphi(-\alpha) + (1-g(x))\varphi(\alpha)) \\ &= g(x)\varphi(-f_{\varphi}^x) + (1-g(x))\varphi(f_{\varphi}^x) \quad (\text{Voir page 11}) \end{aligned}$$

On a donc

$$H(g(x)) = \inf_{\alpha \in \mathbb{R}} g(x)\varphi(-\alpha) + (1-g(x))\varphi(\alpha)$$

$$\text{De plus, } H(t) \leq 1 - \left(\frac{1-t}{2c} \right)^L$$

$$\Rightarrow H(t) \leq 1 - c^{-2} \left(\frac{1}{2} - t \right)^L$$

$$\Rightarrow \left(\frac{1}{2} - t \right)^L \leq c^2 (1 - H(t))$$

$$\Rightarrow \left(\frac{1}{2} - t \right)^L \leq \frac{\ln 2}{2} (1 - H(t))$$

L'inégalité de Zhang permet de conclure :

$$\forall f, \quad L(f) - L^* \leq 2c (A(f) - A^*)^{1/2}$$

$$\Rightarrow \frac{L(f) - L^*}{(A(f) - A^*)^{1/2}} \leq \ln 2.$$

5) On pose $F = \Psi_0 \tilde{F}$, $\tilde{F} = \{(x, y) \mapsto -y \text{ for } f \in F_x\}$

On peut appliquer le théorème 2.10 : avec probabilité $1-\delta$:

$$\sup_{h \in F} \left(E[h(z)] - \frac{1}{n} \sum h(z_i) \right) \leq 2R_n(F) + c \sqrt{\frac{\log(1/\delta)}{2n}}$$

On cherche c . Il suffit de borner toute fonction de F :

$$\forall h \in F, |h(z)| = |\Psi(f_y f(x))| = \log_2 \left(1 + e^{-y \sum w_i g_i(x)} \right)$$

$$\leq \log_2 (1 + e^k) \triangleq c$$

$$(g_i \in \{-1, 1\})$$

De plus, $\Psi'(x) = \underbrace{\frac{c}{1+e^x}}_{< 1} \frac{1}{\ln 2} \leq \frac{1}{\ln 2} \Rightarrow \Psi(x)$ est $\frac{1}{\ln 2}$ -Lipschitz

On a :

$$\sup_{f \in F_x} |\hat{A}_n(f) - A(f)| \leq \frac{2}{\ln 2} R_n(F_x) + \log_2(1 + e^k) \sqrt{\frac{\log(1/\delta)}{2n}}$$

$$\Rightarrow \sup_{f \in F_\lambda} |\widehat{A}_n(f) - A(f)| \leq \underbrace{\left(\frac{2\lambda}{\ln 2} \right)}_{C_1(\lambda)} \sqrt{\underbrace{\frac{\ln \left(\frac{e_n}{\nu} \right)}{n}}_{C_2(\lambda)} + \underbrace{\log_2(1+e^\lambda)}_{C_3(\lambda)} \frac{\ln \frac{1}{\delta}}{2n}} \quad (7)$$

$$6) L(\widehat{f}_{n,\lambda}) - L^* \leq \ln 2 \left(A(\widehat{f}_{n,\lambda}) - A^* \right)^{\frac{1}{k_2}} \stackrel{q=1}{=} h$$

$$\leq \ln 2 \left[\left(A(\widehat{f}_{n,\lambda}) - \inf_{f \in F_\lambda} A(f) \right)^{\frac{1}{k_2}} + \left(\inf_{f \in F_\lambda} A(f) - A^* \right)^{\frac{1}{k_2}} \right]$$

$$\text{Or, } A(\widehat{f}_{n,\lambda}) - A(\bar{f})$$

$$= A(\widehat{f}_{n,\lambda}) - A(\bar{f}) - \widehat{A}_n(\widehat{f}_{\lambda,n}) + \widehat{A}_n(\widehat{f}_{\lambda,n})$$

$$= A(\widehat{f}_{n,\lambda}) - \widehat{A}_n(\widehat{f}_{\lambda,n}) + \widehat{A}_n(\bar{f}) - A(\bar{f}) + \underbrace{\widehat{A}_n(\widehat{f}_{\lambda,n}) - \widehat{A}_n(\bar{f})}_{\leq 0}$$

$$\leq \left(2 \sup_{f \in F_\lambda} |A(f) - \widehat{A}_n(f)| \right) \leq 0$$

$$(\bar{f}, \widehat{f}_{\lambda,n} \in F_\lambda)$$

Cela nous donne la borne suivante :

$$L(\widehat{f}_{n,\lambda}) - L^* \leq \ln 2 \left[\left(2 \sup_{f \in F_\lambda} |A(f) - \widehat{A}_n(f)| \right)^{\frac{1}{k_2}} + (A(\bar{f}) - A^*)^{\frac{1}{k_2}} \right]$$

Utiliser (7) permet de conclure.

$$f^* = \arg \min_f A(f)$$

$$= \arg \min_f E[\varphi(-Y f(x)) \mid X=x]$$

$$= \arg \min_f \underbrace{\varphi(x) \varphi(-f(x)) + (1-\varphi(x)) \varphi(f(x))}_{h_{\varphi(x)} f(x)}$$

Ainsi,

$$f^*(x) = \arg \min_\alpha h_{\varphi(x)}(\alpha).$$

Comme φ est convexe, $h_{\varphi(x)}$ l'est aussi. Si φ est dérivable, on a :

$$x^* \in \arg \min_\alpha h_{\varphi(x)}(\alpha) \Leftrightarrow h'_{\varphi(x)}(x^*) = 0.$$

$$\Leftrightarrow \frac{\varphi(x)}{1-\varphi(x)} = \frac{\varphi'(x^*)}{\varphi'(f(x^*))}.$$

$$\text{Posons } \varphi(x) = \log_2(1 + e^x).$$

$$\text{On a } \varphi'(x) = \frac{e^x}{e^x \ln(2) + \ln(1)}. \text{ On cherche } x^* \text{ tq}$$

$$\begin{aligned}
 \frac{\varphi(\alpha)}{1 - \varphi(\alpha)} &= \frac{e^{\alpha^*}}{e^{\alpha^*} \ln 2 + \ln 2} \cdot \frac{e^{-\alpha^*} \ln 2 + \ln 2}{e^{-\alpha^*}} \\
 &= \frac{e^{\alpha^*} e^{-\alpha^*} \ln 2 + e^{\alpha^*} \ln 2}{e^{-\alpha^*} e^{\alpha^*} \ln 2 + \ln 2 e^{-\alpha^*}} \\
 &= \frac{\ln 2 + e^{\alpha^*} \ln 2}{\ln 2 + e^{\alpha^*} \ln 2} \\
 &= \frac{1 + e^{\alpha^*}}{1 + e^{-\alpha^*}} \\
 &= e^{\alpha^*} \cdot \frac{1 + e^{\alpha^*}}{e^{\alpha^*} + 1} \\
 &= e^{\alpha^*}
 \end{aligned}$$

$$\Leftrightarrow \alpha^* = \ln \left(\frac{\varphi(\alpha)}{1 - \varphi(\alpha)} \right)$$