

$$D_n = (Z_1, \dots, Z_n) \text{ i.i.d.}$$

$$\hat{Q}_n(F) = \mathbb{E} \left[\sup_{f \in F} \left\{ \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(Z_i) \right\} \middle| D_n \right]$$

$$Q_n(F) = \mathbb{E}_{D_n} [Q_n(F)]$$

$$\begin{aligned} P(\varepsilon_i = 1) &= P(\varepsilon_i = -1) = P(-\varepsilon_i = 1) \\ &= P(\varepsilon_i = -1) = \frac{1}{2} \end{aligned}$$

$$\mathbb{E}_{\varepsilon_i} [\varepsilon_i x] = \mathbb{E}_{\varepsilon_i} [-\varepsilon_i x]$$

Ex 1)

1) $c \in \mathbb{R}$.

$$\hat{Q}_n(c\gamma) = \mathbb{E} \left[\sup_{g \in \gamma} \frac{1}{n} \sum \varepsilon_i g(Z_i) \middle| D_n \right]$$

$$= \mathbb{E} \left[\sup_{f \in \gamma} \frac{1}{n} \sum c \varepsilon_i f(Z_i) \middle| D_n \right]$$

$$= \mathbb{E} \left[\sup_{f \in \gamma} \frac{1}{n} \sum |c| \underbrace{\text{sgn}(c)}_{\sim \varepsilon_i} \varepsilon_i f(Z_i) \middle| D_n \right]$$

$$P(\varepsilon_i = -1) = P(\varepsilon_i = 1) = \frac{1}{2}$$

(ε_i prend -1 et 1 avec la même proba.)

Mettre un (-1) devant ε_i change l'ordre dans lequel on va sommer les termes pour calculer l'espérance mais pas sa valeur finale).

$$= \mathbb{E} \left[\sup_{f \in T} \frac{1}{n} \sum \varepsilon_i f(z_i) \mid D_n \right]$$

$$= \mathbb{E} \left[\widehat{R}_n(T) \right]$$

$$\sup (cA) = \sup \{c - x \mid x \in A\}$$

$$= |c| \sup(A).$$

2) On suppose $T_1 \subseteq T_2$.

$$\text{On a } T_1 \subseteq T_2 \Rightarrow \sup(T_1) \leq \sup(T_2).$$

Si $T_1 = T_2 \Rightarrow$ direct

Si $T_1 \subset T_2$. Supposons $\sup_{\varepsilon_n} (T_1) > \sup_{\varepsilon_2} (T_2)$.

Par def du sup, $\forall x \in T_2, x \leq \varepsilon_2$.

Or, $T_1 \subset T_2$, donc $\forall x \in T_1 \Rightarrow x \in T_2$ donc $x \leq \varepsilon_2$.

Par def du sup, $\varepsilon_1 \leq \varepsilon_2 \Rightarrow$ Contradiction.

$$T_1 \subseteq T_2 \Rightarrow \left\{ \frac{1}{n} \sum \varepsilon_i f(z_i) \mid f \in T_1 \right\} \subseteq \left\{ \frac{1}{n} \sum \varepsilon_i f(z_i) \mid f \in T_2 \right\}.$$

$$\Rightarrow \sup \{ \quad \} \leq \sup \{ \quad \}$$

$$\text{Ainsi, } \widehat{R}_n(T_1) = \mathbb{E} \left[\sup_{f \in T_1} \left\{ \frac{1}{n} \sum \varepsilon_i f(z_i) \right\} \mid D_n \right]$$

$$\leq \mathbb{E} \left[\sup_{f \in T_2} \left\{ \frac{1}{n} \sum \varepsilon_i f(z_i) \right\} \mid D_n \right] = \widehat{R}_n(T_2)$$

$$3) A+B = \{a+b \mid a \in A, b \in B\}$$

$$\text{Mq } \sup(A+B) = \sup(A) + \sup(B)$$

$$\forall a, b, \quad a \leq \sup(A+B) - b$$

$$\text{Par def, } \sup(A) \leq \sup(A+B) - b$$

$$b \leq \sup(A+B) - \sup(A) \Rightarrow \sup(B) \leq \sup(A+B) - \sup(A)$$

$$\sup(A) + \sup(B) \leq \sup(A+B)$$

$$\text{Def sup, } \forall s, (\forall a, b, a+b \leq s), \sup(A+B) \leq s.$$

$$\text{Donc, } \sup(A+B) \leq \sup(A) + \sup(B)$$

$$\text{et } // \quad = \quad //$$

$$\begin{aligned} \widehat{Q}_n(T_1 + T_2) &= \mathbb{E} \left[\sup_{f \in T_1 + T_2} \frac{1}{n} \sum \varepsilon_i f(z_i) \mid \mathcal{D}_n \right] \\ &= \mathbb{E} \left[\sup_{\substack{f_1 \in T_1 \\ f_2 \in T_2}} \frac{1}{n} \sum \varepsilon_i (f_1(z_i) + f_2(z_i)) \mid \mathcal{D}_n \right] \\ &= \mathbb{E} \left[\sup_{\substack{f_1 \in T_1 \\ f_2 \in T_2}} \frac{1}{n} \left(\sum \varepsilon_i f_1(z_i) + \sum \varepsilon_i f_2(z_i) \right) \mid \mathcal{D}_n \right] \\ &= \widehat{Q}_n(T_1) + \widehat{Q}_n(T_2) \end{aligned}$$

$$h) \operatorname{conv}(A) = \left\{ x \mid x = \sum_{i=1}^p \lambda_i a_i, a_i \in A, \lambda_i \geq 0, \|x\| = 1 \right\}.$$

$$\arg \max_{\lambda_i \geq 0, \|x\|=1} \sum \lambda_i a_i = \begin{cases} 1 & \text{si } z = \arg \max(a_i) \\ 0 & \text{sinon} \end{cases}$$

$$\hat{Q}_n(\operatorname{conv}(T)) = \mathbb{E} \left[\sup_{\substack{p \in \mathbb{N} \\ (f_p) \in T \\ \lambda_i \geq 0 \\ \|x\|=1}} \frac{1}{n} \sum_{i=1}^n \mathbb{E}_i \left[\sum_{j=1}^p \lambda_j f_j(z_i) \right] \mid D_n \right]$$

$$= \mathbb{E} \left[\frac{1}{n} \sup_{\| \cdot \|} \sum_{j=1}^p \sum_{i=1}^n \mathbb{E}_i \left[\lambda_j f_j(z_i) \right] \mid D_n \right]$$

$$= \mathbb{E} \left[\frac{1}{n} \sup_p \sum_{j=1}^p \lambda_j \left(\sum_{i=1}^n \mathbb{E}_i \left[f_j(z_i) \right] \mid D_n \right) \right]$$

$$= \mathbb{E} \left[\frac{1}{n} \sup_{\substack{p \in \mathbb{N} \\ (f_p) \in T}} \max_{f \in \{f_p\}} \sum_{i=1}^n \mathbb{E}_i \left[f(z_i) \right] \mid D_n \right]$$

$$= \mathbb{E} \left[\frac{1}{n} \sup_{f \in T} \sum_{i=1}^n \mathbb{E}_i \left[f(z_i) \right] \mid D_n \right]$$

$$= \hat{Q}_n(T).$$

Ex 2)

$$1. \mathcal{B}_\infty(C) = \{x \in \mathbb{R}^d \mid \|x\|_\infty \leq C\} \quad \mathcal{G} = \{x \in \mathcal{B}_\infty(C) \mapsto w^T x \mid \|w\|_1 \leq B\}$$

$$\hat{R}_n(\mathcal{G}) = \mathbb{E} \left[\sup_{f \in \mathcal{G}} \frac{1}{n} \sum \varepsilon_i f(z_i) \mid \mathcal{D}_n \right]$$

$$= \mathbb{E} \left[\sup_{\substack{w \in \mathbb{R}^d \\ \|w\|_1 \leq B}} \frac{1}{n} \sum \varepsilon_i w^T z_i \mid \mathcal{D}_n \right]$$

$$= \mathbb{E} \left[\sup_{\substack{w \in \mathbb{R}^d \\ \|w\|_1 \leq B}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^d w^{(j)} \varepsilon_i z_i^{(j)} \mid \mathcal{D}_n \right]$$

$$= \mathbb{E} \left[\sup_{\substack{w \in \mathbb{R}^d \\ \|w\|_1 \leq B}} \frac{1}{n} \sum_{i=1}^d \omega^{(i)} \left| \sum_{i=1}^n \varepsilon_i z_i^{(i)} \right| \mid D_n \right]$$

$$= \mathbb{E} \left[\sup_{\substack{w \in \mathbb{R}^d \\ \|w\|_1 \leq B}} \frac{1}{n} w^T \left(\sum_{i=1}^n \varepsilon_i z_i \right) \mid D_n \right]$$

$$\leq \mathbb{E} \left[\sup_{\substack{w \in \mathbb{R}^d \\ \|w\|_1 \leq B}} \frac{1}{n} \|w\|_1 \left\| \sum_{i=1}^n \varepsilon_i z_i \right\|_\infty \mid D_n \right] \quad \left(\text{Holder} \right. \\ \left. |a \cdot b| \leq \|a\|_1 \|b\|_\infty \right)$$

$$\leq \mathbb{E} \left[\frac{B}{n} \left\| \sum_{i=1}^n \varepsilon_i z_i \right\|_\infty \mid D_n \right]$$

$$= \frac{B}{n} \mathbb{E} \left[\max_j \left| \sum \varepsilon_i z_i^{(j)} \right| \mid D_n \right]$$

$$= \frac{B}{n} \mathbb{E} \left[\sup_{x \in C} \left| \sum \varepsilon_i x_i \right| \mid D_n \right], \quad C = \left\{ \begin{pmatrix} z_n^{(1)} \\ \vdots \\ z_n^{(d)} \end{pmatrix}, \dots, \begin{pmatrix} z_n^{(1)} \\ \vdots \\ z_n^{(d)} \end{pmatrix} \right\}$$

$$= \frac{B}{n} \mathbb{E} \left[\sup_{x \in C \cup (-C)} \sum \varepsilon_i x_i \mid D_n \right]$$

Massart

$$\leq \frac{B \max_{x \in C} \|x\|_2 \sqrt{2 \ln |C \cup (-C)|}}{n}$$

$$\leq \frac{B C \sqrt{2 \ln (2d)}}{\sqrt{n}}$$

$\max_{x \in C}$

$$\|x\|_2 = \max_i \left(\sum z_i^{(j)} \right)^{1/2}$$

$$\leq \left(\sum c^2 \right)^{1/2}$$

$$= (n c^4)^{1/2}$$

$$= \sqrt{n} c$$

$$2) \quad \kappa(x, x') = \langle \phi(x), \phi(x') \rangle.$$

$$K = \left(\kappa(x_i, x_j) \right)_{i \in n, j \in n}$$

$$\mathcal{H} = \left\{ x \mapsto \langle w, \phi(x) \rangle \mid \|w\|_K \leq M \right\}.$$

$$\hat{r}_n(\mathcal{H}) = \mathbb{E} \left[\sup_{\|w\|_K \leq M} \frac{1}{n} \sum \varepsilon_i \langle w, \phi(x_i) \rangle \mid \mathcal{D}_n \right]$$

$$= \frac{1}{n} \mathbb{E} \left[\sup_{\|w\|_K \leq M} \langle w, \sum \varepsilon_i \phi(x_i) \rangle_K \mid \mathcal{D}_n \right]$$

$$\leq \frac{1}{n} \mathbb{E} \left[\sup_{\|w\|_K \leq M} \|w\|_K \left\| \sum \varepsilon_i \phi(x_i) \right\|_K \mid \mathcal{D}_n \right]$$

$$\leq \frac{M}{n} \mathbb{E} \left[\left\| \sum \varepsilon_i \phi(x_i) \right\|_K \mid \mathcal{D}_n \right]$$

$$= \frac{M}{n} \mathbb{E} \left[\left(\left\langle \sum \varepsilon_i \phi(x_i), \sum \varepsilon_i \phi(x_i) \right\rangle_K \right)^{\frac{1}{2}} \mid \mathcal{D}_n \right]$$

$$\leq \frac{M}{n} \mathbb{E} \left[\left\langle \sum \varepsilon_i \phi(x_i), \sum \varepsilon_i \phi(x_i) \right\rangle_K \mid \mathcal{D}_n \right]^{\frac{1}{2}}$$

Sensen

$\psi_{\text{concave}} =$

$$\mathbb{E}[\psi(x)] \leq \psi(\mathbb{E}[x])$$

$\psi_{\text{convex}} =$

$$\mathbb{E}[\psi(x)] \geq \psi(\mathbb{E}[x])$$

$$= \frac{M}{n} \mathbb{E} \left[\sum_{i,j} \langle \varepsilon_i \phi(x_i), \varepsilon_j \phi(x_j) \rangle_K \mid \mathcal{D}_n \right]^{\frac{1}{2}}$$

$$= \frac{M}{n} \mathbb{E} \left[\sum_i \varepsilon_i \sum_j \varepsilon_j \langle \phi(x_i), \phi(x_j) \rangle_K \mid \mathcal{D}_n \right]^{\frac{1}{2}}$$

$$= \frac{M}{n} \mathbb{E} \left[\sum_{i \neq j} \varepsilon_i \varepsilon_j \langle \phi(x_i), \phi(x_j) \rangle_K + \sum_i \varepsilon_i^2 \|\phi(x_i)\|_K^2 \right]^{\frac{1}{2}}$$

$$= \frac{M}{n} \mathbb{E} \left[\sum_{i \neq j} \mathbb{E}[\varepsilon_i \varepsilon_j] \langle \phi(x_i), \phi(x_j) \rangle_K + \sum_i \mathbb{E}[\varepsilon_i^2] \|\phi(x_i)\|_K^2 \right]^{\frac{1}{2}}$$

$$\begin{aligned}
 &= \frac{M}{n} \left(\sum_{i \neq j} \langle \phi(x_i), \phi(x_j) \rangle_{\kappa} \mathbb{E}[\varepsilon_i] \mathbb{E}[\varepsilon_j] + \sum_i \|\phi(x_i)\|_{\kappa}^2 \right)^{\frac{1}{2}} \\
 &= \frac{M}{n} \left(\sum_i \|\phi(x_i)\|_{\kappa}^2 \right)^{\frac{1}{2}} \\
 &= \frac{M \sqrt{\text{trace}(\kappa)}}{n}
 \end{aligned}$$

Ex 3

$$r(\mathcal{G}, n) = \max_{(x_i)_{i \leq n} \in \mathbb{R}^d} \left| \{ (g(x_i))_{i \leq n} \mid g \in \mathcal{G} \} \right|$$

Nombre maximum de manières de classer n points en utilisant des fonctions dans $\mathcal{G}$

$$1) \bar{R}_n(\mathcal{G}) = \mathbb{E}_{D_n} \left[\mathbb{E} \left[\sup_{g \in \mathcal{G}} \frac{1}{n} \sum \varepsilon_i g(z_i) \mid D_n \right] \right]$$

Posons $\mathcal{G}|_{D_n}$ l'ensemble des vecteurs $(g(z_1), \dots, g(z_n))^T$ pour toute fonction dans $\mathcal{G}$: $\mathcal{G}|_{D_n} = \{ (g(z_i))_{i \leq n}^T \mid g \in \mathcal{G} \}$.

$\mathcal{G}_n$ obtient

$$\bar{R}_n(\mathcal{G}) = \mathbb{E}_{D_n} \left[\mathbb{E} \left[\sup_{z \in \mathcal{G}|_{D_n}} \frac{1}{n} \sum \varepsilon_i z_i \mid D_n \right] \right]$$

Comme $g \in \mathcal{G}$ est à valeurs dans $\{-1, 1\}$,

$$\max_{z \in \mathcal{G}_{|D_n}} \|z\|_2 \leq \sqrt{n}.$$

On peut donc appliquer le lemme de Massart:

$$\mathbb{E}_{D_n} \left[\mathbb{E} \left[\sup_{z \in \mathcal{G}_{|D_n}} \frac{1}{n} \sum \varepsilon_i z_i \mid D_n \right] \right] \leq \mathbb{E}_{D_n} \left[\frac{\sqrt{n} \sqrt{2 \ln |\mathcal{G}_{|D_n}|}}{n} \right].$$

Or, $\mathcal{G}_{|D_n} \subseteq \{(g(x_i))_{i \leq n} \mid (x_i)_{i \leq n} \in \mathbb{R}^d, g \in \mathcal{G}\}$. Ainsi,

$$|\mathcal{G}_{|D_n}| \leq \max_{(x_i)_{i \leq n} \in \mathbb{R}^d} |\{(g(x_i))_{i \leq n} \mid g \in \mathcal{G}\}| = \gamma(\mathcal{G}, n).$$

On en déduit

$$\begin{aligned} \mathbb{E}_{D_n} \left[\frac{\sqrt{n} \sqrt{2 \ln |\mathcal{G}_{|D_n}|}}{n} \right] &\leq \frac{\sqrt{n} \sqrt{2 \ln \gamma(\mathcal{G}, n)}}{n} \\ &= \left(\frac{2 \ln \gamma(\mathcal{G}, n)}{n} \right)^{1/2}. \end{aligned}$$

On suppose $V(\mathcal{G}) < +\infty$.

Si $n \leq V(\mathcal{G})$, $\gamma(\mathcal{G}, n) = 2^n$. Ainsi

$$\mathcal{R}_n(\mathcal{G}) \leq \left(\frac{2 \ln(2^n)}{n} \right)^{1/2} \leq \left(\frac{2 \log_2 2^n}{n} \right)^{1/2} = \sqrt{2}.$$

Si $n > v(G)$:

$$\chi(G, n) \leq \left(\frac{en}{v(G)} \right)^{v(G)} \quad \text{Parsons } V = v(G)$$

α_n en déduit

$$\overline{\alpha}_n(G) \leq \left(\frac{2 \ln \left(\frac{en}{v} \right)^v}{n} \right)^{1/2}$$

$$= \left(\frac{2 \left(v + v \ln \left(\frac{n}{v} \right) \right)}{n} \right)^{1/2}$$

$$= \left(\frac{2v \left(\ln \frac{n}{v} + 1 \right)}{n} \right)^{1/2}$$

$$= \left(\frac{v \ln \left(\frac{n}{v} \right)}{n} \left(2 + \frac{2}{\ln \left(\frac{n}{v} \right)} \right) \right)^{1/2}$$

$$= \left(\frac{v \ln n \left(2 - \frac{2 \ln v}{\ln n} + \frac{2}{\ln n} \right)}{n} \right)^{1/2}$$

$$= \left(\frac{v \ln n}{n} \right)^{1/2} \left(\frac{2 \left(1 + \frac{\ln v}{\ln n} + \frac{1}{\ln n} \right)}{n} \right)^{1/2}$$

$$\left(\log_2^n = \frac{\ln n}{\ln 2} \right) \gg \left(\frac{v \log_2 n}{n} \right)^{1/2} C$$

ii
C

E_X h

$$1) \quad \phi: \mathcal{X} \rightarrow \mathcal{Y}, \quad |\phi(x_1) - \phi(x_2)| \leq \kappa |x_1 - x_2|$$

$$\hat{Q}_n(\phi \circ f) = \mathbb{E} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \varepsilon_i (\phi \circ f)(Z_i) \mid D_n \right]$$

$$= \mathbb{E} \left[\frac{1}{n} \sup_{f \in \mathcal{F}} \sum_{i=1}^{n-1} \varepsilon_i (\phi \circ f)(Z_i) + \varepsilon_n (\phi \circ f)(Z_n) \mid D_n \right]$$

$$= \frac{1}{n} \mathbb{E} \left[\sup_{f \in \mathcal{F}} \mu_{n-1}(f) + \varepsilon_n (\phi \circ f)(Z_n) \mid D_n \right]$$

$$= \frac{1}{n} \mathbb{E} \left[\mathbb{E}_{\varepsilon_{n-1}} \left[\mathbb{E}_{\varepsilon_n} [\mu_{n-1}(f) + \varepsilon_n (\phi \circ f)(Z_n) \mid Z_n] \mid D_{n-1} \right] \right] \quad (1)$$

Def sup: $\forall \varepsilon > 0, \exists x \in A, x \geq \sup(A) - \varepsilon$.

Par définition du sup, $\forall \varepsilon > 0, \exists f_1, f_2 \in \mathcal{F}$ t.q.: (sans perte de gen, on suppose la sup > 0)

$$\mu_{n-1}(f_1) + (\phi \circ f_1)(Z_n) \geq (1-\varepsilon) \left(\sup_{f \in \mathcal{F}} \mu_{n-1}(f) + (\phi \circ f)(Z_n) \right)$$

$$\text{et } \mu_{n-1}(f_2) - (\phi \circ f_2)(Z_n) \geq (1-\varepsilon) \left(\sup_{f \in \mathcal{F}} \mu_{n-1}(f) - (\phi \circ f)(Z_n) \right)$$

Ainsi,

$$(1-\varepsilon) \mathbb{E}_{\varepsilon_n} \left[\sup_{f \in \mathcal{F}} \mu_{n-1}(f) + \varepsilon_n (\phi \circ f)(Z_n) \right] = (1-\varepsilon) \left[\frac{1}{2} \sup_{f \in \mathcal{F}} \mu_{n-1}(f) + (\phi \circ f)(Z_n) \right. \\ \left. + \frac{1}{2} \sup_{f \in \mathcal{F}} \mu_{n-1}(f) - (\phi \circ f)(Z_n) \right]$$

$$\leq \frac{1}{2}(\mu_{n,n}(f_n) + (\phi \circ f_n)(z_n)) + \frac{1}{2}(\mu_{n,n}(f_n) - (\phi \circ f_n)(z_n))$$

$$= \frac{1}{2}(\mu_{n,n}(f_n) + \mu_{n,n}(f_n) + (\phi \circ f_n)(z_n) - (\phi \circ f_n)(z_n))$$

Posons $s = \text{sign}(f_n(z_n) - f_n(z_n))$. Par propriété de Lipschitz, on a

$$\phi(f_n(z_n)) - \phi(f_n(z_n)) \leq |\phi(f_n(z_n)) - \phi(f_n(z_n))| \leq \kappa s (f_n(z_n) - f_n(z_n)).$$

L'inégalité précédente implique donc :

$$(1-\varepsilon) \mathbb{E}_{\varepsilon_n} \left[\sup_{f \in \mathcal{F}} \mu_{n,n}(f) + \varepsilon_n (\phi \circ f)(z_n) \right] \leq \frac{1}{2} (\mu_{n,n}(f_n) + \mu_{n,n}(f_n) + \kappa s (f_n(z_n) - f_n(z_n)))$$

$$= \frac{1}{2} (\mu_{n,n}(f_n) + \kappa s f_n(z_n)) + \frac{1}{2} (\mu_{n,n}(f_n) - \kappa s f_n(z_n))$$

$$\leq \frac{1}{2} \sup_{f \in \mathcal{F}} \mu_{n,n}(f) + \kappa s f(z_n) + \frac{1}{2} \sup_{f \in \mathcal{F}} \mu_{n,n}(f) - \kappa s f(z_n)$$

$$= \mathbb{E}_{\varepsilon_n} \left[\sup_{f \in \mathcal{F}} \mu_{n,n}(f) + \underbrace{\varepsilon_n \kappa}_{\sim \varepsilon_n} f(z_n) \right]$$

$$= \mathbb{E}_{\varepsilon_n} \left[\sup_{f \in \mathcal{F}} \mu_{n,n}(f) + \varepsilon_n \kappa f(z_n) \right]$$

Puisque cette inégalité est vraie pour tout $\varepsilon > 0$, cela implique que

$$\mathbb{E}_{\varepsilon_n} \left[\sup_{f \in \mathcal{F}} \mu_{n,n}(f) + \varepsilon_n (\phi \circ f)(z_n) \right] \leq \mathbb{E}_{\varepsilon_n} \left[\sup_{f \in \mathcal{F}} \mu_{n,n}(f) + \varepsilon_n \kappa f(z_n) \right].$$

On en déduit l'inégalité suivante,

$$\mathbb{E} \left[\mathbb{E} \left[\sup_{f \in F} u_{n-1}(f) + \varepsilon_n (\phi \circ f)(z_n) \mid \mathcal{Z}_n \right] \mid \mathcal{D}_{n-1} \right] \quad (1)$$

$$\leq \mathbb{E} \left[\mathbb{E} \left[\sup_{f \in F} u_{n-1}(f) + \varepsilon_n K f(z_n) \mid \mathcal{Z}_n \right] \mid \mathcal{D}_{n-1} \right]$$

$$= \mathbb{E} \left[\sup_{f \in F} u_{n-1}(f) + \varepsilon_n K f(z_n) \mid \mathcal{D}_n \right].$$

Procéder de la même manière pour tous les autres $(\varepsilon_i)_{i \leq n}$ termine la preuve.

2) Posons $F' = \{ (x, y) \mapsto y f(x) \mid f \in F \}$

et $\tilde{F} = \{ \phi_p \circ f \mid f \in F' \}$.

$\forall f \in \tilde{F}$, f est $\{0, 1\}$ -valuée. On applique donc un théorème du cours (TD2):

$$\forall f \in \tilde{F}, \mathbb{E}[f(z)] \leq \underbrace{\frac{1}{n} \sum_{i=1}^n f(z_i)}_{\text{a.p. } 1-\delta} + 2 \mathcal{R}_n(\tilde{F}) + \left(\frac{\ln^{1/8}}{2n} \right)^{\frac{1}{2}}$$

$\Leftrightarrow$

$$\forall f \in F, \mathbb{E}[\phi_p(y f(x))] \leq \hat{L}_{n,p}(f) + 2 \mathcal{R}_n(\phi \circ F') + \left(\frac{\ln^{1/8}}{2n} \right)^{\frac{1}{2}}$$

Or, $\forall n \in \mathbb{N}$, $\mathbb{1}_{\{n \leq 0\}} \in \Phi_p(n)$. Ainsi

$$\mathbb{E}[\mathbb{1}_{\{Y f(x) \leq 0\}}] \leq \mathbb{E}[\Phi_p(Y f(x))] \text{ et donc}$$

$$L(f) \leq \widehat{L}_{n,p}(f) + 2 \mathcal{R}_n(\Phi_p \circ F') + \left(\frac{\ln \frac{1}{\delta}}{2n}\right)^{\frac{1}{2}}$$

De plus,
$$\Phi_p'(x) = \begin{cases} 0 & \text{si } x \leq -p \\ -x & \text{si } -p \leq x \leq p \\ 0 & \text{si } x \geq p \end{cases}$$

Φ_p' est donc $\frac{1}{p}$ -Lipschitz.

On obtient ainsi :

$$L(f) \leq \widehat{L}_{n,p}(f) + \frac{2}{p} \mathcal{R}_n(F') + \ln\left(\frac{1}{\delta}\right)^{\frac{1}{2}}$$

Finalement,
$$\begin{aligned} \mathcal{R}_n(F') &= \frac{1}{n} \mathbb{E} \left[\sup_{f \in F} \sum \varepsilon_i y_i f(x_i) \right] \\ &= \frac{1}{n} \mathbb{E} \left[\sup_{f \in F} \sum \varepsilon_i f(x_i) \right] \\ &= \mathcal{R}_n(F), \end{aligned}$$

ce qui donne le résultat attendu. La 2^e partie se montre de manière analogue en utilisant la 2^e inégalité du cours.

Ex 5)

Soit K un noyau t.q. $\forall x, x', |K(x, x')| \leq B^2$.

Soit F le RKHS associé à K . On pose $\inf_{f \in F} A(f) = \inf_f A(f)$ et $A(f) = \mathbb{E} \left[\max \{ 0, 1 - Y f(X) \} \right]$.

On considère l'estimateur suivant :

$$\hat{f}_n^\lambda = \underset{f \in F}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^n \max \{ 0, 1 - Y_i f(X_i) \} + \lambda \|f\|_F^2 \right\}$$

$\forall \lambda > 0$.

On note $\bar{A}_n^\lambda(f) = \frac{1}{n} \sum_{i=1}^n \max \{ 0, 1 - Y_i f(X_i) \} + \lambda \|f\|_F^2$.

1) $\bar{A}_n^\lambda(0_F) = \frac{1}{n} \sum_{i=1}^n \max \{ 0, 1 - 0 \} = 1$.

La fonction $\hat{f}_n^\lambda$ minimise $\bar{A}_n^\lambda$. Puisque $0_F \in F$, on en déduit que

$$\bar{A}_n^\lambda(\hat{f}_n^\lambda) \leq \bar{A}_n^\lambda(0_F) = 1.$$

Or, $\frac{1}{n} \sum_{i=1}^n \max \{ 0, 1 - Y_i f(X_i) \} \geq 0, \forall f \in F$. Ainsi :

$$\lambda \|\hat{f}_n^\lambda\|_F^2 \leq \bar{A}_n^\lambda(\hat{f}_n^\lambda) \leq 1 \Rightarrow \|\hat{f}_n^\lambda\|_F^2 \leq \frac{1}{\lambda}.$$

2) Posons $A(f) = E[Y(Y f(x))]$ avec $Y(x) = \max\{0, 1 - x\}$.



Posons $\lambda > 0$ et $\widehat{F}_\lambda = \underset{f \in F}{\operatorname{argmin}} \widehat{A}_n(f)$

On définit l'ensemble $\widetilde{F} = \{z = (x, y) \mapsto y f(x) \mid f \in \widehat{F}_\lambda\}$.

Avec probabilité $1 - \frac{\delta}{2}$,

$$\forall f \in \widehat{F}_\lambda, |A(f) - \widehat{A}_n(f)| \leq 2 \mathcal{R}_n(Y \circ \widetilde{F}) + c \sqrt{\frac{\ell_n(2\delta)}{2n}}$$

avec c tel que

$$c = |\ell(Y f(x)) - Y(Y' f(x))| = |\max\{0, 1 - Y f(x)\} \cdot \max\{0, 1 - Y'(x)\}|$$

$$\leq \max\{0, 1 - Y f(x)\} \quad \downarrow \leq \operatorname{sgn} Y' f \operatorname{sgn} f(x)$$

$$\leq 1 + |f(x)| = 1 + \langle f, k(x, \cdot) \rangle_F$$

Th 2.10 avec

$$\phi(z_1, \dots, z_n) = \sup_{f \in \widehat{F}} \{A(f) - \widehat{A}_n(f)\}$$

$$\leq 1 + \|f\|_F \|k(x, \cdot)\|_F$$

$$\leq 1 + \|f\|_F B$$

$$\leq 1 + \frac{B}{\lambda}$$

On sait que γ est τ -Lipschitz. On a donc

$$|A(f) - \hat{A}_n(f)| \leq 2R_n(\tilde{F}) + c \sqrt{\frac{\ln(\frac{2}{\delta})}{2n}}.$$

Aussi,

$$Q_n(\tilde{F}) = \mathbb{E} \left[\mathbb{E} \left[\sup_{f \in \tilde{F}_\lambda} \frac{1}{n} \sum_{i=1}^n \underbrace{\varepsilon_i}_{\sim \varepsilon_i} \gamma_i f(X_i) \mid D_n := (X_i, \gamma_i) \right] \right]$$

$$= \mathbb{E} \left[\mathbb{E} \left[\sup_{f \in \tilde{F}_\lambda} \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \mid D_n \right] \right]$$

$$\left. \begin{array}{l} K_{i,j} = k(X_i, X_j) \\ \text{Er}^2_{qL} \end{array} \right) = \mathbb{E} \left[\frac{\sup_f \|f\|_F \sqrt{\text{Tr}(K)}}{\sup_f \|f\|_F \sqrt{n Q^2}} \right] \quad \begin{array}{l} \gamma_i \in \{-1, 1\} \\ \text{(Mussart)} \end{array}$$

$$\leq \frac{\sup_f \|f\|_F Q}{\sqrt{n}}$$

$$\leq \frac{Q}{\sqrt{n\lambda}} \quad (q^{c-1})$$

Finalement, on obtient :

$$A(f) \leq \hat{A}_n(f) + \frac{2Q}{\sqrt{n\lambda}} + \left(1 + \frac{Q}{\sqrt{\lambda}}\right) \sqrt{\frac{\ln \frac{2}{\delta}}{2n}}, \quad \forall f \in \hat{\tilde{F}}^\lambda$$

3) On pose $(\lambda_n)_{n \in \mathbb{N}}$ tq $\lambda_n \rightarrow 0$ et $n\lambda_n \rightarrow \infty$.

D'après le développement effectué à la q^e précédente, on a,
avec prob $1 - \frac{\delta}{2}$,

$$-C_n \leq A(\hat{f}_n^{\lambda_n}) - \widehat{A}_n(\hat{f}_n^{\lambda_n}) \leq C_n,$$

$$\text{où } C_n = \frac{29}{\sqrt{n\lambda_n}} \cdot \left(1 + \frac{9}{\sqrt{\lambda_n}}\right) \sqrt{\frac{\ln(2/\delta)}{2n}}$$

Posons $f^* = \underset{f}{\operatorname{arg\,min}} A(f)$. On obtient, avec prob $1 - \frac{\delta}{2}$,

$$A(\hat{f}_n^{\lambda_n}) + \underbrace{\widehat{A}_n(f^*) - \widehat{A}_n(\hat{f}_n^{\lambda_n})}_{\geq 0} + \underbrace{A(f^*) - \widehat{A}_n(f^*)}_{-C_n \leq \cdot \leq C_n} - A(f^*) \leq C_n$$

$$\Rightarrow A(\hat{f}_n^{\lambda_n}) - A(f^*) \leq 2C_n.$$

Ainsi,

$$\mathbb{P}(A(\hat{f}_n^{\lambda_n}) - A(f^*) > 2C_n) \leq \frac{\delta}{2}.$$

Posons $\delta \triangleq \frac{2}{n^2}$, $2C_n \xrightarrow{n \rightarrow \infty} 0$. On en déduit que $\forall \varepsilon > 0$, $\exists N$,

$\forall n \geq N$, $2C_n < \varepsilon$. On obtient

$$S_N + \sum_{n=N}^{\infty} \mathbb{P}(A(\hat{f}_n^{\lambda_n}) - A(f^*) > \varepsilon) \leq S_N + \sum_{n=N}^{\infty} \frac{1}{n^2}$$

Où
$$S_N = \sum_{n=0}^{N-1} P(A(\hat{f}_n^{\lambda_n}) - A(f^*) > \varepsilon) < \infty.$$

De plus, la série $\sum \frac{1}{n^2}$ converge. On en conclut, $\forall \varepsilon > 0$,

$$\sum_{n=0}^{\infty} P(A(\hat{f}_n^{\lambda_n}) - A(f^*) > \varepsilon) < \infty. \text{ On utilise Borel-Cantelli pour conclure.}$$

On a

$$L(\text{sgn}(\hat{f}_n^{\lambda_n})) = P(\text{sgn}(\hat{f}_n^{\lambda_n}) \neq Y) = P(Y \hat{f}_n^{\lambda_n}(x) \leq 0)$$

$$= E[\mathbb{1}_{\{Y \hat{f}_n^{\lambda_n} \leq 0\}}]$$

- si $Y \hat{f}_n^{\lambda_n}(x) \leq 0$,

$$\varphi(Y \hat{f}_n^{\lambda_n}(x)) \geq 1 \geq \mathbb{1}_{\{Y \hat{f}_n^{\lambda_n} \leq 0\}}$$

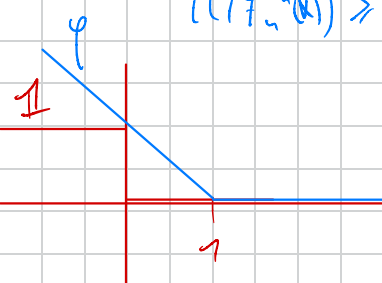
- si $Y \hat{f}_n^{\lambda_n}(x) \geq 1$,

$$\varphi(Y \hat{f}_n^{\lambda_n}(x)) = 0 = \mathbb{1}_{\{Y \hat{f}_n^{\lambda_n} \leq 0\}}$$

- si $0 \leq Y \hat{f}_n^{\lambda_n}(x) \leq 1$,

$$\varphi(Y \hat{f}_n^{\lambda_n}(x)) \geq 0 = \mathbb{1}_{\{Y \hat{f}_n^{\lambda_n} \leq 0\}}$$

$$\leq E[\varphi(Y \hat{f}_n^{\lambda_n})] \\ = A(\hat{f}_n^{\lambda_n})$$



h) On cherche à montrer que SVM est uniformément stable:

$$D_n = \left((x_1, y_1), \dots, (x_n, y_n) \right)$$

$$D'_n = \left((x'_1, y'_1), \dots, (x'_n, y'_n) \right)$$

$$\text{t.q. } \forall i \neq k \quad (x_i, y_i) = (x'_i, y'_i)$$

$$\hat{f}_n^{\lambda_n} \in \underset{f \in F}{\operatorname{argmin}} \hat{A}_n(f), \quad \hat{f}_n^{\lambda'_n} \in \underset{f \in F}{\operatorname{argmin}} \hat{A}'_n(f)$$

$$\forall f \in F^{\lambda_n, x, y}, \quad \varphi(y, f(x)) \leq \Lambda \triangleq 1 + \frac{B}{\sqrt{\lambda}}$$

$$\left| \varphi\left(\hat{f}_n^{\lambda_n}\right) - \varphi\left(\hat{f}_n^{\lambda'_n}\right) \right| \stackrel{(\varphi \text{ 2-Lipschitz})}{\leq} \left| \hat{f}_n^{\lambda_n} - \hat{f}_n^{\lambda'_n} \right|_{\infty} \leq B \left\| \hat{f}_n^{\lambda_n} - \hat{f}_n^{\lambda'_n} \right\|_F \leq \frac{B^2}{\lambda_n} \triangleq \gamma_n$$

$$\forall f \in F, \forall x, \| \phi(x, \cdot) \|_F = \left\| \langle k(x, \cdot), \cdot \rangle_F \right\|_F$$

$$\leq \| k(x, \cdot) \|_F \| f \|_F$$

$$= \sqrt{\langle k(x, \cdot), k(x, \cdot) \rangle} \| f \|_F$$

$$= \sqrt{\kappa(x, x)} \| f \|_F$$

$$\leq B \| f \|_F$$

$$\leq B \| f \|_F$$

$$\Rightarrow \| f \|_{\infty} \leq B \| f \|_F$$

(Proposition 11.1,

Foundations of ML
Mohri et al.)

On peut démontrer que, avec probabilité $1 - \delta$, ①

$$\left| A\left(\frac{\tilde{f}^{\lambda_n}}{f_n}\right) - \tilde{A}_n\left(\frac{\tilde{f}^{\lambda_n}}{f_n}\right) \right| \leq \gamma_n + (2n\gamma_n + \Delta) \sqrt{\frac{\log(2/\delta)}{2n}}.$$

Montrons que $\gamma_n \rightarrow 0$ et $\sqrt{n}\gamma_n \rightarrow 0$. (Supposons que $\sqrt{n}\lambda_n \rightarrow +\infty$)

$$\gamma_n = \frac{\beta^2}{\lambda_n} \rightarrow 0.$$

$$\sqrt{n}\gamma_n = \frac{\beta^2}{\lambda_n \sqrt{n}} \rightarrow 0.$$

Ainsi, $\gamma_n + (2n\gamma_n + \Delta) \sqrt{\frac{\log(2/\delta)}{2n}} \rightarrow 0$. Appliquer la

même stratégie qu'à la question précédente termine la preuve.

Preuve de ①.

On cherche à majorer

$$\begin{aligned} \textcircled{1} \quad |A(\bar{f}_n) - A(\bar{f}'_n)| &= \left| \mathbb{E}_{X,Y} [\varphi(\bar{f}_n)] - \mathbb{E}_{X,Y} [\varphi(\bar{f}'_n)] \right| \\ &= \left| \mathbb{E} [\varphi(\bar{f}_n) - \varphi(\bar{f}'_n)] \right| \\ &\leq \mathbb{E} [|\varphi(\bar{f}_n) - \varphi(\bar{f}'_n)|] \\ &\leq \gamma_n \end{aligned}$$

$$\begin{aligned} \textcircled{2} \quad |\hat{A}_n(\hat{f}) - \hat{A}_n(\hat{f}'_n)| &= \frac{1}{n} \left| \sum_{i=1}^n \varphi(y_i, \hat{f}(x_i)) - \varphi(y_i, \hat{f}'_n(x_i)) \right| \\ &\leq \frac{1}{n} \sum_{i \neq k} |\varphi(y_i, \hat{f}(x_i)) - \varphi(y_i, \hat{f}'_n(x_i))| \\ &\quad + \frac{1}{n} |\varphi(y_k, \hat{f}(x_k)) - \varphi(y_k, \hat{f}'_n(x_k))| \\ &\leq \frac{(n-1)\gamma_n}{n} + \frac{\Delta}{n} \leq \gamma_n + \frac{\Delta}{n} \end{aligned}$$

En utilisant ① et ②, on peut obtenir l'inégalité suivante

$$\begin{aligned}
 |A(\hat{f}_n) - \hat{A}_n(\hat{f}_n) - A(\hat{f}'_n) + \hat{A}_n(\hat{f}'_n)| &= |A(\hat{f}_n) - A(\hat{f}'_n) + \hat{A}_n(\hat{f}'_n) - \hat{A}_n(\hat{f}_n)| \\
 &\leq |A(\hat{f}_n) - A(\hat{f}'_n)| + |\hat{A}_n(\hat{f}'_n) - \hat{A}_n(\hat{f}_n)| \\
 &\leq 2\gamma_n + \frac{\Lambda}{n} \triangleq c.
 \end{aligned}$$

On peut appliquer l'inégalité de Mc Diarmid:

$$P(|A(\hat{f}_n) - \hat{A}_n(\hat{f}_n) - \mathbb{E}[A(\hat{f}_n) - \hat{A}_n(\hat{f}_n)]| \geq \varepsilon) \leq 2 \exp\left(\frac{-2\varepsilon^2}{nc^2}\right)$$

$$\begin{aligned}
 \text{On pose } \delta &\triangleq 2 \exp\left(\frac{-2\varepsilon^2}{nc^2}\right) = 2 \exp\left(\frac{-2\varepsilon^2}{n(2\gamma_n + \frac{\Lambda}{n})^2}\right) \\
 &= 2 \exp\left(\frac{-2n\varepsilon^2}{(2n\gamma_n + \Lambda)^2}\right)
 \end{aligned}$$

$$\Rightarrow \ln\left(\frac{\delta}{2}\right) = \frac{-2n\varepsilon^2}{(2n\gamma_n + \Lambda)^2}$$

$$\sqrt{\frac{\ln\left(\frac{2}{\delta}\right)}{2n}} (2n\gamma_n + \Lambda) = \varepsilon.$$

Ainsi,

$$P(|A(\hat{f}_n) - \hat{A}_n(\hat{f}_n) - \mathbb{E}[A(\hat{f}_n) - \hat{A}_n(\hat{f}_n)]| \leq (2n\gamma_n + \Lambda) \sqrt{\frac{\ln\left(\frac{2}{\delta}\right)}{2n}}) \geq 1 - \delta.$$

$\mathcal{O}_1$,

$$\begin{aligned}\mathbb{E}_{\mathcal{D}_n} [A(\hat{f}_n) - \hat{A}_n(\hat{f}_n)] &= \mathbb{E}_{\mathcal{D}_n} \left[\frac{1}{n} \sum_{i=1}^n \varphi(y_i, \hat{f}_n(x_i)) - \mathbb{E}_z [\varphi(y, \hat{f}_n(x))] \right] \\&= \mathbb{E}_{\mathcal{D}_n, z} \left[\frac{1}{n} \sum_{i=1}^n \varphi(y_i, \hat{f}_n(x_i)) - \varphi(y, \hat{f}_n(x)) \right] \\&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mathcal{D}_n, z} [\varphi(y_i, \hat{f}_n(x_i)) - \varphi(y, \hat{f}_n(x))] \\&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mathcal{D}_n, z} [\varphi(y, \hat{f}'_n(x)) - \varphi(y, \hat{f}_n(x))] \\&\leq \gamma_n.\end{aligned}$$

(Où $\hat{f}'_n(x)$ est apprise sur l'ensemble des points de $\mathcal{D}_n$ à l'exception de $z_i = z$. La "symétrie" de l'espérance permet de mélanger les termes).

Finalement, avec probabilité $1 - \delta$,

$$|A(\hat{f}_n) - \hat{A}_n(\hat{f}_n)| \leq \gamma_n + (2\gamma_n + \Delta) \sqrt{\frac{\log \frac{2}{\delta}}{2n}}.$$