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

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

$$R_n(F) = \mathbb{E}_{D_n}\left[\hat{R}_n(F)\right]$$

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

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

Ex 1)

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

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

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

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

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

(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).

$$= \mathbb{E}\left[\sup_{f \in \mathcal{T}} \frac{1}{n} |c| \, \varepsilon_i \, f(z_i) \,\middle|\, D_n \right]$$

$$= |c| \, \hat{R}_n(\mathcal{T})$$

$\sup(|c|A) = \sup\{|c| \cdot x \in A\}$
$= |c| \sup(A).$

2) On suppose $T_1 \subseteq T_2$.

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 $\underbrace{\sup(T_1)}_{t_1} > \underbrace{\sup(T_2)}_{t_2}$.

Par def du sup, $\forall x \in T_2$, $x \leq t_2$.

Or, $T_1 \subset T_2$, donc $\forall x \in T_1 \Rightarrow x \in T_2$ donc $x \leq t_2$.

Par def du sup, $t_1 \leq t_2 \Rightarrow$ Contradiction.

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

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

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

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

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

Mq $\sup(A+B) = \sup(A) + \sup(B)$

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

Par def, $\quad \sup(A) \leq \sup(A+B) - b$

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

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

Def sup, $\forall s, \left(\forall a,b,\ a+b \leq s \right), \sup(A+B) \leq s$.

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

et $\quad // \quad = \quad //$

$$\hat{R}_n(T_1+T_2) = \mathbb{E}\left[\sup_{g \in T_1+T_2} \frac{1}{n} \sum \varepsilon_i\, g(z_i) \mid D_n \right]$$

$$= \mathbb{E}\left[\sup_{\substack{f_1 \in T_1 \\ f_2 \in T_2}} \frac{1}{n} \sum \varepsilon_i \left(f_1(z_i) + f_2(z_i) \right) \mid D_n \right]$$

$$= \mathbb{E}\left[\sup_{\substack{f_1 \in T_1 \\ f_2 \in T_2}} \frac{1}{n} \sum \varepsilon_i f_1(z_i) + \varepsilon_i f_2(z_i) \mid D_n \right]$$

$$= \hat{R}_n(T_1) + \hat{R}_n(T_2)$$

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

$$\underset{\substack{\lambda_i \geq 0 \\ \|\lambda\| = 1}}{\arg\max} \sum \lambda_i a_i = \begin{cases} 1 & \text{si } i = \arg\max (a_j) \\ 0 & \text{sinon} \end{cases}$$

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

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

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

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

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

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

5) $\phi : \mathbb{R} \to \mathbb{R}, \quad |\phi(x_1) - \phi(x_2)| \leq k |x_1 - x_2|$

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

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

$$= \frac{1}{n} \mathbb{E}\left[\sup_{f \in \mathcal{T}} u_{n-1}(f) + \varepsilon_n (\phi \circ f)(z_n) \mid D_n \right].$$

$$= \frac{1}{n} \mathbb{E}_{\varepsilon_1, \dots, \varepsilon_{n-1}}\left[\mathbb{E}_{\varepsilon_n}\left[u_{n-1}(f) + \varepsilon_n (\phi \circ f)(z_n) \mid z_n \right] \mid D_{n-1} \right] \quad (1)$$

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

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

$$u_{n-1}(f_1) + (\phi \circ f_1)(z_n) \geq (1-\varepsilon)\left(\sup_{f \in \mathcal{T}} u_{n-1}(f) + (\phi \circ f)(z_n) \right)$$

$$\text{et} \quad u_{n-1}(f_2) - (\phi \circ f_2)(z_n) \geq (1-\varepsilon)\left(\sup_{f \in \mathcal{T}} u_{n-1}(f) - (\phi \circ f)(z_n) \right)$$

Ainsi,

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

$$\leq \frac{1}{2}\left(u_{n-1}(f_1) + (\phi \circ f_1)(z_n) \right) + \frac{1}{2}\left(u_{n-1}(f_2) - (\phi \circ f_2)(z_n) \right)$$

$$= \frac{1}{2}\left(u_{n-1}(f_1) + u_{n-1}(f_2) + (\phi \circ f_1)(z_n) - (\phi \circ f_2)(z_n) \right)$$

Posons $s = \mathrm{sgn}\left(f_1(z_n) - f_2(z_n)\right)$. Par propriété de Lipschitz, on a

$$\phi(f_1(z_n)) - \phi(f_2(z_n)) \leq |\phi(f_1(z_n)) - \phi(f_2(z_n))| \leq \kappa s \left(f_1(z_n) - f_2(z_n)\right).$$

L'inégalité précédente implique donc :

$$(1-\varepsilon) \mathop{\mathbb{E}}_{\varepsilon_n}\left[\sup_{f \in \mathcal{F}} u_{n-1}(f) + \varepsilon_n (\phi \circ f)(z_n)\right] \leq \frac{1}{2}\left(u_{n-1}(f_1) + u_{n-1}(f_2) + s\kappa\left(f_1(z_n) - f_2(z_n)\right)\right)$$

$$= \frac{1}{2}\left(u_{n-1}(f_1) + s\kappa f_1(z_n)\right) + \frac{1}{2}\left(u_{n-1}(f_2) - s\kappa f_2(z_n)\right)$$

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

$$= \mathop{\mathbb{E}}_{\varepsilon_n}\left[\sup_{f \in \mathcal{F}} u_{n-1}(f) + \underbrace{\varepsilon_n s}_{\sim \varepsilon_n} \kappa f(z_n)\right]$$

$$= \mathop{\mathbb{E}}_{\varepsilon_n}\left[\sup_{f \in \mathcal{F}} u_{n-1}(f) + \varepsilon_n \kappa f(z_n)\right]$$

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

$$\mathop{\mathbb{E}}_{\varepsilon_n}\left[\sup_{f \in \mathcal{F}} u_n(f) + \varepsilon_n (\phi \circ f)(z_n)\right] \leq \mathop{\mathbb{E}}_{\varepsilon_n}\left[\sup_{f \in \mathcal{F}} u_{n-1}(f) + \varepsilon_n \kappa f(z_n)\right].$$

On en déduit l'inégalité suivante,

$$\mathbb{E}_{\varepsilon_1 \dots \varepsilon_{n-1}}\left[\mathbb{E}_{\varepsilon_n}\left[\sup_{f \in \mathcal{T}} u_{n-1}(f) + \varepsilon_n (\phi \circ f)(z_n) \,\middle|\, z_n \right] \,\middle|\, D_{n-1} \right] \quad (1)$$

$$\leqslant \mathbb{E}_{\varepsilon_1 \dots \varepsilon_{n-1}}\left[\mathbb{E}_{\varepsilon_n}\left[\sup_{f \in \mathcal{T}} u_{n-1}(f) + \varepsilon_n K f(z_n) \,\middle|\, z_n \right] \,\middle|\, D_{n-1} \right]$$

$$= \mathbb{E}\left[\sup_{f \in \mathcal{T}} u_{n-1}(f) + \varepsilon_n K f(z_n) \,\middle|\, D_n \right].$$

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

Ex 2)

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

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

$$= \mathbb{E}\left[\sup_{\substack{w \in \mathbb{R}^d \\ \|w\|_1 \leqslant B}} \frac{1}{n} \sum \varepsilon_i w^T z_i \,\middle|\, D_n \right]$$

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

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

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

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

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

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

$$= \frac{B}{n} \mathbb{E}\left[\sup_{x \in C} \left| \sum \varepsilon_i x_i \right| \,\middle|\, D_n \right], \quad C = \left\{ \begin{pmatrix} z_1^{(1)} \\ \vdots \\ z_n^{(1)} \end{pmatrix}, \dots, \begin{pmatrix} z_1^{(d)} \\ \vdots \\ z_n^{(d)} \end{pmatrix} \right\}$$

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

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

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

$$\max_{x \in C} \|x\|_2 = \max_j \left(\sum z_i^{2(j)} \right)^{\frac{1}{2}} \leq \left(\sum c^2 \right)^{\frac{1}{2}} = (n c^2)^{\frac{1}{2}} = \sqrt{n}\, c$$

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

$K = \left(k(X_i, X_j) \right)_{i \leq n, j \leq n}$

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

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

$$= \frac{1}{n} \mathbb{E}\left[\sup_{\|w\|_k \leq M} \langle w, \sum \varepsilon_i \phi(X_i) \rangle \mid D_n \right]$$

$$\leq \frac{1}{n} \mathbb{E}\left[\sup_{\|w\|_k \leq M} \|w\|_k \| \sum \varepsilon_i \phi(X_i) \|_k \mid D_n \right]$$

$$\leq \frac{M}{n} \mathbb{E}\left[\| \sum \varepsilon_i \phi(X_i) \|_k \mid D_n \right]$$

$$= \frac{M}{n} \mathbb{E}\left[\left(\langle \sum \varepsilon_i \phi(X_i), \sum \varepsilon_i \phi(X_i) \rangle \right)^{1/2} \mid D_n \right]$$

Jensen

φ concave: $\mathbb{E}[\varphi(X)] \leq \varphi(\mathbb{E}[X])$

φ convexe: $\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X])$

$$\leq \frac{M}{n} \mathbb{E}\left[\langle \sum \varepsilon_i \phi(X_i), \sum \varepsilon_i \phi(X_i) \rangle \mid D_n \right]^{1/2}$$

$$= \frac{M}{n} \mathbb{E}\left[\sum_{i,j} \langle \varepsilon_i \phi(X_i), \varepsilon_j \phi(X_j) \rangle \mid D_n \right]^{1/2}$$

$$= \frac{M}{n} \mathbb{E}\left[\sum_i \varepsilon_i \sum_j \varepsilon_j \langle \phi(X_i), \phi(X_j) \rangle \mid D_n \right]^{1/2}$$

$$= \frac{M}{n} \mathbb{E}\left[\sum_{i \neq j} \varepsilon_i \varepsilon_j \langle \phi(X_i), \phi(X_j) \rangle + \sum_i \|\phi(X_i)\|_k^2 \right]^{1/2}$$

$$= \frac{M}{n} \left[\sum \mathbb{E}[\varepsilon_i \varepsilon_j] \langle \phi(X_i), \phi(X_j) \rangle + \sum_i \|\phi(X_i)\|_k^2 \right]^{1/2}$$

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

$$= \frac{M}{n}\left(\sum_i \|\phi(X_i)\|_K^2\right)^{1/2}$$

$$= \frac{M\sqrt{\operatorname{trace}(K)}}{n}.$$

Ex 3)

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

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

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

Posons $\mathcal{G}_{|D_n}$ l'ensemble des vecteurs $(g(Z_1), \dots, g(Z_n))^T$ pour toute fonction dans $\mathcal{G}$: $\mathcal{G}_{|D_n} = \{(g(Z_i))_{i\leq n}^T \mid g\in\mathcal{G}\}$.

On obtient

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

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

$$\mathbb{E}_{D_n}\left[\frac{\sqrt{n}\sqrt{2 \ln |\mathcal{G}_{|D_n}|}}{n}\right] \leq \frac{\sqrt{n}\sqrt{2 \ln \gamma(\mathcal{G}, n)}}{n}$$

$$= \left(\frac{2 \ln \gamma(\mathcal{G}, n)}{n}\right)^{1/2}.$$

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

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

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

Si $n > V(\mathcal{G})$:

$$\gamma(\mathcal{G}, n) \leq \left(\frac{en}{V(\mathcal{G})}\right)^{V(\mathcal{G})}.$$ Posons $V = V(\mathcal{G})$

On en déduit

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

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

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

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

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

$$= \left(\frac{V\ln n}{n}\right)^{1/2}\underbrace{\left(\frac{2\left(1 + \frac{\ln V}{\ln n} + \frac{1}{\ln n}\right)}{n}\right)^{1/2}}_{=C}$$

$\left(\log_2 n = \frac{\ln n}{\ln 2} \geq \ln n \quad \forall n \geq 1\right)$ $\leq \left(\frac{V\log_2 n}{n}\right)^{1/2} C$

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

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

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

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

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

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

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

Posons $\lambda > 0$ et $\widehat{F_\lambda} = \underset{f \in F}{\arg\min}\ \hat{A}_n(f)$

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

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

$$\forall f \in \widehat{F_\lambda},\ |A(f) - \hat{A}_n(f)| \leq 2 R_n(\varphi \circ \tilde{F}) + c\sqrt{\frac{\ln(2/\delta)}{2n}}$$

avec c tel que

$$c = |\varphi(y f(x)) - \varphi(y' f(x'))| = |\max\{0, 1 - y f(x)\} - \max\{0, 1 - y' f(x')\}|$$

$$\leq \max\{0, 1 - y f(x)\}$$

$\Updownarrow$ Si $\operatorname{sgn} Y \neq \operatorname{sgn} f(X)$

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

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

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

$$\leq 1 + \frac{D}{\sqrt{\lambda}}$$

Th 2.10 avec

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

On sait que φ est 1-Lipschitz. On a donc

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

Aussi,

$$R_n(\tilde{F}) = \mathbb{E}\left[\mathbb{E}\left[\sup_{f\in\hat{F}_\lambda} \frac{1}{n}\sum \underbrace{\varepsilon_i Y_i}_{\sim\varepsilon_i} f(X_i) \;\middle|\; D_n := (X_i, Y_i)\right]\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\sup_{f\in\hat{F}_\lambda} \frac{1}{n}\sum \varepsilon_i f(X_i) \;\middle|\; D_n\right]\right]$$

$Y_i \in \{-1, 1\}$

$\left. \begin{array}{l} K_{i,j} = k(X_i, X_j) \\ E_y 2 \, q \, L \end{array} \right)$

$$= \mathbb{E}\left[\frac{\sup_A \|f\|_F \sqrt{T_r(K)}}{n}\right]$$

$$\leq \frac{\sup\|f\|_F \sqrt{n B^2}}{n} = \frac{\sup\|f\|_F \, B}{\sqrt{n}}$$

$$\leq \frac{B}{\sqrt{n\lambda}} .$$

$(q^{\circ} 1)$

Finalement, on obtient :

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

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

D'après le développement effectué à la q° précédente, on a, avec prob $1-\delta$:

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

où $$C_n = \frac{2B}{\sqrt{n\lambda}} + \left(1 + \frac{B}{\sqrt{\lambda}}\right)\sqrt{\frac{\ln(2/\delta)}{2n}}$$

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

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

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

Ainsi,

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

Posons $\delta \triangleq \frac{1}{n^2}$, $2C_n \xrightarrow[n\to\infty]{} 0$. On en déduit que $\forall \varepsilon > 0, \exists N,$
$\forall n \geq N, 2C_n < \varepsilon$. On obtient

$$R_N + \sum_{n=N}^{\infty} \mathbb{P}\left(A(\hat{f}_n^{\lambda}) - A(f^*) > \varepsilon\right) \leq R_N + \sum_{n=N}^{\infty} \frac{1}{n^2},$$

Où $\quad R_N = \sum_{n=0}^{N-n} \mathbb{P}\left(A(\hat{f}_n^{\lambda_n}) - A(f^*) > \varepsilon \right) < \infty.$

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

$$\sum_{n=0}^{\infty} \mathbb{P}\left(A(\hat{f}_n^{\lambda_n}) - A(f^*) > \varepsilon \right) < \infty.$$

On a

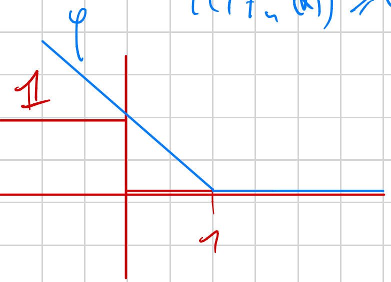
$$L(\mathrm{sgn}(\hat{f}_n^{\lambda_n})) = \mathbb{P}\left(\mathrm{sgn}(\hat{f}_n^{\lambda_n}(X)) \neq Y \right) = \mathbb{P}\left(Y \hat{f}_n^{\lambda_n}(X) \leq 0 \right)$$

$$= \mathbb{E}\left[\mathbb{1}_{\{Y \hat{f}_n^{\lambda_n}(X) \leq 0\}} \right]$$

$$\leq \mathbb{E}\left[\varphi\left(Y \hat{f}_n^{\lambda_n}(X) \right) \right]$$

$$= A(\hat{f}_n^{\lambda_n})$$

- si $Y \hat{f}^{\lambda_n}(X) \leq 0$,
  $\varphi(Y \hat{f}_n^{\lambda_n}(X)) \geq 1 \geq \mathbb{1}_{\{\dots\}}$
- si $Y \hat{f}_n^{\lambda_n}(X) \geq 1$,
  $\varphi(Y \hat{f}_n^{\lambda_n}(X)) = 0 = \mathbb{1}_{\{\dots\}}$
- si $0 \leq Y \hat{f}_n^{\lambda_n}(X) \leq 1$,
  $\varphi(Y \hat{f}_n^{\lambda_n}(X)) \geq 0 = \mathbb{1}_{\{\dots\}}$



4) On cherche à montrer que SVM est uniformément stable.

$$D_n = \left((X_1, Y_1), \dots, (X_n, Y_n) \right)$$

$$D_n' = \left((X_1', Y_1'), \dots, (X_n', Y_n') \right) \quad \text{tq. } \forall i \neq k \ (X_i, Y_i) = (X_i', Y_i')$$

$$\hat{f}_n^{\lambda_n} \in \underset{f \in F}{\arg\min} \ \hat{A}_n(f), \quad \hat{f}_n^{\lambda_n'} \in \underset{f \in F}{\arg\min} \ \hat{A}_n'(f)$$

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

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

(φ 1-Lipschitz) (Tikhonov regularization)

$$\forall f \in F, \ \forall x, \ |f(x)| = |\langle k(x, \cdot), f \rangle_F|$$
$$\leq \|k(x, \cdot)\|_F \|f\|_F$$
$$= \sqrt{\langle k(x, \cdot), k(x, \cdot) \rangle} \|f\|_F$$
$$= \sqrt{k(x, x)} \|f\|_F$$
$$\leq B \|f\|_F$$
$$\leq B \|f\|_F$$
$$\Rightarrow \|f\|_\infty \leq B \|f\|_F$$

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

$$\left| A\left(\hat{f}_n^{\lambda_n}\right) - \hat{A}_n\left(\hat{f}_n^{\lambda_n}\right) \right| \leq \gamma_n + (2n\gamma_n + 1)\sqrt{\frac{\log(2/\delta)}{2n}}.$$

Montrons que $\gamma_n \longrightarrow 0$ et $\sqrt{n}\,\gamma_n \longrightarrow 0$: (supposons que $\sqrt{n}\,\lambda_n \longrightarrow 0$)

$$\gamma_n = \frac{B^2}{\lambda n} \longrightarrow 0.$$

$$\sqrt{n}\,\gamma_n = \frac{B^2}{\lambda \sqrt{n}} \longrightarrow 0.$$

Ainsi, $\gamma_n + (2n\gamma_n + 1)\sqrt{\frac{\log(2/\delta)}{2n}} \longrightarrow 0$. Appliquer la même stratégie qu'à la question précédente termine la preuve.

Preuve de ①:

On cherche à majorer

① $|A(\hat{f}_n) - A(\hat{f}_n')| = \left| \mathbb{E}_{X,Y}[\varphi(\hat{f}_n)] - \mathbb{E}_{X,Y}[\varphi(\hat{f}_n')] \right|$.

$$= \left| \mathbb{E}[\varphi(\hat{f}_n) - \varphi(\hat{f}_n')] \right|$$

$$\leq \mathbb{E}[|\varphi(\hat{f}_n) - \varphi(\hat{f}_n')|]$$

$$\leq \gamma_n$$

② $|\hat{A}_n(\hat{f}_n) - \hat{A}_n(\hat{f}_n')| = \frac{1}{n} \left| \sum_{i=1}^{n} \varphi(Y_i, \hat{f}_n(X_i)) - \varphi(Y_i', \hat{f}_n'(X_i')) \right|$

$$\leq \frac{1}{n} \sum_{i \neq k} |\varphi(Y_i, \hat{f}_n(X_i)) - \varphi(Y_i, \hat{f}_n'(X_i))|$$

$$+ \frac{1}{n} |\varphi(Y_k', \hat{f}_n(X_k)) - \varphi(Y_k', \hat{f}_n'(X_k'))|$$

$$\leq \frac{(n-1)\gamma_n}{n} + \frac{\Delta}{n} \leq \gamma_n + \frac{\Delta}{n}$$

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

$$|A(\hat{f}_n) - \hat{A}_n(\hat{f}_n) - A(\hat{f}'_n) + \hat{A}_n(\hat{f}'_n)| = |A(\hat{f}_n) - A(\hat{f}'_n) + \hat{A}_n(\hat{f}'_n) - \hat{A}_n(\hat{f}_n)|$$

$$\leq |A(\hat{f}_n) - A(\hat{f}'_n)| + |\hat{A}_n(\hat{f}'_n) - \hat{A}(\hat{f}_n)|$$

$$\leq 2\gamma_n + \frac{\Delta}{n} \triangleq c.$$

On peut appliquer l'inégalité de Mc Diarmid :

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

On pose $\delta \triangleq 2\exp\left(\frac{-2\varepsilon^2}{nc^2}\right) = 2\exp\left(\frac{-2\varepsilon^2}{n\left(2\gamma_n + \frac{\Delta}{n}\right)^2}\right)$

$$= 2\exp\left(\frac{-2n\varepsilon^2}{(2n\gamma_n + \Delta)^2}\right)$$

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

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

Ainsi,

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

Or,

$$\mathbb{E}_{D_n}\left[\left|A(\hat{f}_n) - \hat{A}_n(\hat{f}_n)\right|\right] = \mathbb{E}_{D_n}\left[\frac{1}{n}\sum_{i=1}^{n} \varphi(y_i, \hat{f}_n(x_i)) - \mathbb{E}_{z}\left[\varphi(y, \hat{f}_n(x))\right]\right]$$

$$= \mathbb{E}_{D_n, z}\left[\frac{1}{n}\sum_{i=1}^{n} \varphi(y_i, \hat{f}_n(x_i)) - \varphi(y, \hat{f}_n(x))\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{D_n, z}\left[\varphi(y_i, \hat{f}_n(x_i)) - \varphi(y, \hat{f}_n(x))\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{D_n, z}\left[\varphi(y, \hat{f}'_n(x)) - \varphi(y, \hat{f}_n(x))\right]$$

$$\leq \gamma_n .$$

(Où $f'_n(x)$ est apprise sur l'ensemble des points de D_n à l'exception de $z_i = z$. La "symmétrie" de l'espérance permet de mélanger les termes).

Finalement, avec probabilité $1-\delta$,

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