

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

$$\widehat{R}_n(F) = \mathbb{E} \left[ \sup_{f \in F} \left\{ \frac{1}{n} \sum_{i=1}^n \epsilon_i f(Z_i) \right\} \mid D_n \right]$$

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

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

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

$\mathbb{E}[x]$

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

$$\widehat{R}_n(c\gamma) = \mathbb{E} \left[ \sup_{g \in \mathcal{G}} \frac{1}{n} \sum \epsilon_i g(Z_i) \mid D_n \right]$$

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

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

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

( $\epsilon_i$  prend -1 et 1 avec la même proba.)

Mettre un ( $f$ ) devant  $\epsilon_i$  change l'ordre

dans lequel on va sommer les termes

pour calculer l'espérance mais pas la valeur finale).

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

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

$$\sup_{f \in \mathcal{T}} (c_f A) = \sup \{ (c_f - x) c_A \}$$

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

2) On suppose  $T_1 \subseteq T_2$ .

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

S:  $T_1 = T_2 \Rightarrow$  direct

S:  $T_1 \subset T_2$ . Supposons  $\sup_{\mathcal{T}_1} \geq \sup_{\mathcal{T}_2}$ .

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.

$$T_1 \subseteq T_2 \Rightarrow \left\{ \frac{1}{n} \sum_{i=1}^n |f(z_i) - f_i| \mid f \in T_1 \right\} \subseteq \left\{ \frac{1}{n} \sum_{i=1}^n |f(z_i) - f_i| \mid f \in T_2 \right\}.$$

$$\Rightarrow \sup \{ \| \cdot \| \} \leq \sup \{ \| \cdot \| \}$$

$$\text{Ainsi, } \widehat{R}_n(T_1) = \mathbb{E} \left[ \sup_{f \in T_1} \left\{ \frac{1}{n} \sum_{i=1}^n |f(z_i) - f_i| \right\} \mid D_n \right]$$

$$\left( \mathbb{E} \left[ \sup_{f \in T_2} \left\{ \frac{1}{n} \sum_{i=1}^n |f(z_i) - f_i| \right\} \mid D_n \right] \right) = \widehat{R}_n(T_2)$$

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

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

$$\forall a, b, a \leq \sup(A + \mathcal{B}) - b$$

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

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

$$\sup(A) + \sup(\mathcal{B}) \leq \sup(A + \mathcal{B})$$

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

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

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

$$\hat{Q}_{n_1}(T_1 + T_2) = \mathbb{E} \left[ \sup_{g_i \in T_1 + T_2} \sum_n \mathbb{E}_i g_i(z_i) \mid D_n \right]$$

$$= \mathbb{E} \left[ \sup_{\substack{f_1 \in T_1 \\ f_2 \in T_2}} \sum_n \mathbb{E}_i (f_1(z_i) + f_2(z_i)) \mid D_n \right]$$

$$= \mathbb{E} \left[ \sup_{\substack{f_1 \in T_1 \\ f_2 \in T_2}} \sum_n (\mathbb{E}_i f_1(z_i) + \mathbb{E}_i f_2(z_i)) \mid D_n \right]$$

$$= \hat{Q}_{n_1}(T_1) + \hat{Q}_{n_2}(T_2)$$

$$h) \text{Conv}(A) = \left\{ x \mid x = \sum_{i=1}^p \lambda_i a_i, a_i \in A, \lambda_i \geq 0, \|\lambda\|_1 = 1 \right\}.$$

$$\arg \max_{\substack{\lambda: \sum_i \lambda_i a_i \\ \|\lambda\|_1 = 1}} \sum_i \lambda_i a_i = \begin{cases} 1 & \text{si } i = \arg \max_i (\alpha_i) \\ 0 & \text{sinon} \end{cases}$$

$$\hat{Q}_n(\text{Conv}(T)) = \mathbb{E} \left[ \sup_{p \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^n E: \sum_{j=1}^p \lambda_j f_j(z_i) \mid D_n \right]$$

$(f_p) \in T$

$\lambda_j \geq 0$

$\|\lambda\|_1 = 1$

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

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

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

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

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

On cherche p fonctions dans

T mais on ne garde que celle qui maximise la somme des  $E_i$ .

Ex 2]

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

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

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

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

$$= \mathbb{E} \left[ \sup_{w \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n w^{(i)} \left[ \sum_{i=1}^n \epsilon_i z_i^{(i)} \mid D_n \right] \right]$$

$\|w\|_1 \leq B$

$$= \mathbb{E} \left[ \sup_{w \in \mathbb{R}^d} \frac{1}{n} \omega^T \left( \sum_{i=1}^n \epsilon_i z_i \right) \mid D_n \right]$$

$\|w\|_1 \leq B$

$$\leq \mathbb{E} \left[ \sup_{w \in \mathbb{R}^d} \frac{1}{n} \|w\|_1 \left\| \sum_{i=1}^n \epsilon_i z_i \right\|_\infty \mid D_n \right] \quad \begin{matrix} \text{(Holder)} \\ |(a,b)| \leq (\|a\|_1 \|b\|_\infty) \end{matrix}$$

$\|w\|_1 \leq B$

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

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

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

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

Masurk

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

$$\max_{x \in C} \|x\|_2 = \max_{z \in \mathbb{R}^d} \left( \sum z_i^2 \right)^{\frac{1}{2}}$$

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

$$\leq \left( \sum c_i^2 \right)^{\frac{1}{2}}$$

$$= (n C^2)^{\frac{1}{2}}$$

$$= \sqrt{n} C$$

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

$$\kappa = (\kappa(x_i, x_j))_{i, j \in \mathbb{N}}$$

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

$$\hat{\mathbb{P}}_n(\mathcal{H}) = \mathbb{E} \left[ \sup_{\|w\|_\infty \leq M} \left\| \sum \varepsilon_i \langle w, \phi(x_i) \rangle \right\|_K \mid D_n \right]$$

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

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

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

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

$$\stackrel{\text{Jensen}}{\leq} \frac{M}{n} \mathbb{E} \left[ \left\langle \sum \varepsilon_i \phi(x_i), \sum \varepsilon_i \phi(x_i) \right\rangle_K^{\frac{1}{2}} \mid D_n \right]$$

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

$\Psi$  convex:  
 $\mathbb{E}[\Psi(X)] \geq \Psi(\mathbb{E}[X])$

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

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

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

$$\begin{aligned}
 &= \frac{M}{n} \left( \sum_{i \neq j} \langle \phi(x_i), \phi(x_j) \rangle_k \mathbb{E}[e_i] \mathbb{E}[e_j] + \sum_i \| \phi(x_i) \|_k^2 \right)^{\frac{1}{2}} \\
 &= \frac{M}{n} \left( \sum_i \| \phi(x_i) \|_k^2 \right)^{\frac{1}{2}} \\
 &= \frac{\sqrt{M \text{trace}(k)}}{n}
 \end{aligned}$$

Ex 3

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

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

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

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

$\mathcal{O}_n$  obtient

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

Comme  $y \in \mathcal{G}$  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 :

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

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,

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

On en déduit

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

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

S:  $n \leq V(\mathcal{G}) \Rightarrow \mathcal{J}(g, n) = 2^n$ . Ainsi

$$R_n(g) \leq \left( \frac{2 \ln (2^n)}{n} \right)^{1/2} \leq \left( \frac{2 \log 2^n}{n} \right)^{1/2} = \sqrt{2}$$

Si  $n > V(G)$  :

$$\chi(G_{\lceil n \rceil}) \leq \left( \frac{e_n}{V(G)} \right)^{V(G)}. \text{ Posons } V = V(G)$$

On en déduit

$$\overline{\chi}_{\lceil n \rceil}(G) \leq \left( \frac{2 \ln \left( \frac{e_n}{V} \right)^V}{n} \right)^{\frac{1}{2}}$$

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

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

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

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

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

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

Ex h

$$1) \quad \phi: \mathbb{R} \rightarrow \mathbb{R}, \quad |\phi(x_1) - \phi(x_2)| \leq K |x_1 - x_2|$$

$$\begin{aligned} \widehat{R}_n(\phi \circ \tau) &= \mathbb{E} \left[ \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \varepsilon_i (\phi \circ f)(z_i) \mid D_n \right] \\ &= \mathbb{E} \left[ \frac{1}{n} \sup_{f \in \mathcal{F}} \sum_{i=1}^{n-1} \varepsilon_i (\phi \circ f)(z_i) + \varepsilon_n (\phi \circ f)(z_n) \mid D_n \right] \\ &= \frac{1}{n} \mathbb{E} \left[ \sup_{f \in \mathcal{F}} u_{n-1}(f) + \varepsilon_n (\phi \circ f)(z_n) \mid D_n \right]. \\ &= \frac{1}{n} \mathbb{E} \left[ \mathbb{E}_{\varepsilon_n} [u_{n-1}(f) + \varepsilon_n (\phi \circ f)(z_n) \mid z_n] \mid D_{n-1} \right] \quad (1) \end{aligned}$$

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  $\geq 0$ )

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

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

Ainsi,

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

$$\begin{aligned} &\quad (1-\varepsilon) \\ &\quad + \frac{1}{2} \sup_{f \in \mathcal{F}} u_{n-1}(f) - (\phi \circ f)(z_n) \end{aligned}$$

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

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

Posons  $s = \text{Sign}(f_1(z_n) - f_2(z_n))$ . Par propriété de Lipschitz, on a

$$|\phi(f_1(z_n)) - \phi(f_2(z_n))| \leq |\phi(f_1(z_n)) - \phi(f_1(z_m))| + |\phi(f_1(z_m)) - \phi(f_2(z_m))| \leq \kappa s (f_1(z_n) - f_2(z_n)).$$

L'inégalité précédente implique donc :

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

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

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

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

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

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

$$\mathbb{E} \left[ \sup_{f \in T} u_n(f) + \varepsilon_n (\phi \circ f)(z_n) \right] \leq \mathbb{E} \left[ \sup_{f \in T} u_{n-n}(f) + \varepsilon_n s\kappa f(z_n) \right].$$

On en déduit l'inégalité suivante.

$$\begin{aligned} & \mathbb{E} \left[ \mathbb{E} \left[ \sup_{f \in F} u_{n,n}(f) + \mathbb{E}_n (\phi \circ f)(z_n) \mid D_{n-1} \right] \right] \quad (1) \\ & \leq \mathbb{E} \left[ \mathbb{E}_{\tilde{\mathcal{E}}_n} \mathbb{E} \left[ \sup_{f \in F} u_{n,n}(f) + \mathbb{E}_n (\phi \circ f)(z_n) \mid z_n \right] \mid D_{n-1} \right] \\ & = \mathbb{E} \left[ \sup_{f \in F} u_{n,n}(f) + \mathbb{E}_n (\phi \circ f)(z_n) \mid D_n \right]. \end{aligned}$$

Procéder de la même manière pour tous les autres  $(\mathcal{E}_i)_{i \geq n}$  termine la preuve.

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

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

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

$$\begin{aligned} \forall f \in \tilde{F}, \quad & \mathbb{E}[f(z)] \leq \sum_{i=1}^n f(z_i) + R_n(\tilde{F}) + \left( \frac{\ln^{1/\delta}}{2n} \right)^{\frac{1}{\delta}} \\ \Leftrightarrow \quad & \end{aligned}$$

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

Or,  $\forall n \in \mathbb{N}$ ,  $\exists f_{n,0} \in \Phi_p(n)$ . Ainsi

$\mathbb{E}[\mathbb{I}_{\{Y f_{n,0} \leq 0\}}] \leq \mathbb{E}[\Phi_p(Y f_{n,0})]$  et donc

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

Dès plus,  $\Phi'_p(x) = \begin{cases} 0 & \text{si } p \leq x \\ -1/p & \text{si } 0 \leq x \leq p \\ 0 & \text{si } x \leq 0 \end{cases}$ .

$\Phi_p(x)$  est donc  $\frac{1}{p}$ -Lipschitz.

On obtient ainsi :

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

$$\begin{aligned} \text{Finalement, } Q_n(F') &= \frac{1}{n} \mathbb{E} \left[ \sup_{\substack{(x,y) \\ \in \mathcal{E}}} \sum_{f \in F} \varepsilon_i^* q_i^* f(x_i) \right] \\ &= \frac{1}{n} \mathbb{E} \left[ \sup_{\substack{(x,y) \\ \in \mathcal{E}}} \sum_{f \in F} \varepsilon_i^* f(x_i) \right] \\ &= Q_n(F), \end{aligned}$$

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

Ex 5

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

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

On considère l'estimateur suivant :

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

$\forall \lambda > 0$ .

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

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

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



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

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

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

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

avec  $c$  tel que

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

$$\left\langle \max\{0, 1 - Y f(x)\} \right\rangle \text{ s. sign } f$$

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

$$\langle 1 + \|f\|_F \|\kappa(x, \cdot)\|_F \rangle$$

$$\langle 1 + \|f\|_F \rangle$$

$$\langle 1 + \frac{\eta}{\sqrt{n}} \rangle$$

Mc Diarmid avec

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

On sait que  $\Psi$  est 1-Lipschitz. On a donc

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

Aussi,

$$\begin{aligned} Q_{L_n}(\tilde{F}) &= \mathbb{E}\left[\mathbb{E}\left[\sup_{f \in \tilde{F}_\lambda} \frac{1}{n} \sum_i \underbrace{\varepsilon_i}_{\sim \mathcal{E}_i} Y_i f(X_i) \mid D_n = (X_i, Y_i)\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sup_{f \in \tilde{F}_\lambda} \frac{1}{n} \sum_i \varepsilon_i f(X_i) \mid D_n\right]\right] \quad Y_i \in \{-1, 1\} \\ &= \mathbb{E}\left[\frac{\sup_{f \in \tilde{F}_\lambda} \|f\|_F \sqrt{T_n(K)}}{\sqrt{n}}\right] \\ &\leq \frac{\sup_{f \in \tilde{F}_\lambda} \|f\|_F \sqrt{Q^2}}{\sqrt{n}} = \frac{\sup_{f \in \tilde{F}_\lambda} \|f\|_F Q}{\sqrt{n}} \end{aligned}$$

$$\leq \frac{Q}{\sqrt{n\lambda}} \cdot (q^{c_n})$$

Finalement, on obtient :

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

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

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

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

$$\text{ou } C_n = \frac{2\beta}{\sqrt{n\lambda_n}} + \left(1 + \frac{\beta}{\sqrt{\lambda_n}}\right) \sqrt{\frac{\ln(2/\delta)}{2n}}$$

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

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

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

$$\text{Donc } A(f^*) - \hat{A}_n(f^*) \leq 0 \vee -C_n \leq A(f^*) - \hat{A}_n(f^*).$$

Dans le cas 1, l'inéq est bornée par  $C_n$ . Dans le 2<sup>e</sup>, par  $2C_n$ . Or,  $C_n \geq 0$   
(car  $0 \leq \delta \leq 2$ , afin que  $0 \leq \frac{1-\delta}{2} \leq 1$ ) et donc  $C_n \leq 2C_n$ .

Ainsi,

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

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

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

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

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

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

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

On a

$$L(\text{sgn}(\hat{f}_n^{\lambda_n})) = P(\text{sgn}(\hat{f}_n^{\lambda_n}(x)) \neq Y) = P(Y \hat{f}_n^{\lambda_n}(x) \leq 0) \\ = \mathbb{E}[\mathbb{I}_{\{Y \hat{f}_n^{\lambda_n}(x) \leq 0\}}]$$

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

$$\psi(Y \hat{f}_n^{\lambda_n}(x)) \geq 1 \geq \mathbb{I}_{\{.\}}$$

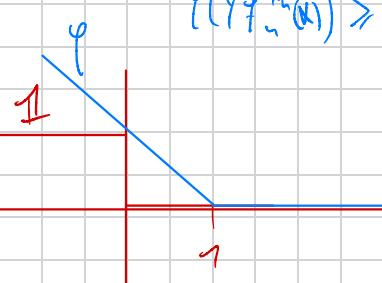
- si  $Y \hat{f}_n^{\lambda_n}(x) \geq 1$ ,

$$\psi(Y \hat{f}_n^{\lambda_n}(x)) = 0 = \mathbb{I}_{\{.\}}$$

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

$$\psi(Y \hat{f}_n^{\lambda_n}(x)) \geq 0 = \mathbb{I}_{\{.\}}$$

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



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

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

$$D_n' = \left( (X'_1, Y'_1), \dots, (X'_n, Y'_n) \right)$$

$$\text{tq. } \forall i \neq k \quad (X_i, Y_i) = (X'_k, Y'_k)$$

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

$$\forall f \in \hat{F}_{n'}, \forall y, |\varphi(y, f(x))| \leq \Lambda \triangleq 1 + \frac{\beta}{\sqrt{\lambda}}$$

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

( $\varphi$  1-Lipschitz)

$$\forall f \in F, \forall x, \|f(x)\| = \|\langle L_k(x, \cdot), f \rangle_F\|$$

$$\leq \|\langle L_k(x, \cdot) \rangle_F\| \|f\|_F$$

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

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

$$\leq \beta \|f\|_F$$

$$\leq \beta \|f\|_F$$

$$\Rightarrow \|f\|_\infty \leq \beta \|f\|_F$$

(Proposition 11.1,

Foundations of ML

Mohri et al. 1<sup>st</sup> ed)

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

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

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

$$Y_n = \frac{\sigma^2}{\lambda_n} \longrightarrow 0.$$

$$\sqrt{n}Y_n = \frac{\sigma^2}{\lambda_n\sqrt{n}} \longrightarrow 0.$$

Ainsi,  $Y_n + (2nY_n + 1) \sqrt{\frac{\log(2/\delta)}{2n}} \longrightarrow 0$ . Appliquer la

même stratégie que à la question précédente termine la preuve.

## Preuve de ① :

$\Omega_n$  cherche à majorer

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

$$\begin{aligned}
 ② \quad |\widehat{A}_n(\hat{f}) - \widehat{A}_n(\hat{f}'_n)| &= \frac{1}{n} \left| \sum_{i=1}^n \left( Y_i, \hat{f}_n(X_i) \right) - \left( Y'_i, \hat{f}'_n(X'_i) \right) \right| \\
 &\leq \frac{1}{n} \sum_{i=1}^n \left| Y_i, \hat{f}_n(X_i) - Y'_i, \hat{f}'_n(X'_i) \right| \\
 &\quad + \frac{1}{n} \left| Y'_i, \hat{f}_n(X_k) - Y'_i, \hat{f}'_n(X'_k) \right| \\
 &\leq \frac{(n-1)\gamma_n}{n} + \frac{\Delta}{n} \leq \gamma_n + \frac{\Delta}{n}
 \end{aligned}$$

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

$$\begin{aligned}
|A(\hat{f}_n) - \hat{A}_n(\hat{f}_n) - E[\hat{f}_n] + \hat{A}_n(\hat{f}_n)| &= |A(\hat{f}_n) - A(\hat{f}_n) + \hat{A}_n(\hat{f}_n) - \hat{A}_n(\hat{f}_n)| \\
&\leq |A(\hat{f}_n) - A(\hat{f}_n)| + |\hat{A}_n(\hat{f}_n) - \hat{A}_n(\hat{f}_n)| \\
&\leq 2\gamma_n + \frac{\Delta}{n} \stackrel{\Delta}{=} c.
\end{aligned}$$

On peut appliquer l'inégalité de Mc Diarmid:

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

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

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

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

Ainsi,

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

$\Omega_1$ ,

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

$\leq \gamma_n$ .

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

Finalement, avec probabilité  $1 - \delta$ ,

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