①

Leçon n°4 : EMMTT Lien forte proba - espérance, Complexité de Rademacher, application aux modèles linéaires

Contexte : Contrôle de $2 \sup_{f \in F} |\hat{M}(f) - M(f)|$

(Erreur d'estimation)

I. Lien forte Proba - Espérance

$$\hat{M}(f) = \frac{1}{n} \sum_{i=1}^{n} l(y_i, f(x_i))$$

Hypothèse : $\forall f \in F, \forall (x, y) \in \text{Supp}(\text{loi des données})$,

$$0 \leq l(y_i, f(x_i)) \leq l_\infty .$$

Remarque : En changeant un couple (x_i, y_i) par n'importe quoi qui reste dans le support de la loi des données, $\hat{M}(f)$ change au plus de $\frac{l_\infty}{n}$ (vrai $\forall f$).

Et donc, $\left(\sup_{f \in F} |\hat{M}(f) - M(f)| \right)$ change au plus de $\frac{l_\infty}{n}$.

②

Il est alors possible d'appliquer l'inégalité de McDarmid à $\sup_{f \in F} |\hat{M}(f) - M(f)|$, ce qui donne, avec probabilité $1-\delta$, (pour $\delta \in (0,1)$),

$$\left| \sup_{f \in F} |\hat{M}(f) - M(f)| - \mathbb{E}\left(\sup_{f \in F} |\hat{M}(f) - M(f)| \right) \right|$$
$$\leq \frac{l_\alpha}{\sqrt{2n}} \sqrt{\log \frac{1}{\delta}}$$

<u>Donc</u> Pour obtenir des bornes avec forte proba, il suffit d'être capable de donner des bornes en espérance.

(3)

II Complexité de Rademacher

__Idée__ Pour donner des bornes sur $\mathbb{E}\left(\sup_{f\in F} \left|\tilde{R}(f) - R(f)\right|\right)$, il existe un outil générique appelé la complexité de Rademacher.

__Reformulation__ On reparamétrise les fonctions f (et F) par des fonctions h (et H). Nous sommes alors intéressés par

$$\sup_{h\in H} \left| \mathbb{E}\left(\cancel{\sup_{h\in H}}\, h(z)\right) - \frac{1}{n}\sum_{i=1}^{n} h(z_i) \right|$$

__Définition :__ $$R_n(H) \equiv \mathbb{E}_{\varepsilon, z}\left(\sup_{h\in H} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i h(z_i)\right)$$

où $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ est une suite de variables de Rademacher indépendantes entre elles et de toutes les autres quantités.

$R_n(H)$ s'appelle la complexité de Rademacher de la classe H.

Théorème (Propriété de symétrisation).

Les deux inégalités suivantes sont vérifiées :

- $$\mathbb{E}\left(\sup_{h\in H}\left(\frac{1}{n}\sum_{i=1}^{n} h(z_i) - \mathbb{E}(h(z))\right)\right) \le 2R_n(H)$$

- $$\mathbb{E}\left(\sup_{h\in H}\left(\mathbb{E}(h(z)) - \frac{1}{n}\sum_{i=1}^{n} h(z_i)\right)\right) \le 2R_n(H).$$

preuve : L'idée est d'introduire des variables aléatoires $z_1', \dots, z_n'$ qui sont indépendantes entre elles et des autres quantités du problème.

Alors $$\mathbb{E}_{z_1,\dots,z_n}\left(\sup_{h\in H}\left(\frac{1}{n}\sum_{i=1}^{n} h(z_i) - \mathbb{E}_z(h(z))\right)\right)$$

$$= \mathbb{E}_{z_1,\dots,z_n}\left(\sup_{h\in H}\left(\frac{1}{n}\sum_{i=1}^{n} h(z_i) - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{z_1',\dots,z_n'}(h(z_i'))\right)\right)$$

$$= \mathbb{E}_{z_1,\dots,z_n}\left(\sup_{h\in H}\left(\mathbb{E}_{z_1',\dots,z_n'}\left(\frac{1}{n}\sum_{i=1}^{n}(h(z_i) - h(z_i'))\right)\right)\right)$$

$$\le \cancel{\sup_{h\in H}}\ \mathbb{E}_{\substack{z_1,\dots,z_n\\ z_1',\dots,z_n'}}\left(\sup_{h\in H}\left(\frac{1}{n}\sum_{i=1}^{n}(h(z_i) - h(z_i'))\right)\right)$$

(5)

De plus, z_i et z_i' ont des lois interchangeables (donc $h(z_i)-h(z_i')$ est symétrique) et ε_i est symétrique.

donc

$$\mathbb{E}\left(\sup_{h\in H}\left(\frac{1}{n}\sum_{i=1}^{n} h(z_i) - \mathbb{E}(h(z))\right)\right)$$

$$\leq \mathbb{E}_{\substack{z_1,\dots,z_n\\ z_1',\dots,z_n'\\ \varepsilon_1,\dots,\varepsilon_n}}\left(\sup_{h\in H}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left(h(z_i)-h(z_i')\right)\right)\right)$$

$$\leq \mathbb{E}_{\text{''}}\left(\sup_{\text{''}}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i h(z_i)\right)\right) + \mathbb{E}_{\text{''}}\left(\sup_{\text{''}}\left(\frac{1}{n}\sum_{i=1}^{n}-\varepsilon_i h(z_i')\right)\right)$$

$$= 2R_n(H)$$

□

III Analyse des modèles linéaires contraints via la complexité de Rademacher

1) ~~Calcul de la complexité de Rademacher~~

↳ À présenter après le calcul de complexité.

Lemme : Lorsque la fonction $\ell(y,\cdot)$ est G-Lipschitz p.s., alors

$$\mathbb{E}\left(\sup_{f\in F}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i \ell(y_i, f(x_i))\right) \leq G.\,\mathbb{E}\left(\sup_{f\in F}\sum_{i=1}^{n}\varepsilon_i f(x_i)\right)$$

preuve : Admis, voir Bach 2024 □

(6)

Modèles linéaires contraints :

$$\mathcal{F} = \left\{ f_\theta(x) = \theta^T \varphi(x),\ \Omega(\theta) \leq D \right\}$$

Ω : norme sur $\mathbb{R}^d$.

$$R_n(F) = \mathbb{E}\left[\sup_{\Omega(\theta) \leq D} \left(\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \theta^T \varphi(x_i) \right) \right]$$

$$= \mathbb{E}\left[\sup_{\Omega(\theta) \leq D} \frac{1}{n} \varepsilon^T \Phi \theta \right] \quad \text{(Notations leçon 2)}$$

$$= \frac{D}{n} \mathbb{E}\left(\Omega^*\left(\Phi^T \varepsilon \right) \right) \quad (\Omega^* : \text{norme duale de } \Omega).$$

Lorsque $\Omega(\cdot) = \|\cdot\|_2$:

$$R_n(F) = \frac{D}{n} \mathbb{E}\left(\|\Phi^T \varepsilon\|_2 \right)$$

$$\underset{\text{Jensen}}{\leq} \frac{D}{n} \sqrt{\mathbb{E}\left(\|\Phi^T \varepsilon\|_2^2 \right)}$$

$$= \frac{D}{n} \sqrt{\mathbb{E}\left(\varepsilon^T \Phi \Phi^T \varepsilon \right)}$$

$$= \frac{D}{n} \sqrt{\mathbb{E}\left(\operatorname{tr}\left(\varepsilon^T \Phi \Phi^T \varepsilon \right) \right)}$$

$$= \frac{D}{n} \sqrt{\mathbb{E}\left(\operatorname{tr}\left(\Phi^T \varepsilon \varepsilon^T \Phi \right) \right)}$$

7

$$= \frac{D}{n}\sqrt{\mathbb{E}\left(\mathbb{E}\left(\mathrm{tr}\left(\Phi^T \varepsilon \varepsilon^T \Phi\right) \middle| \Phi\right)\right)}$$

$$= \frac{D}{n}\sqrt{\mathbb{E}\left(\mathrm{tr}\left(\Phi^T \underbrace{\mathbb{E}\left(\varepsilon \varepsilon^T \middle| \Phi\right)}_{= I} \Phi\right)\right)}$$

$$= \frac{D}{n}\sqrt{\mathbb{E}\left(\mathrm{tr}\left(\Phi^T \Phi\right)\right)} = \frac{D}{n}\sqrt{\sum_{i=1}^{n} \mathbb{E}\left(\Phi^T \Phi\right)_{ii}}$$

$$= \frac{D}{n}\sqrt{\sum_{i=1}^{n} \mathbb{E}\left(\|\varphi(x_i)\|_2^2\right)}$$

$$= \frac{D}{\sqrt{n}}\sqrt{\mathbb{E}\left(\|\varphi(x)\|_2^2\right)}$$

donc $\boxed{R_n(F) \leq \underbrace{\frac{D}{\sqrt{n}}\sqrt{\mathbb{E}\left(\|\varphi(x)\|_2^2\right)}}_{\text{Indep de la dimension.}}}$

Indep de la dimension.

<u>Conclusion:</u> Lorsque $\ell(y, \cdot)$ est G-Lipschitz p.s.

Sur $\bigcup_{f \in F} f(\mathrm{Supp}(x))$, alors

$$\mathbb{E}\left(R\left(f_{\hat{\theta}}\right)\right) \leq \inf_{\|\theta\|_2 \leq D} R\left(f_\theta\right) + \frac{4 G D \sqrt{\mathbb{E}\left(\|\varphi(x)\|_2^2\right)}}{\sqrt{n}}$$