Entropie et robustesse

Olivier Catoni
CNRS - DMA

en collaboration avec

Jean-Yves Audibert
CERTIS - DI

Jeudi 29 janvier 2009

Agrégation d'estimateurs

$$\varphi : \mathcal{X} \to \mathbb{R}^d$$

$$(X_i, Y_i)_{i=1}^n \sim \mathbb{P}^{\otimes n}, \quad \mathbb{P} \in \mathcal{M}_+^1(\mathcal{X} \times \mathbb{R}),$$

Θ convexe fermé de $\mathbb{R}^d$,

$$\inf_{\theta \in \Theta} \mathbb{E}\left[(\langle \theta, \varphi(X)\rangle - Y)^2\right] ?$$

$$\mathcal{F} = \left\{ f_\theta : f_\theta(X) = \langle \theta, \varphi(X)\rangle, \theta \in \Theta \right\},$$

$$\widehat{f} \in \arg\min_{f_\theta \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \left[f_\theta(X_i) - Y_i \right]^2 + \lambda \|\theta\|^2$$

« régression ridge » : $\Theta = \mathbb{R}^d$

« estimateur des moindres carrés » : $\lambda = 0, \Theta = \mathbb{R}^d$.

Questions :
comportement quand φ n'est pas bornée ?
quand Y/X n'a pas de moment exponentiel ?
Le risque empirique

$$r(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left[f_\theta(X_i) - Y_i \right]^2$$

est-il la bonne quantité sur laquelle se fonder ?

Théorème

Supposons que

$$\mathbb{E}\big[\|\varphi(X)\|^4\big] < +\infty, \quad \mathbb{E}\Big\{\|\varphi(X)\|^2\big[f^*(X) - Y\big]^2\Big\} < +\infty,$$

alors pour tout $\epsilon > 0$, il existe n_ϵ tel que pour tout $n > n_\epsilon$, avec probabilité au moins $1 - \epsilon$,

$$R_\lambda(\widehat{f}) - R_\lambda(f^*) \leq \frac{30D}{n}\widetilde{\mathbb{E}}\big\{\big[f^*(X) - Y\big]^2\big\} + \frac{1000\log(3/\epsilon)}{n} \sup_{v \in \mathbb{R}^d} \widetilde{\widetilde{\mathbb{E}}}_v\big\{\big[f^*(X) - Y\big]^2\big\},$$

$$\text{où } D = \sum_{k=1}^{d} \frac{\nu_k}{\nu_k + \lambda}\mathbb{1}(\nu_k > 0), \{\nu_1, \dots, \nu_d\} = \mathrm{Sp}\big(\mathbb{E}\big[\varphi(X)\varphi(X)^T\big)\big],$$

$$R_\lambda(f_\theta) = \mathbb{E}\big\{\big[f_\theta(X) - Y\big]^2\big\} + \lambda\|\theta\|^2, \quad f^* \in \arg\min_{\Theta} R_\lambda.$$

Précisions

$$\text{Si } Q_\lambda = \mathbb{E}\big[\varphi(X)\varphi(X)^T\big] + \lambda I,$$

$$\widetilde{\mathbb{E}}(Z) = \frac{\mathbb{E}\big\{\|Q_\lambda^{-1/2}\varphi(X)\|^2 Z\big\}}{\mathbb{E}\big\{\|Q_\lambda^{-1/2}\varphi(X)\|^2\big\}},$$

$$\widetilde{\widetilde{\mathbb{E}}}_v(Z) = \frac{\mathbb{E}\big[\langle v, \varphi(X)\rangle^2 Z\big]}{\mathbb{E}\big[\langle v, \varphi(X)\rangle^2\big]}.$$

Théorème

$$\text{\emph{Supposons} } \lambda = 0,$$

$$\sup_{v \in \mathbb{R}^d} \frac{\operatorname{ess\,sup} \langle v, \varphi(X) \rangle^2}{\mathbb{E}\left[\langle v, \varphi(X)\rangle^2\right]} < B < \infty,$$

$$\mathbb{E}\left\{\left[Y - f^*(X)\right]^4\right\} < +\infty,$$

$$2/n \le \epsilon \le 1,$$

$$n > 1280\, B^2 \left[3Bd' + \log(2\epsilon^{-1}) + \frac{(4Bd')^2}{n}\right].$$

Avec probabilité au moins $1 - \epsilon$,

$$R(\widehat{f}) - R(f^*) \le 640\, B \left\{2R(f^*) + \mathbb{E}\left\{\left[f^*(X) - Y\right]^4\right\}^{1/2}\right\} \times \left[\frac{3Bd' + \log(2/\epsilon)}{n} + \left(\frac{4Bd'}{n}\right)^2\right].$$

Théorème

$$\textit{Supposons}\ \mathbb{E}\Big\{\big[f^*(X) - Y\big]^2 \,\big|\, X\Big\} \le H_1,$$

$$\sup_{v\in\mathbb{R}^d} \frac{\mathbb{E}\big[\langle v, \varphi(X)\rangle^4\big]^{1/2}}{\mathbb{E}\big[\langle v, \varphi(X)\rangle^2\big] + \lambda\|v\|^2} \le H_2,$$

$$\mathrm{Sp}(Q) \in [q_{\min}, q_{\max}].$$

Il existe un estimateur, obtenu par une méthode de troncature, $\widetilde{f}$, qui vérifie avec probabilité $1-\epsilon$,

$$R_\lambda(\widetilde{f}) - R_\lambda(f^*) \le \frac{32 H_1 q_{\max} d}{(q_{\min}+\lambda)n} + \big[4H_1 + H_2^2(q_{\max}+\lambda)\|\Theta\|^2\big]\bigg\{\frac{576 H_2^2 (q_{\max}+\lambda)^2 d^2}{(q_{\min}+\lambda)^2 n^2} + \frac{48\log(2/\epsilon)}{n}\bigg\}.$$

$$
\text{Description : } \rho_{\theta_1}(d\theta) = \left(\frac{\beta}{2\pi}\right)^{d/2} \exp\left(-\frac{\beta}{2}\|\theta-\theta_1\|^2\right) d\theta,
$$

$$
W_i(\theta,\theta') = \bigl[\langle\theta,\varphi(X_i)\rangle - Y_i\bigr]^2 - \bigl[\langle\theta',\varphi(X_i)\rangle - Y_i\bigr]^2,
$$

$$
\widetilde{\theta} \in \arg\min_{\theta_1\in\Theta} \sup_{\theta_2\in\Theta} -\frac{1}{n}\sum_{i=1}^{n} \int \rho_{\theta_1}(d\theta') \log\Bigl\{ \int \rho_{\theta_2}(d\theta) \Bigl[1
$$

$$
-\alpha W_i(\theta',\theta) + \frac{\alpha^2}{2} W_i^2(\theta',\theta)\Bigr]\Bigr\} + \alpha\lambda\bigl[\|\theta_1\|^2 - \|\theta_2\|^2\bigr],
$$

pour des valeurs bien choisies de α et β.

Théorème

$$\textit{Supposons } \Theta \textit{ borné},$$
$$\sup_{f_1, f_2 \in \mathcal{F}, x \in \mathcal{X}} |f_1(x) - f_2(x)| \leq H,$$
$$\text{ess sup}\, \mathbb{E}\Big\{ \big[f^*(X) - Y \big]^2 \,|\, X \Big\} \leq \sigma^2 < +\infty.$$

Il existe alors un estimateur $\widetilde{f}$ *tel qu'avec probabilité* $1 - \epsilon$

$$R(\widetilde{f}) - R(f^*) \leq (2\sigma + H)^2 \frac{16{,}6d + 12{,}5 \log(2\epsilon^{-1})}{n}$$

Description : soit

$$\frac{d\widetilde{\pi}}{d\pi}(\theta) = \frac{\left\{\int_\Theta \pi(d\theta') \prod_{i=1}^{n} \left[1 - \alpha W_i(\theta,\theta') + \frac{\alpha^2}{2} W_i(\theta,\theta')\right]^{-1}\right\}^{-1}}{\int_\Theta \pi(d\theta'') \left\{\int_\Theta \pi(d\theta') \prod_{i=1}^{n} \left[1 - \alpha W_i(\theta'',\theta') + \frac{\alpha^2}{2} W_i(\theta'',\theta')\right]^{-1}\right\}^{-1}},$$

où π est la mesure de Lebesgue sur Θ,

posons $\widetilde{\theta} = \int_\Theta \theta \widetilde{\pi}(d\theta)$

et $\widetilde{f} = f_{\widetilde{\theta}}$.

Les propriétés de ce type d'estimateur, dans sa version randomisée où on tire $\widetilde{\theta}$ suivant $\widetilde{\pi}$, se généralisent à toute fonction de perte (pas nécessairement convexe) vérifiant une inégalité de marge.

Deux éléments importants des preuves :
pour un échantillon i.i.d. $(Z_i)_{i=1}^n$ d'une v.a. réelle Z, avec probabilité $1 - \epsilon$

$$
\begin{aligned}
n\alpha\mathbb{E}(Z) - \frac{n\alpha^2}{2}\mathbb{E}(Z^2) &\leq -n\log\Big[1 - \alpha\mathbb{E}(Z) + \frac{\alpha^2}{2}\mathbb{E}(Z^2)\Big] \\
&\leq -\sum_{i=1}^n \log\Big(1 - \alpha Z_i + \frac{\alpha^2}{2}Z_i^2\Big) + \log(\epsilon^{-1}) \\
&\leq \sum_{i=1}^n \log\Big(1 + \alpha Z_i + \frac{\alpha^2}{2}Z_i^2\Big) + \log(\epsilon^{-1}) \\
&\leq \alpha\sum_{i=1}^n Z_i + \frac{\alpha^2}{2}\sum_{i=1}^n Z_i^2 + \log(\epsilon^{-1}).
\end{aligned}
$$

Inégalité d'entropie :

$$\int \rho(d\theta) h(\theta) \leq \log\left[\int \pi(d\theta) \exp\left[h(\theta)\right]\right] + \mathcal{K}(\rho, \pi).$$

les deux ensemble : avec probabilité $1-\epsilon$, *pour tout* $\theta_2 \in \Theta$

$$\begin{aligned} - n \log\left\{\int \rho_{\theta_2}(d\theta)\left(1 - \alpha \mathbb{E}\left[W_i(\theta,\theta^*) + \frac{\alpha^2}{2} \mathbb{E}\left[W_i^2(\theta,\theta^*)\right]\right)\right\} \\ \leq \sum_{i=1}^n \log\left\{\int \rho_{\theta_2}(d\theta)\left(1 + \alpha W_i(\theta,\theta^*) + \frac{\alpha^2}{2} W_i^2(\theta,\theta^*)\right)\right\} \\ + \mathcal{K}(\rho_{\theta_2}, \rho_{\theta^*}) + \log(\epsilon^{-1}). \end{aligned}$$