Correction Examen

2020 - 2021

Exercice 1 :

1. D'après le cours, le risque de Bayes associé à l'erreur de classification est égal à :

$$\mathbb{E}\left[\min\left(\eta(X), 1-\eta(X)\right)\right]$$

Comme $\frac{x}{x+\theta} < \frac{1}{2} \iff 2x < x+\theta \iff x<\theta$,

on a :

$$\mathbb{E}\left[\min\left(\eta(X), 1-\eta(X)\right)\right] = \mathbb{E}\left[\frac{X}{X+\theta} \mathbb{1}_{\{X \leq \theta\}}\right] + \mathbb{E}\left[\underbrace{\left(1-\frac{X}{X+\theta}\right)}_{\frac{\theta}{X+\theta}} \mathbb{1}_{\{X>\theta\}}\right]$$

$$= \frac{1}{\alpha\theta}\int_0^\theta \frac{x}{x+\theta}\,dx + \frac{1}{\alpha\theta}\int_\theta^{\alpha\theta} \frac{\theta}{x+\theta}\,dx$$

$$= \frac{1}{\alpha\theta}\int_0^\theta -\frac{\theta}{x+\theta}+1\,dx + \frac{1}{\alpha}\left[\ln(x+\theta)\right]_\theta^{\alpha\theta}$$

$$= \frac{1}{\alpha} - \frac{1}{\alpha}\left[\ln(x+\theta)\right]_0^\theta + \frac{1}{\alpha}\ln\left(\frac{(1+\alpha)\theta}{2\theta}\right)$$

$$= \frac{1}{\alpha} - \frac{1}{\alpha}\ln\left(\frac{2\theta}{\theta}\right) + \frac{1}{\alpha}\ln\left(\frac{(1+\alpha)}{2}\right)$$

$$= \boxed{\frac{1}{\alpha}\left(1+\ln\left(\frac{1+\alpha}{4}\right)\right)}.$$

2. (a) ②

$$L(g) = \mathbb{E}\left[\ell(Y, g(X))\right] = \mathbb{P}(Y=1, g(X)=0) + \mathbb{P}(Y=0, g(X)=1) + \lambda \mathbb{P}(g(X)=1)$$

$$= \int \underbrace{\mathbb{P}(Y=1 / \cancel{g} = x)}_{\eta(x)} \cdot 1_{\{g(x)=0\}} \mathbb{P}_X(dx)$$
$$+ \int \underbrace{\mathbb{P}(Y=0 / X=x)}_{(1-\eta(x))} 1_{\{g(x)=1\}} \mathbb{P}_X(dx)$$
$$+ \cancel{\int \mathbb{P}(} \lambda\, \mathbb{E}[1_{\{g(X)=1\}}]$$

$$= \mathbb{E}\left[\eta(X) 1_{\{g(X)=0\}} + (1-\eta(X)) 1_{\{g(X)=1\}} + \lambda 1_{\{g(X)=1\}}\right]$$

d'où le résultat.

(b) Pour déterminer g^*, on résoud dans un premier temps l'inégalité $1+\lambda-p \leq p \iff p \geq \frac{1+\lambda}{2}$.
$(p \in [0,1])$

On pose alors :

$$g^*(x) = \begin{cases} 0 & \text{si } \eta(x) \leq \frac{1+\lambda}{2} \\ 1 & \text{si } \eta(x) > \frac{1+\lambda}{2}. \end{cases}$$

Par construction, $\forall x \in \mathcal{X}$,

1) $\forall g$, $(1+\lambda-\eta(x)) 1_{\{g(x)=1\}} + \eta(x) 1_{\{g(x)=0\}} \geq \min(1+\lambda-\eta(x), \eta(x))$

2) $(1+\lambda-\eta^*(x)) 1_{\{g^*(x)=1\}} + \eta^*(x) 1_{\{g^*(x)=0\}} = \min(1+\lambda-\eta(x), \eta(x))$

On en déduit donc que :

$$L(g^*) = \mathbb{E}\left[\min(1+\lambda-\eta(\cancel{x}X), \eta(X))\right] \leq L(g) \quad \forall g.$$

puis que :

$g^* = \text{Argmin } L(g)$ et risque optimal $= L(g^*)$.

Exercice 2 :

$$\mathcal{L}(\theta) = \frac{1}{2N} \sum_{k=1}^{N} \Psi(Y_i, \langle x_i, \theta \rangle) + \lambda \overbrace{\left(\alpha \|\theta\|_1 + (1-\alpha) \frac{\|\theta\|_2^2}{2} \right)}^{\mathcal{P}_\alpha(\theta)}$$

1) Si $\Psi(y, z) = y \frac{e^z}{1+e^z} + (1-y)\frac{1}{1+e^z}$, alors

$$\mathcal{L}(\theta) = \underbrace{\frac{1}{2N} \sum_{k=1}^{N} Y_i \frac{e^{\langle x_i, \theta \rangle}}{1+e^{\langle x_i, \theta \rangle}} + (1-Y_i) \frac{1}{1+e^{\langle x_i, \theta \rangle}}}_{A} + \underbrace{\mathcal{P}_\alpha(\theta)}_{B}.$$

A un facteur 2 près, A correspond à la log. vraisemblance associée à la regression logistique, i.e $\pi(x, \theta) = \frac{e^{\langle x, \theta \rangle}}{1+e^{\langle x, \theta \rangle}}$

tandis que $\mathcal{P}_\alpha(\theta)$ désigne une pénalité de type Elastic Net. On a donc ici un estimateur associé à la regression logistique pénalisée avec pénalité Elastic Net.

Sur R, ça donne :

glmnet (X , Y , family = "binomial", s = λ, alpha = α)

matrice de design — vecteur des réponses

2. (a) $\theta \longmapsto \langle x, \theta \rangle$ est convexe, ϕ et $u \to u^2$ le sont également donc par composition, $\theta \to (y - \phi(\langle x, \theta \rangle))^2$ l'est également. Par ailleurs $\theta \to \|\theta\|_1$ et $\theta \to \|\theta\|_2^2$ sont clairement convexes. On a donc une somme de fonctions convexes qui est donc convexe.

(b). $\partial_{\theta_j}\left[(y - \phi(\langle x, \theta \rangle))^2\right] = 2(\phi(\langle x, \theta \rangle - y) . \partial_{\theta_j} (\phi(\langle x, \theta$

Or, $\partial_{\theta_j}(\phi(\langle x, \theta \rangle)) = \phi'(\langle x, \theta \rangle) . x_j$.

Donc $\partial_{\theta_j}\left[(y - \phi(\langle x, \theta \rangle)^2\right] = 2(\phi(\langle x, \theta \rangle - y) . \phi'(\langle x, \theta \rangle) . x_j$.

On en déduit donc que :

$$\partial_{\theta_j} \mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \phi'(\langle X_i, \theta \rangle) . \underset{\substack{\uparrow \\ j\text{ième coord.}}}{X_i^j} (\phi(\langle X_i, \theta \rangle - Y_i) *$$

$$+ \lambda(1-\alpha)\theta_j + \underbrace{\begin{cases} \lambda\alpha \operatorname{sgn}(\theta_j) & \text{si } \theta_j \neq 0 \\ \lambda\alpha [-1, 1] & \text{si } \theta_j = 0. \end{cases}}_{②}$$

(c) Si $\phi(x) = x$, alors,

$$\partial_{\theta_j} \mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^{N} X_i^j (\cancel{\phi}\langle X_i, \theta \rangle - Y_i) + \lambda(1-\alpha)\theta_j + ② .$$

Ainsi par le cours, $\theta \in \operatorname{Argmin} \mathcal{L} \iff 0 \in \partial_{\theta_j} \mathcal{L}(\theta)$.

$$\iff \forall j, \begin{cases} \frac{1}{N} \sum_{i=1}^{N} X_i^j (\cancel{\phi}\langle X_i, \theta \rangle - Y_i) + \lambda(1-\alpha)\theta_j + \lambda\alpha \operatorname{sgn}(\theta_j) & \text{si } \theta_j \neq 0 \\ \frac{1}{N} \sum_{i=1}^{N} X_i^j (\langle X_i, \theta \rangle - Y_i) + \lambda\alpha[-1, 1] \supset 0. & \text{si } \theta_j = 0. \end{cases}$$

(Matriciellement, ça donne :) ⑤

$$\Leftrightarrow \begin{cases} \frac{1}{N} x_j^T (Y - X\theta) = \lambda(1-\alpha)\theta_j + \lambda\alpha \,\mathrm{sgn}(\theta_j) \text{ si } \theta_j \neq 0 \quad \forall j \\ \left| \frac{1}{N} x_j^T (X\theta - Y) \right| \leq \lambda\alpha \quad \text{si } \theta_j = 0. \end{cases}$$

(où x_j = la jième colonne de X)

(d) $X^T X = I_p \Rightarrow$ l'équation ci-dessus devient

$$\begin{cases} \frac{x_j^T Y}{N} = \left(\frac{1}{N} + \lambda(1-\alpha)\right) \cdot \theta_j + \lambda\alpha \,\mathrm{sgn}(\theta_j) \quad \text{si } \theta_j \neq 0 \\ \left| \frac{1}{N} \langle x_j, y \rangle \right| \leq \lambda\alpha \quad \text{si } \theta_j = 0. \end{cases}$$

Ainsi, si $\frac{\langle x_j, Y\rangle}{N} > \lambda\alpha$, alors $\theta_j > 0$ et :

$$\theta_j = \left(\frac{1}{N} + \lambda(1-\alpha)\right)^{-1} \left(\frac{\langle x_j, Y\rangle}{N} - \lambda\alpha\right),$$

si $\frac{\langle x_j, Y\rangle}{N} < -\lambda\alpha$, alors $\theta_j < 0$ et :

$$\theta_j = \left(\frac{1}{N} + \lambda(1-\alpha)\right)^{-1} \left(\lambda\alpha + \frac{\langle x_j, Y\rangle}{N}\right)$$

$$= -\left(1 + \lambda N(1-\alpha)\right)^{-1} \left(\frac{|\langle x_j, Y\rangle|}{N} - \lambda\alpha\right)$$

et enfin si $\left|\frac{1}{N}\langle x_j, Y\rangle\right| \leq \lambda\alpha$, alors $\theta_j = 0$.

La formule suit.

Exercice 3.

1. Voir TD
2. Voir TD (Exprimer en fonction de $\mathrm{Corr}(\hat{f}_{b_1}, \hat{f}_{b_2}) =: \rho$.
3. Du point de vue biais-variance, la réponse est ~~pas~~ a priori "oui" car par construction, le "biais est identique" et d'après la question 2 la variance est toujours plus petite. Plus précisément, $\forall x$, $\mathbb{E}\left[\left(\frac{1}{B}\sum_{b=1}^{B} \hat{f}_b(x) - f(x)\right)^2\right] \leq \mathrm{Biais}^2(\hat{f}_b) + \mathrm{Var}(\hat{f}_b)$

Néanmoins $\hat{f}_b$ n'est pas exactement l'estimateur de base car il est basé sur un échantillon Bootstrap et non sur l'échantillon d'apprentissage initial mais on peut supposer que l'influence sur le biais est assez mineure. (6)