Devoir final

Mathieu Rita
16 décembre 2018

1 SOFT-THRESHOLDING

1.1 QUESTION 1

On définit l'estimateur de Lasso comme :

$$\hat{\beta}^L = argmin_{\beta \in \mathbb{R}^n} \left(\frac{1}{n} \left\| Y - \beta \right\|_{l_2}^2 + \lambda \left\| \beta \right\|_{l_1} \right) \tag{1.1}$$

Nous cherchons donc, en vertu de la définition des normes, à minimiser la fonction :

$$\begin{aligned} f : \mathrm{R}^n &\to \mathbb{R} \\ \beta &\mapsto \textstyle\sum_{i=1}^n \frac{1}{n}(Y_i - \beta_i)^2 + \lambda|\beta_i| \end{aligned}$$

Comme il s'agit d'une somme de termes positifs, cela revient à minimiser chacun des termes de la somme.

Nous cherchons donc :

$$\hat{\beta}_i^L = argmin_{\beta_i \in \mathbb{R}} \left(\frac{1}{n}(Y_i - \beta_i)^2 + \lambda|\beta_i| \right) \tag{1.2}$$

$$= \quad argmin_{\beta_i \in \mathbb{R}} \left(\frac{1}{n} Y_i^2 - \frac{2}{n} Y_i \beta_i + \frac{1}{n} \beta_i^2 + \lambda|\beta_i| \right) \tag{1.3}$$

$$= \quad argmin_{\beta_i \in \mathbb{R}} \left(-\frac{2}{n} Y_i \beta_i + \frac{1}{n} \beta_i^2 + \lambda|\beta_i| \right) \tag{1.4}$$

Nous cherchons donc à minimiser, pour tout $i \in [|1, n|]$, les fonctions :

$$f_i(\beta_i) = -\frac{2}{n} Y_i \beta_i + \frac{1}{n}\beta_i^2 + \lambda|\beta_i| \tag{1.5}$$

Les fonctions f_i n'étant pas dérivable en 0, on sépare l'étude en deux cas :

Cas $\beta_i > 0$:

Dans ce cas, $|\beta_i| = \beta_i$ et :

$$f_i'(\beta_i) = -\frac{2}{n} Y_i + \frac{2}{n}\beta_i + \lambda \tag{1.6}$$

f_i étant un polynôme de degré 2 à coefficient maximal positif, l'existence d'un minimum est garantie. Il est obtenu lorsque la condition $f_i'(\hat{\beta}_i^L) = 0$ est vérifiée, c'est-à-dire pour :

$$\hat{\beta}_i^L = Y_i - \frac{n}{2}\lambda \tag{1.7}$$

Ainsi : si $\beta_i > 0$: $\hat{\beta}_i^L = max(0, Y_i - \frac{n}{2}\lambda)$

Cas $\beta_i < 0$:

Dans ce cas, $|\beta_i| = -\beta_i$ et :

$$f_i'(\beta_i) = -\frac{2}{n} Y_i + \frac{2}{n}\beta_i - \lambda \tag{1.8}$$

f_i étant un polynôme de degré 2 à coefficient maximal positif, l'existence d'un minimum est garantie. Il est obtenu lorsque la condition $f_i'(\hat{\beta}_i^L) = 0$ est vérifiée, c'est-à-dire pour :

$$\hat{\beta}_i^L = Y_i + \frac{n}{2}\lambda \tag{1.9}$$

Ainsi : si $\beta_i < 0$: $\hat{\beta}_i^L = max(0, Y_i + \frac{n}{2}\lambda)$

On peut réunir ces cas avec l'écriture simplifiée :

$$\hat{\beta}_i^L = sign(Y_i)\left(|Y_i| - \frac{n}{2}\lambda\right)_+ \tag{1.10}$$

$$\hat{\beta}_i^L = \eta_{\frac{n}{2}\lambda}(Y_i) \tag{1.11}$$

1.2 QUESTION 2

Soit $n \geq 2$. On rappelle qu'on a, pour tout $i \in [|1, n|]$:

$$\left\{\begin{array}{l} Y_i = \beta_i + \epsilon_i \\ \hat{\beta_i^S} = sign(Y_i)\,(|Y_i| - \tau)_+ \\ \tau = A\sigma\sqrt{log(n)} \end{array}\right.$$

MONTRONS QUE : Pour un $A > 0$ assez grand, avec une probabilité au moins égale à $1 - n^{1-\frac{A^2}{8}}$:

$$\underline{||\hat{\beta}_S - \beta||_{l_2}^2 \leq C_A \sigma^2 ||\beta||_0 log(n)} \tag{1.12}$$

PREUVE

1/ OBTENTION DE L'INÉGALITÉ SOUS LA CONDITION : POUR TOUT $i \in [|1, n|]$, $|\epsilon_i| < \tau$

$$||\hat{\beta}_S - \beta||_2^2 = \sum_{i=1}^{n} (\hat{\beta_S^i} - \beta^i)^2 \tag{1.13}$$

Séparons les cas où $\beta_i = 0$ et $\beta_i \neq 0$ et définissons :

$$I = \{i \in [|1, n|]; \beta_i \neq 0\} \tag{1.14}$$

On a ainsi :

$$||\hat{\beta}_S - \beta||_2^2 = \underbrace{\sum_{i \in I} (\hat{\beta_S^i} - \beta^i)^2}_{S_1} + \underbrace{\sum_{i \notin I} (\hat{\beta_S^i} - \beta^i)^2}_{S_2} \tag{1.15}$$

Etudions chacune des deux sommes pour obtenir la majoration recherchée :

<u>(1) Cas $\beta_i \neq 0$:</u>

On a, pour $\beta_i \neq 0$:

$$\hat{\beta_S^i} - \beta^i = \left\{\begin{array}{ll} \beta^i + \epsilon^i - \tau - \beta^i = \epsilon^i - \tau & \text{si } Y_i > 0 \\ \beta^i + \epsilon^i + \tau - \beta^i = \epsilon^i + \tau & \text{sinon} \end{array}\right.$$

Or, sous la condition : <u>pour tout $i \in [|1, n|]$, $|\epsilon_i| < \tau$</u> :

$$\left\{\begin{array}{l} (\epsilon^i - \tau)^2 < 4\tau^2 \\ (\epsilon^i + \tau)^2 < 4\tau^2 \end{array}\right.$$

Donc, pour tout $i \in I$:

$$(\hat{\beta_S^i} - \beta^i)^2 \leq 4\tau^2 \tag{1.16}$$

Ce qui nous amène à l'inégalité suivante, si pour tout $i \in [|1,n|]$, $|\epsilon_i| < \tau$:

$$S_1 \leq \text{Card(I)}4\tau^2 \leq ||\beta||_{l_0}4\tau^2 \leq \underbrace{4A^2\sigma^2||\beta||_{l_0}\text{log(n)}}_{\tau = A\sigma\sqrt{log(n)}} \tag{1.17}$$

(2) Cas $\beta_i = 0$:

Si $\beta_i = 0$, on a :

$$\hat{\beta}^i_S - \beta^i = sign(\epsilon_i)\,(|\epsilon_i| - \tau)_+ \tag{1.18}$$

Ainsi, sous la condition pour tout $i \in [|1,n|]$, $|\epsilon_i| < \tau$:

$$\hat{\beta}^i_S - \beta^i = 0 \tag{1.19}$$

et donc, si pour tout $i \in [|1,n|]$, $|\epsilon_i| < \tau$:

$$S_2 = 0 \tag{1.20}$$

Nous obtenons donc en dénitif : si pour tout $i \in [|1,n|]$, $|\epsilon_i| < \tau$:

$$||\hat{\beta}_S - \beta||^2_{l_2} \leq \underbrace{C_A}_{=4A^2}\sigma^2||\beta||_{l_0}log(n) \tag{1.21}$$

2/ Obtention de la probabilité avec laquelle l'inéquation est réalisée

Il nous reste à prouver que l'inégalité est obtenue avec une probabilité au moins égale à $1 - n^{1-\frac{A^2}{8}}$.

Pour que l'inégalité soit obtenue, il faut que la condition "pour tout $i \in [|1,n|]$, $|\epsilon_i| < \tau$" soit réalisée. Nous introduisons donc l'événement :

$$\Omega = \bigcap_{j=0}^{n} \{|\epsilon^j| < \tau\} \tag{1.22}$$

Et montrons que : $\mathbb{P}(\Omega) \geq 1 - n^{1-\frac{A^2}{8}}$

On a, grâce à l'indépendance des ϵ^j :

$$\begin{aligned}
\mathbb{P}(\Omega) &= 1 - \mathbb{P}(\Omega^C) && (1.23)\\
&= 1 - \mathbb{P}\left(\bigcup_{j=0}^{n}\{|\epsilon^j| > \tau\}\right) && (1.24)\\
\mathbb{P}(\Omega) &= 1 - n.\mathbb{P}(|\epsilon| > \tau) && (1.25)\\
& && (1.26)
\end{aligned}$$

Si $\frac{\tau}{\sigma} \geq 1$ ie. $A \geq \frac{1}{\sqrt{log(n)}}$ ie. $A \geq \frac{1}{\sqrt{log(2)}}$, on peut proposer la majoration de $\mathbb{P}(|\epsilon| > \tau)$ suivante :

$$\begin{aligned}
\mathbb{P}(|\epsilon| > \tau) &= \frac{2}{\sigma\sqrt{2\pi}} \int_{\tau}^{\infty} e^{\frac{-x^2}{2\sigma^2}}\, \mathrm{d}x && (1.27)\\
&\leq \int_{\tau}^{\infty} \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}\, \mathrm{d}x && (1.28)\\
\mathbb{P}(|\epsilon| > \tau) &\leq e^{-\frac{\tau^2}{2\sigma^2}} && (1.29)
\end{aligned}$$

Avec $\tau = A\sigma\sqrt{log(n)}$ (et en assimilant $log(n)$ à $ln(n)$) :

$$\mathbb{P}(|\epsilon| > \tau) \leq n^{\frac{-A^2}{2}} \leq n^{\frac{-A^2}{8}} \tag{1.30}$$

Donc :

$$\mathbb{P}(\Omega) \geq 1 - n^{1-\frac{A^2}{8}} \tag{1.31}$$

3/ Conclusion

Le résultat (1.21) couplé au résultat (1.31) nous permet d'aboutir au résultat (1.12) demandé.