Contrôle : Statistique pour les big data

Patrick Tardivel,
Université de Bourgogne

Quelques rappels

Soit $X \in \mathbb{R}^{n\times p}$, $y \in \mathbb{R}^n$ et $\lambda > 0$, l'ensemble $S_{X,\lambda\|.\|_1}(y)$ des solutions LASSO est défini par

$$S_{X,\lambda\|.\|_1}(y) := \underset{b\in\mathbb{R}^p}{\text{Argmin}}\ \frac{1}{2}\|y - Xb\|_2^2 + \lambda\|b\|_1.$$

On rappelle que $\hat\beta \in S_{X,\lambda\|.\|_1}(y)$ si seulement si l'une des caractérisations suivantes est satisfaite

$$\|X^T(y-X\hat\beta)\|_\infty \le \lambda \text{ et } \forall i \in \text{supp}(\hat\beta), X_i^T(y-X\hat\beta) = \lambda\text{sign}(\hat\beta_i) \Leftrightarrow \|X^T(y-X\hat\beta)\|_\infty \le \lambda \text{ et } \hat\beta^T X^T(y-X\hat\beta) = \lambda\|\hat\beta\|_1.$$

Exercice 1 (questions de cours)

1. *Soit $z \in \mathbb{R}$, on considère la fonction*

$$f : t \in \mathbb{R} \mapsto \frac{1}{2}(z-t)^2 + \lambda|t|$$

En quel point la fonction f atteint son minimum ? Pour la preuve, on pourra sans perte de généralité supposer que $z > 0$.
Soit $y \in \mathbb{R}^p$ et $X \in \mathbb{R}^{n\times p}$. Lorsque $X = I_p$, rappeler (sans justifier), l'expression explicite de la solution du problème LASSO :

$$S_{I_p,\lambda\|.\|_1}(y) := \underset{b\in\mathbb{R}^n}{\text{argmin}}\ \frac{1}{2}\|y-b\|_2^2 + \lambda\|b\|_1$$

Lorsque X est une matrice telle que $X^TX = I_p$, rappeler (sans justifier), l'expression explicite de la solution du problème LASSO $S_{X,\lambda\|.\|_1}(y)$.

2. *Soit $X \in \mathbb{R}^{n\times p}$, $y \in \mathbb{R}^n$, $\lambda > 0$ et $\hat\beta$ un élément arbitraire de $S_{X,\lambda\|.\|_1}(y)$. On pose $\hat r = y - X\hat\beta$. Montrer que $\hat r$ est la solution du problème suivant*

$$\hat r = \underset{r\in\mathbb{R}^n}{\text{argmin}}\ \|y-r\|_2 \text{ sous la contrainte } \|X^T r\|_\infty \le \lambda.$$

3. *Soit $Z \in \mathbb{R}^{n\times p}$ une matrice dont les coefficients sont iid $\mathcal{N}(0,1)$ et $x^* \in \mathbb{R}^p$ tel que $\|x^*\|_0 = k$. Pour chacun des cas suivant justifier si la probabilité que x^* soit identifiable relativement à Z et la norme ℓ_1 est proche de 0 ou proche 1.*
   a) Lorsque $k = 100, n = 500$ et $p = 2000$.
   b) Lorsque $k = 600, n = 1000$ et $p = 2000$.
   Pour cette question on pourra utiliser la courbe de transition de Donoho-Tanner donnée ci-dessous :

Exercice 2 *Soit $A \in \mathbb{R}^{2\times 3}$ la matrice*

$$A := \begin{pmatrix} 1 & 0 & 2/3 \\ 0 & 1 & 2/3 \end{pmatrix}.$$

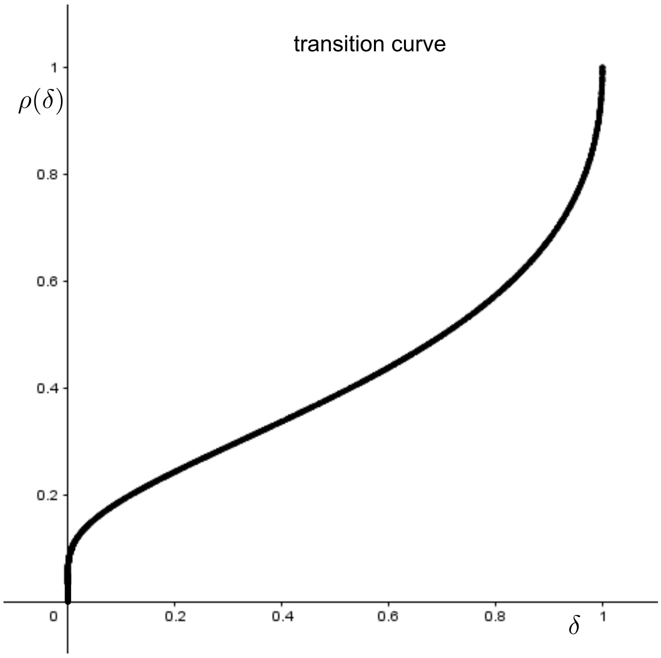

FIGURE 1 – Courbe de transition de Donoho-Tanner

1. *Tracer sur la boule unité de la norme N_A définie par*

$$\forall b \in \mathbb{R}^2, N_A(b) := \min\{\|x\|_1 : Ax = b\}.$$

2. *On considère un vecteur $x = (x_1, x_2, x_3)$ tel que $x_1 > 0, x_2 > 0$ et $x_3 = 0$. Est-ce que x est identifiable relativement à A et à la norme ℓ_1 ?*

3. *On considère un vecteur $x = (x_1, x_2, x_3)$ tel que $x_1 > 0, x_2 < 0$ et $x_3 = 0$. Est-ce que x est identifiable relativement à A et à la norme ℓ_1 ?*

Le but du problème suivant est de démontrer la Proposition 1 issue de l'article [1]. Cette proposition propose une méthode simple pour identifier certaines composantes nulles du LASSO.

Proposition 1 (Méthode de dépistage des zéros) *Soit $X = (X_1|\ldots|X_p) \in \mathbb{R}^{n\times p}$, $y \in \mathbb{R}^n$ tel que $\|X^T y\|_\infty \neq 0$, $\lambda > 0$ tel que $\lambda < \|X^T y\|_\infty$ et $\hat\beta$ une solution LASSO de $S_{X,\lambda\|\cdot\|_1}(y)$. Si l'inégalité suivante est satisfaite :*

$$|X_j^T y| < \lambda \left(1 + \frac{\|X_j\|_2\|y\|_2}{\|X^T y\|_\infty}\right) - \|X_j\|_2\|y\|_2$$

alors $\hat\beta_j = 0$.

Problème : méthode de dépistage des zéros

Soit $X = (X_1|\ldots|X_p) \in \mathbb{R}^{n\times p}$, $y \in \mathbb{R}^n$ tel que $\|X^T y\|_\infty \neq 0$, $\lambda > 0$ tel que $\lambda < \|X^T y\|_\infty$ et $\hat\beta \in S_{X,\lambda\|\cdot\|_1}(y)$. On pose $\hat r = y - X\hat\beta$ et on rappelle que

$$\hat r := \underset{r\in\mathbb{R}^n}{\operatorname{argmin}} \|y - r\|_2 \text{ sous la contrainte } \|X^T y\|_\infty \leq \lambda. \tag{1}$$

1. En utilisant le vecteur $r_0 = \lambda y/\|X^T y\|_\infty$ et l'expression (1) montrer l'inégalité suivante :

$$\|y - \hat r\|_2 \leq \left(1 - \frac{\lambda}{\|X^T y\|_\infty}\right)\|y\|_2.$$

2. Soit $a \in \mathbb{R}^n, b \in \mathbb{R}^n$ et $\gamma \geq 0$. Montrer que l'on a l'égalité suivante

$$\sup\{|a^T v| : \|b - v\|_2 \leq \gamma\} = |a^T b| + \|a\|_2\gamma$$

3. En prenant $\gamma = \left(1 - \frac{\lambda}{\|X^T y\|_\infty}\right)\|y\|_2$ dans la question 2) et en utilisant la question 1), montrer que

$$|X_j^T \hat r| \leq |X_j^T y| + \|X_j\|_2\|y\|_2\left(1 - \frac{\lambda}{\|X^T y\|_\infty}\right).$$

4. En utilisant la question 3), montrer que si

$$|X_j^T y| < \lambda \left(1 + \frac{\|X_j\|_2\|y\|_2}{\|X^T y\|_\infty}\right) - \|X_j\|_2\|y\|_2$$

alors $\hat\beta_j = 0$.

Références

[1] Laurent El Ghaoui, Vivian Viallon, and Tarek Rabbani. Safe feature elimination for the lasso and sparse supervised learning problems. *arXiv preprint arXiv :1009.4219*, 2010.