Feuille 5 : correction

Exercice 1 (Critère pour l'unicité de la solution Lasso)

1. Soient $\widehat{\beta}^{(1)}$ et $\widehat{\beta}^{(2)}$ deux solutions de $(\star)$ et soit $\widehat{\beta} = \frac{\widehat{\beta}^{(1)}+\widehat{\beta}^{(2)}}{2}$. Supposons $\mathbf{X}\widehat{\beta}^{(1)} \neq \mathbf{X}\widehat{\beta}^{(2)}$. Alors, par convexité stricte de $\|\cdot\|_2^2$, on a
$$\|Y - \widehat{\beta}\mathbf{X}\|_2^2 < \frac{1}{2}\|Y - \mathbf{X}\widehat{\beta}^{(1)}\|_2^2 + \frac{1}{2}\|Y - \mathbf{X}\widehat{\beta}^{(2)}\|_2^2\,.$$
De plus, par l'inégalité triangulaire, $\|\widehat{\beta}\|_1 \leq \frac{1}{2}\left(\|\widehat{\beta}^{(1)}\|_1 + \|\widehat{\beta}^{(2)}\|_1\right)$. On obtient donc
$$\mathcal{L}(\widehat{\beta}) < \frac{1}{2}\left(\mathcal{L}(\widehat{\beta}^{(1)}) + \mathcal{L}(\widehat{\beta}^{(2)})\right)\,,$$
ce qui est en contradiction avec l'optimalité de $\widehat{\beta}^{(1)}$ et $\widehat{\beta}^{(2)}$. Donc $\mathbf{X}\widehat{\beta}^{(1)} = \mathbf{X}\widehat{\beta}^{(2)}$.
2. On a
$$\partial\mathcal{L}(\beta) = \left\{-2\mathbf{X}^T(T - \mathbf{X}\beta) + \lambda z\,,\ z \in \partial\|\beta\|_1\right\}\,,$$
avec
$$\partial\|\beta\|_1 = \{z \in \mathbb{R}^p,\ z_j = \mathsf{sign}(\beta_j) \text{ si } \beta_j \neq 0,\ \|z\|_\infty \leq 1\}\,.$$
Par optimalité de $\widehat{\beta}$, on a $0 \in \partial\mathcal{L}(\widehat{\beta})$. Comme $\lambda > 0$, cela signifie que
$$\widehat{z} = \frac{2}{\lambda}\mathbf{X}^T(Y - \mathbf{X}\widehat{\beta}) \in \partial\|\beta\|_1\,.$$
En particulier, $\|\widehat{z}\|_\infty \leq 1$, et si $|\widehat{z}_j| < 1$, alors $\widehat{\beta}_j = 0$. On a donc $\widehat{\beta}_{J^c} = 0$. Cela implique que $\mathbf{X}\widehat{\beta} = \mathbf{X}_J\widehat{\beta}_J$, et donc que $(\mathbf{X}^T\mathbf{X}\widehat{\beta})_J = \mathbf{X}_J^T\mathbf{X}_J\widehat{\beta}_J$.
3. Sur J^c, toute solution $\widehat{\beta}$ est nulle. Et si $\mathbf{X}_J^T\mathbf{X}_J$ est inversible, alors $\widehat{\beta}_J = (\mathbf{X}_J^T\mathbf{X}_J)^{-1}\mathbf{X}_J^T\widehat{Y}$ avec $\widehat{Y} = \mathbf{X}\widehat{\beta}$.

Exercice 2 (Projection sur la boule ℓ^1)

1. Soit F_λ la fonction qui à $\alpha \in \mathbb{R}^p$ associe $F_\lambda(\alpha) = \|\beta - \alpha\|_2^2 + 2\lambda\|\alpha\|_1$. La fonction F_λ est convexe et
$$\partial F_\lambda(\alpha) = \{-2(\beta - \alpha) + 2\lambda z,\ z \in \partial\|\alpha\|_1\}\,.$$
Montrons que $0 \in \partial F_\lambda(S_\lambda(\beta))$. Si $|\beta_j| > \lambda$, alors $S_\lambda(\beta)_j = \beta_j - \mathsf{sign}(\beta_j)$ et $\mathsf{sign}(S_\lambda(\beta)_j) = \mathsf{sign}(\beta_j)$. Donc
$$-2(\beta_j - S_\lambda(\beta)_j) + 2\lambda\,\mathsf{sign}(S_\lambda(\beta)_j) = 0\,.$$
Si $|\beta_j| \leq \lambda$, alors $S_\lambda(\beta)_j = 0$. En prenant $z_j = \frac{\beta_j}{\lambda}$, on obtient
$$-2\beta_j + 2\lambda z_j = 0\,.$$
2. Par dualité langrangienne,
$$S(\beta) \in \operatorname{argmin}_{\alpha\in\mathbb{R}^p, \|\alpha\|_1\leq R}\left\{\|\beta - \alpha\|_2^2\right\} \Leftrightarrow S(\beta) \in \operatorname{argmin}_{\alpha\in\mathbb{R}^p}\left\{F_{\lambda_\star}(\alpha)\right\}\,,$$
avec $\lambda_\star$ tel que $\|S(\beta)\|_1 = R$.

3. Soit $J_\star = \max\{J \in [\![1,p]\!] \;|\beta(J)| > \lambda_\star\}$. On a

$$R = \|S_{\lambda_\star}(\beta)\|_1 = \sum_{j, |\beta_j| > \lambda_\star} |\beta_j| \left(1 - \frac{\lambda}{|\beta_j|}\right) = \sum_{j \leq J_\star} |\beta_{(j)}| - J_\star \lambda_\star \,.$$

Par définition de $J_\star$, on a $|\beta_{(J_\star+1)}| \leq \lambda_\star < |\beta_{(J_\star)}|$. Ainsi

$$R > \sum_{j \leq J_\star} |\beta_{(j)}| - J_\star |\beta_{(J_\star)}| \qquad \text{et} \qquad R \leq \sum_{j \leq J_\star+1} |\beta_{(j)}| - (J_\star+1)|\beta_{(J_\star+1)}| \,.$$

Donc

$$J_\star = \max\left\{J \in [\![1,p]\!]\,,\; \sum_{j \leq J} |\beta_{(j)}| - J|\beta_{(J)}| < R\right\}\,.$$

Et on a

$$\lambda_\star = J_\star^{-1}\left(\sum_{j \leq J_\star} |\beta_{(j)}| - R\right)\,,$$

Exercice 3 (Elastic Net)

1. Pour $\beta_j \neq 0$, on a

$$\frac{\partial}{\partial \beta_j} \mathcal{L}(\beta) = -2\mathbf{X}_j^T(Y - \mathbf{X}\beta) + 2\lambda\beta_j + \mu\,\mathsf{sign}(\beta_j)\,.$$

En utilisant que $\mathbf{X}\beta = \sum_k \mathbf{X}_k \beta_k$ et que $\mathbf{X}_j^T\mathbf{X}_j = 1$, on obtient

$$\begin{aligned}\frac{\partial}{\partial \beta_j} \mathcal{L}(\beta) &= 2\mathbf{X}_j^T\mathbf{X}_j\beta_j - 2\mathbf{X}_j^T\left(Y - \sum_{k \neq j} \mathbf{X}_k\beta_k\right) + 2\lambda\beta_j + \mu\,\mathsf{sign}(\beta_j) \\ &= 2\left((1+\lambda)\beta_j - R_j + \frac{\mu}{2}\,\mathsf{sign}(\beta_j)\right)\,.\end{aligned}$$

2. Notons $\mathcal{L}_j$ la fonction qui à β_j associe $\mathcal{L}(\beta)$. La sous-différentielle de $\mathcal{L}_j$ s'écrit

$$\partial\mathcal{L}_j(\beta_j) = \left\{2\left((1+\lambda)\beta_j - R_j + \frac{\mu}{2}z\right)\,,\; z \in \partial|\beta_j|\right\}\,.$$

Pour $\beta_j = \frac{R_j}{1+\lambda}\left(1 - \frac{\mu}{2|R_j|}\right)_+$, on a
— si $\beta_j > 0$, alors $R_j > 0$, et

$$(1+\lambda)\beta_j - R_j + \frac{\mu}{2}\,\mathsf{sign}(\beta_j) = \frac{\mu}{2}\frac{R_j}{|R_j} + \frac{\mu}{2}\,\mathsf{sign}(R_j) = 0\,.$$

— si $\beta_j = 0$, alors $2|R_j| \leq \mu$, et en prenant $z = \frac{2R_j}{\mu}$, on a

$$-R_j + \frac{\mu}{2}\frac{2R_j}{\mu} = 0\,.$$

Ainsi $0 \in \partial\mathcal{L}_j(\beta_j)$ donc β_j minimise $\mathcal{L}_j$.

3. On peut approcher $\widehat{\beta}_{\lambda,\mu}$ par descente de gradient coordonnée par coordonnée. On part d'un vecteur quelconque $\widehat{\beta}^{(0)}$, puis, pour $k = 1, \ldots, N$, on optimise $\mathcal{L}$ coordonnée par coordonnée :

$$\begin{aligned}\widehat{\beta}_1^{(k)} &\leftarrow \text{argmin}_{\beta_1}\,\mathcal{L}\left(\beta_1, \widehat{\beta}_2^{(k-1)}, \ldots, \widehat{\beta}_p^{(k-1)}\right) \\ \widehat{\beta}_2^{(k)} &\leftarrow \text{argmin}_{\beta_2}\,\mathcal{L}\left(\widehat{\beta}_1^{(k)}, \beta_2, \widehat{\beta}_3^{(k-1)}, \ldots, \widehat{\beta}_p^{(k-1)}\right) \\ &\vdots \qquad \vdots \\ \widehat{\beta}_p^{(k)} &\leftarrow \text{argmin}_{\beta_p}\,\mathcal{L}\left(\widehat{\beta}_1^{(k)}, , \ldots, \widehat{\beta}_{p-1}^{(k)}, \beta_p\right)\end{aligned}$$