

Modèle linéaire et modèles dérivés, partie 2 : examen

Exercice 1

On considère la régression logistique

$$\forall i = 1, \dots, n, Y_i \sim \mathcal{B}(1, p(X_i, \beta^*)) \text{ avec } p(X_i, \beta) = \frac{\exp(X_i \beta)}{1 + \exp(X_i \beta)} \quad \forall \beta \in \mathbb{R}^p$$

où les Y_i sont supposés indépendants, les X_i sont dans \mathbb{R}^p et fixés et on note

$$Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} X_1 \\ \dots \\ X_n \end{pmatrix}, \quad \vec{p}(\beta) = \begin{pmatrix} p(X_1, \beta) \\ \dots \\ p(X_n, \beta) \end{pmatrix}$$

On veut dans cet exercice expliciter l'algorithme "iterated reweighted least squares" pour la régression logistique pénalisée ridge.

1. Écrire la $-\log$ vraisemblance au point β associée à ces observations, notée $\ell(\beta)$.
2. Calculer le gradient et la matrice hessienne de ℓ au point β . Montrer qu'on peut récrire la matrice hessienne sous la forme $X^\top W X$, où W est une matrice diagonale à coefficients positifs que vous déterminerez.
3. Montrer que, à partir d'une valeur courante β^{old} , l'algorithme de Newton-Raphson pour la régression logistique propose l'update

$$\beta^{\text{new}} = (X^\top W X)^{-1} X^\top W z$$

$$\text{avec } z = X \beta^{\text{old}} + W^{-1} (Y - \vec{p}(\beta^{\text{old}})).$$

4. Montrer que β^{new} est la solution du problème des moindres carrés repondérés

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n W_{ii} (z_i - X_i \beta)^2.$$

5. Ecrire l'update $\beta^{\text{new,ridge}}$ à partir d'une valeur $\beta^{\text{old,ridge}}$ dans le cas de la régression logistique pénalisée ridge

$$\ell(\beta) + \frac{\lambda}{2} \|\beta\|^2.$$

6. Puis montrer qu'il s'écrit

$$(X^\top W X + \lambda \text{Id}_p)^{-1} X^\top W \left(X \beta^{\text{old,ridge}} + W^{-1} (Y - \vec{p}(\beta^{\text{old,ridge}})) \right).$$

7. En déduire le problème de moindres carrés repondérés pénalisés ridge dont il est solution.

Exercice 2

On considère un modèle linéaire

$$Y = X\beta^* + \varepsilon$$

où $Y \in \mathbb{R}^n$ est centré ($\bar{Y} = 0$) et X est de taille $n \times p$ et a ses colonnes centrées et réduites

$$\bar{X}^j = 0 \text{ et } \frac{1}{n} \|X^j\|^2 = 1,$$

$\beta^* \in \mathbb{R}^p$ est inconnu, et ε est un vecteur centré et de variance proportionnelle à l'identité. Pour tout ensemble d'indices $\mathcal{A} \subset \{1, \dots, p\}$, on note $X_{\mathcal{A}}$ la matrice qui contient les colonnes de X dont les indices sont dans \mathcal{A} . Pour tout vecteur $a = (a_1, \dots, a_k)^\top$, on définit

$$[a]_j = a_j \text{ et } \text{signe}(a) = (\text{signe}(a_1), \dots, \text{signe}(a_k))^\top.$$

On va détailler l'algorithme "least angle regression" introduit par Efron et al. en 2004 (Ann. Statist.). On introduit pour cela un chemin de solution $\beta(\lambda)$, avec $\lambda \in \mathbb{R}_+$, que l'on cherche dans cet exercice à déterminer.

1. Montrer que, sous les hypothèses de l'exercice, $\frac{1}{n} |Y^\top X^j|$ est égal, à une constante indépendante de j près, à la corrélation empirique entre X^j et Y .
2. On pose $\lambda_1 = \frac{1}{n} \max_j |Y^\top X^j|$ et $\beta(\lambda) = \vec{0}$ pour tout $\lambda \geq \lambda_1$. Montrer que, pour tout $\lambda \geq \lambda_1$ et si $X^\top X = \text{Id}_p$, $\beta(\lambda)$ est une solution du problème de lasso

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2n} \|Y - X\beta\|^2 + \lambda \|\beta\|_1.$$

On définit $\mathcal{A}_1 = \{j, |Y^\top X^j| = n\lambda_1\}$ et on suppose qu'il existe un unique $j_1 \in \{1, \dots, p\}$ tel que $\mathcal{A}_1 = \{j_1\}$. Pour $\lambda_1 > \lambda \geq \lambda_2$ (défini dans la question suivante), on pose

$$\begin{aligned} [\beta(\lambda)]_{j_1} &= (X_{\mathcal{A}_1}^\top X_{\mathcal{A}_1})^{-1} (X_{\mathcal{A}_1}^\top Y - n\lambda t_{\mathcal{A}_1}) \text{ avec } t_{\mathcal{A}_1} = \text{signe}(X_{\mathcal{A}_1}^\top Y) \text{ et} \\ [\beta(\lambda)]_k &= 0 \text{ pour } k \neq j_1. \end{aligned}$$

3. Montrer que $\lambda \mapsto \beta(\lambda)$ est continue et linéaire et que $X\beta(\lambda) = X_{\mathcal{A}_1} [\beta(\lambda)]_{j_1}$, pour $\lambda_1 \geq \lambda \geq \lambda_2$.
4. Pour $\lambda_1 > \lambda \geq \lambda_2$, que vaut

$$|(X^{j_1})^\top (Y - X\beta(\lambda))| ?$$

On fait décroître λ jusqu'à $\lambda_2 \leq \lambda_1$ de telle sorte que à cette valeur il existe un $j_2 \neq j_1$ pour lequel

$$|(X^{j_2})^\top (Y - X\beta(\lambda_2))| = n\lambda_2.$$

5. On suppose dans cette question que $(X^{j_1})^\top (Y - X\beta(\lambda_2)) \geq 0$ et $(X^{j_2})^\top (Y - X\beta(\lambda_2)) \geq 0$. Calculer λ_2 .

Pour $\lambda_2 > \lambda \geq \lambda_3$, on pose alors $\mathcal{A}_2 = \{j_1, j_2\}$ et

$$\begin{aligned} [\beta(\lambda)]_{j_1, j_2} &= (X_{\mathcal{A}_2}^\top X_{\mathcal{A}_2})^{-1} (X_{\mathcal{A}_2}^\top Y - n\lambda t_{\mathcal{A}_2}) \text{ avec } t_{\mathcal{A}_2} = \text{signe}(X_{\mathcal{A}_2}^\top (Y - X\beta(\lambda_2))) \\ [\beta(\lambda)]_k &= 0 \text{ pour } k \neq j_1, j_2. \end{aligned}$$

6. Montrer que

$$X\beta(\lambda) = X_{\mathcal{A}_2}(X_{\mathcal{A}_2}^\top X_{\mathcal{A}_2})^{-1}(X_{\mathcal{A}_2}^\top Y - n\lambda t_{\mathcal{A}_2}) = X_{\mathcal{A}_2}\hat{\beta}_{\mathcal{A}_2} - n\lambda u_{\mathcal{A}_2}$$

où $\hat{\beta}_{\mathcal{A}_2}$ est l'estimateur des moindres carrés classiques quand la matrice de design est $X_{\mathcal{A}_2}$ et $u_{\mathcal{A}_2}$ est un vecteur à déterminer. En déduire que $\lambda \mapsto \beta(\lambda)$ est linéaire pour $\lambda_2 > \lambda \geq \lambda_3$.

7. Calculer $X_{\mathcal{A}_2}^\top u_{\mathcal{A}_2}$. En déduire que $u_{\mathcal{A}_2}$ est équi-angulaire entre X^{j_1} et X^{j_2} , c'est-à-dire que les cosinus des angles qu'il forme avec ces deux vecteurs sont égaux (au signe près).

8. Montrer que

$$\|Y - X\beta(\lambda)\|^2 = \|Y - X_{\mathcal{A}_2}\hat{\beta}_{\mathcal{A}_2}\|^2 + \|n\lambda u_{\mathcal{A}_2}\|^2.$$

En déduire que les corrélations empiriques entre les résidus $Y - X\beta(\lambda)$ et les variables X^{j_1} et X^{j_2} restent égales (au signe près) et sont monotones quand λ décroît.

9. Montrer que $\lambda \mapsto \beta(\lambda)$ est continue en λ_2 .