TD 12

Les questions marquées d'un astérisque **(*)** *sont facultatives.*

Exercice 1. (Lasso slow rates) Soient n et d deux entiers naturels strictement positifs. Considérons le modèle linéaire $Y = X\beta^\star + \varepsilon$ où $Y \in \mathbb{R}^n$ et $\beta^\star \in \mathbb{R}^d$, ainsi que l'estimateur Lasso $\hat{\beta} \in \arg\min_{\beta \in \mathbb{R}^d}(\|Y - X\beta\|_2^2 + \lambda_n \|\beta\|_1)$, où λ_n est un réel strictement positif.

1. Montrer que $\|X(\hat{\beta} - \beta^\star)\|_2^2 \leqslant 2|\varepsilon^\top X(\hat{\beta} - \beta^\star)| + \lambda_n(\|\beta^\star\|_1 - \|\hat{\beta}\|_1)$.
   Indication : utiliser que $\hat{\beta}$ minimise le critère du Lasso. En particulier, il est meilleur que tous les autres β possibles.
2. Montrer que $|\varepsilon^\top X(\hat{\beta} - \beta^\star)| \leqslant \|X^\top \varepsilon\|_\infty \|\hat{\beta} - \beta^\star\|_1 \leqslant \|X^\top \varepsilon\|_\infty (\|\hat{\beta}\|_1 + \|\beta^\star\|_1)$.
3. En déduire que si $\lambda_n \geqslant 2\|X^\top \varepsilon\|_\infty$, alors $\|X(\hat{\beta} - \beta^\star)\|_2^2 \leqslant 2\lambda_n \|\beta^\star\|_1$.

Application. On suppose à présent que les n coordonnées de ε sont des variables i.i.d. de loi $\mathcal{N}(0, \sigma^2)$. On suppose qu'il existe une constante c telle que, en notant $X_{\cdot j}$ la j-ième colonne de la matrice X, on ait $\max_{1 \leqslant j \leqslant d} \|X_{\cdot j}\|_2 \leqslant c\sqrt{n}$.

4. Montrer que si $\|X\|_\infty \leqslant c$, alors la condition $\max_{1 \leqslant j \leqslant d} \|X_{\cdot j}\|_2 \leqslant c\sqrt{n}$ est satisfaite.
5. Quelle est la loi de $X_{\cdot j}^\top \varepsilon$?
6. Soit $a > 0$. Trouver t tel que $\mathbb{P}(|X_{\cdot j}^\top \varepsilon| \geqslant t) \leqslant \frac{2}{dn^a}$.
   Indication : admettez l'inégalité de concentration suivante : si $Z \sim \mathcal{N}(0, \sigma^2)$, alors pour tout $x > 0$, $\mathbb{P}(|Z| \geqslant x) \leqslant 2\exp\left(\frac{-x^2}{2\sigma^2}\right)$.
7. Toujours avec le même a, en déduire un t tel que $\mathbb{P}(\|X^\top \varepsilon\|_\infty \geqslant t) \leqslant \frac{2}{n^a}$.
8. En déduire qu'avec probabilité au moins $1 - 2/n^a$,
$$\frac{1}{n}\|X(\hat{\beta} - \beta^\star)\|_2^2 \leqslant 4\sqrt{2}\sigma c \|\beta^\star\|_1 \sqrt{\frac{\log d + a \log n}{n}}.$$

Exercice 2. Soient $\mu \in \mathbb{R}$, $\sigma > 0$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$. On observe $X_1, \ldots, X_n$ i.i.d. de loi $\mathcal{N}(\theta, 1)$ (on suppose donc la variance des X_i sachant θ connue).

1. Quelle est la loi a posteriori de θ ?
2. Calculer un estimateur du maximum a posteriori de θ.
3. Donner un intervalle de crédibilité de niveau $1 - \alpha$ de θ.

Exercice 3. (Régression linéaire et Ridge, cadre bayésien)

1. Etant donné un paramètre (aléatoire) θ suivant une loi de densité p et des observations $X = (X_1, \ldots, X_n)$ telles que conditionnellement à θ, X a pour densité $x \mapsto q(x|\theta)$, montrer que l'estimateur du maximum a posteriori de θ maximise
$$\theta \longmapsto \log q(X|\theta) + \log p(\theta).$$

Considérons le modèle suivant. Soient $a, \sigma > 0$. Comme loi *a priori*, on suppose $\beta \sim \mathcal{N}(0, a^2 I_d)$. On observe la matrice X de taille $n \times d$ et le vecteur $Y = X\beta + \varepsilon \in \mathbb{R}^n$ avec $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ indépendant de β.

2. Montrer que l'estimateur du maximum a posteriori minimise la fonction
$$\beta \longmapsto \frac{1}{\sigma^2}\|Y - X\beta\|_2^2 + \frac{1}{a^2}\|\beta\|_2^2.$$
3. En déduire que l'estimateur du maximum a posteriori correspond à l'estimateur Ridge pour un paramètre de régularisation à préciser.
4. Que se passe-t-il quand $a \to 0$? Quand $a \to \infty$? Commenter.
5. **(*)** Quelle est la loi a posteriori de β ?