Exercice : (Décomposition biais - variance) :

Montrer que :

$$\underbrace{\mathbb{E}\left[\|X\beta - X\hat{\beta}\|_2^2\right]}_{\text{Risque}} = \underbrace{\left(\|\mathbb{E}[X\hat{\beta}] - X\beta\|_2\right)^2}_{\text{biais}^2} + \underbrace{\mathbb{E}\left[\|X\hat{\beta} - \mathbb{E}[X\hat{\beta}]\|_2^2\right]}_{\text{variance}}$$

Exercice : (régression ridge) :

$$\ell(\beta) = \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$$

- Montrer que $\underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}}\ \ell(\beta) = \left(X^T X + \lambda I_p\right) X^T y$

indication :
$$\begin{cases} \dfrac{\partial}{\partial x}(x^T c) = \dfrac{\partial}{\partial x}(c^T x) = c \\[2ex] \dfrac{\partial}{\partial x}(x^T A x) = (A + A^T)\, x \end{cases}$$

- Montrer que $\left(X^T X + \lambda I_p\right) \succ 0$ (définie strictement positive)

Conclure ??

Exercice : (ridge)

Montrer que la régression ridge revient à faire une régression linéaire multiple en augmentant les lignes de X et le vecteur y. Préciser cette augmentation.

Exercice (ridge) :

On s'intéresse à la régression ridge avec une seule variable explicative et on obtient un coefficient égal à $\underline{\underline{a}}$.

On ajoute une variable $X^* = X$ la copie exacte de X.

- Déduire le coefficient de régression ridge de X et X^* en fonction de $\underline{a}$.
- Même question si on augmente X de $\underline{\underline{m}}$ colonnes identiques à X.

Exercice (convexité des moindres carrés)

$f : \mathbb{R}^p \to \mathbb{R}$ est convexe SSI

$$f[\omega a + (1-\omega) b] \leq \omega f[a] + (1-\omega) f[b]$$

pour tout $\omega \in [0,1]$ et $a, b \in \mathbb{R}^p$

1 – Montrer que $f : a \mapsto \|y - Xa\|_2^2$ est convexe

2 – " " f est strictement convexe SSI $X^T X$ est inversible.

Exercice (Point de vue Bayésien de ridge et LASSO) :

On considère le modèle linéaire

$$y = X\beta + u \quad \begin{cases} y \in \mathbb{R}^n, & X \in \mathbb{R}^{n \times p} \\ \beta \in \mathbb{R}^p \end{cases}$$

le bruit $u \sim \mathcal{N}_n[0_n, \sigma^2 I_{n \times n}]$

$$y \sim \mathcal{N}_n(X\beta, \sigma^2 I_{n \times n})$$

Cadre Bayésien : se donner une loi a priori de β, on parle de vraisemblance des données sachant le paramètre β.

$$y | \beta \sim \mathcal{N}_n(X\beta, \sigma^2 I_{n \times n}),$$

On va étudier l'estimateur du maximum a posteriori

$$\hat{\beta}_{map} \in \underset{\beta \in \mathbb{R}^p}{\operatorname{argmax}} \; \ell[\beta | y]$$

où la "vraisemblance a posteriori" $\ell[\beta|y]$ est le log de la densité de $\alpha|y$ comme une fonction de β.

On suppose que le vecteur de coefficients de régression $\underline{\beta}$ est indépendant du bruit $\underline{\varepsilon}$ et densité $\underline{g}$.

$\ell[\beta|y]$ est proportionnelle à $f_\beta(y)\, g(\beta)$

où f_β est la densité de $\mathcal{N}_n(X\beta, \sigma^2 I_{n\times n})$.

Nous allons établir l'équivalence entre l'estimateur $\hat{\beta}_{map}$ et les estimateurs fréquentistes $\hat{\beta}_{ridge}$ et $\hat{\beta}_{LASSO}$.

1. Montrer que lorsque $\beta \sim \mathcal{N}_p(0_p, \tau^2 I_{p\times p})$, $\hat{\beta}_{map}$ coïncide avec

$$\hat{\beta}_{ridge} \in \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 \right\}$$

avec $\lambda = \dfrac{\sigma^2}{\tau^2}$.

2. Montrer que lorsque β suit une loi de Laplace multivariée de densité $\dfrac{e^{-\frac{\|\beta\|_1}{\tau}}}{(2\tau)^p}$, l'estimateur $\hat{\beta}_{map}$ coïncide avec

$$\hat{\beta}_{LASSO} \in \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \text{ avec } \lambda = \frac{2\sigma^2}{\tau}.$$

Exercice (elasticnet vs LASSO) :

On considère la régression elasticnet

$$\min_{\beta} \|y - X\beta\|^2 + \lambda \left(\alpha \|\beta\|_2^2 + (1-\alpha) \|\beta\|_1 \right)$$

- Montrer que la régression elasticnet revient à faire une régression LASSO avec une version augmentée de X et y.