Exercice (Décomposition biais - variance):

$$\mathbb{E}\left[\|X\beta - X\hat{\beta}\|_2^2\right] = \mathbb{E}\left[\| \underbrace{X\beta - \mathbb{E}[X\hat{\beta}]}_{a} + \underbrace{\mathbb{E}[X\hat{\beta}] - X\hat{\beta}}_{b} \|_2^2\right]$$

$$= \mathbb{E}\Big[\|X\beta - \mathbb{E}[X\hat{\beta}]\|_2^2 + 2\langle X\beta - \mathbb{E}[X\hat{\beta}], \mathbb{E}[X\hat{\beta}] - X\hat{\beta}\rangle$$

$$+ \|\mathbb{E}[X\hat{\beta}] - X\hat{\beta}\|_2^2\Big]$$

$$= \underbrace{\mathbb{E}\left[\|X\beta - \mathbb{E}[X\hat{\beta}]\|_2^2\right]}_{\text{biais}^2} + 2\,\mathbb{E}\left[\langle X\beta - \mathbb{E}[X\hat{\beta}], \mathbb{E}[X\hat{\beta}] - X\hat{\beta}\rangle\right]$$

$$+ \underbrace{\mathbb{E}\left[\|X\hat{\beta} - \mathbb{E}[X\hat{\beta}]\|_2^2\right]}_{\text{Variance}}$$

Nous avons ici :

$$2\,\mathbb{E}\left[\langle \underset{\substack{\uparrow \\ \text{non aleatoire}}}{X\beta} - \underset{\substack{\uparrow \\ \text{non aléatoire}}}{\mathbb{E}[X\hat{\beta}]}, \mathbb{E}[X\hat{\beta}] - X\hat{\beta}\rangle\right] = 0.$$

Exercice (ridge) :

On écrit $\tilde{X} = \begin{bmatrix} X \\ \sqrt{\lambda} I_p \end{bmatrix}$ et $y = \begin{bmatrix} y \\ 0_p \end{bmatrix}$

I_p : identité $p \times p$. 0_p : $p \times 1$ de zéros

$$\tilde{X}^T X = \left[X^T \ \sqrt{\lambda} I_p \right] \begin{bmatrix} X \\ \sqrt{\lambda} I_p \end{bmatrix} = X^T X + \lambda I_p$$

et $\tilde{X}^T y = \left(X^T \ \sqrt{\lambda} I_p \right) = X^T y$.

$$\hat{\beta} = \left(\tilde{X}^T \tilde{X} + \lambda I_p \right)^{-1} \tilde{X}^T y$$
$$= \left(X^T X + \lambda I_p \right)^{-1} X^T y .$$

Exercice (ridge) :

$X \in \mathbb{R}^{n \times 1}$ et $a = \hat{\beta} = \dfrac{X^T y}{X^T X + \lambda}$.

Lorsque une copie $X^* = X$ est ajoutée, on veut résoudre

$$\min_{(\beta_1, \beta_2) \in \mathbb{R}^2} \| y - \underbrace{X^T \beta_1 - X^{*T}}_{=} \beta_2 \|_2^2 + \lambda \|\beta_1\|_2^2 + \lambda \|\beta_2\|_2^2$$

Par symétrie, on sait que $\beta_1 = \beta_2$ solution de

$$\min_{\beta} \| y - 2 X^T \beta \|_2^2 + 2\lambda \|\beta\|_2^2$$

$$\hat{\beta}_1 = \hat{\beta}_2 = \frac{X^T y}{2 X^T X + \lambda}$$

$m \geq 2$

$$\hat{\beta}_k = \frac{X^T y}{m X^T X + \lambda}, \quad k = 1, \dots, m.$$

Exercice (régression ridge) :

$$\ell(\beta) = [y - X\beta]^T [y - X\beta] + \lambda \beta^T \beta$$

$$= y^T y - \beta^T X^T y - y^T X \beta + \beta^T X^T X \beta + \lambda \beta^T \beta$$

$$= y^T y - \beta^T X^T y - \beta^T X^T y + \beta^T X^T X \beta + \beta^T [\lambda I_p] \beta$$

$$= y^T y - 2 \beta X^T y + \beta^T [X^T X + \lambda I_p] \beta$$

$$\frac{\partial \ell(\beta)}{\partial \beta} = -\cancel{2} X^T y + \cancel{2} (X^T X + \lambda I_p) \beta = 0$$

$$= [X^T X + \lambda I_p] \beta = X^T y$$

$$\Leftrightarrow \quad \hat{\beta}_{ridge} = [X^T X + \lambda I_p]^{-1} X^T y$$

Montrons que si $\lambda > 0$ alors $(X^T X + \lambda I_p) \succ 0$

Soit $a \in \mathbb{R}^p \setminus \{0_p\}$

$$a^T (X^T X + \lambda I_p) a = a X^T X a + \lambda a^T I_p a$$

$$= a X^T X a + \lambda \|a\|_2^2 > 0 \quad \text{(inversible)}.$$

Exercice (Convexité des moindres carrés) :

1) $f : a \longmapsto \|y - Xa\|_2^2 \qquad a \in \mathbb{R}^p$

Soit $w \in [0,1]$ et $a, b \in \mathbb{R}^p$

$$f[wa + (1-w)b] - w f[a] - (1-w) f[b]$$

$$= \|y - X(wa - (1-w)b)\|_2^2 - w\|y - Xa\|_2^2 - (1-w)\|y - Xb\|_2^2$$

$$= \|w(y - Xa) + (1-w)(y - Xb)\|_2^2 - w\|y - Xa\|_2^2 - (1-w)\|y - Xb\|_2^2$$

$$= \|w(y - Xa)\|_2^2 + \|(1-w)(y - Xb)\|_2^2 + 2\langle w(y - Xa), (1-w)(y - Xb)\rangle$$

$$- w\|a - Xa\|_2^2 - (1-w)\|y - Xb\|_2^2 \quad \left(\|a+b\|_2^2 = \|a\|_2^2 + \|b\|_2^2 + 2\langle a, b\rangle\right)$$

$$= w^2\|y - Xa\|_2^2 + (1-w)^2\|y - Xb\|_2^2 + 2w(1-w)\langle y - Xa, y - Xb\rangle$$

$$- w\|y - Xa\|_2^2 - (1-w)\|y - Xb\|_2^2$$

$$= w^2\left(\|y - Xa\|_2^2 + \|y - Xb\|_2^2 - 2\langle y - Xa, y - Xb\rangle\right)$$

$$- w\left(\|y - Xa\|_2^2 + \|y - Xb\|_2^2 - 2\langle y - Xa, y - Xb\rangle\right)$$

$$= w^2\|y - Xa - (y - Xb)\|_2^2 - w\|y - Xa - (y - Xb)\|_2^2$$

$$= (w^2 - w)\|Xb - Xa\|_2^2 \leq 0.$$

2) $\dfrac{\partial^2}{\partial \beta^2}\|y - X\beta\|_2^2 = 2X^TX \implies f$ est strictement convexe SSI $X^TX \succ 0$.

Exercice (point de vue Bayésien de Ridge et LASSO)

Ce n'est que du calcul:

1. $$f_\beta(y) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\|y - X\beta\|_2^2}{2\sigma^2}}$$

et $$g(\beta) = \frac{1}{(2\pi\tau^2)^{p/2}} e^{-\frac{\|\beta\|_2^2}{2\tau^2}}$$

Nous avons:

$$\log\left[f_\beta[y]\, g[\beta]\right] = \log\left[\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\|y - X\alpha\|_2^2}{2\sigma^2}} \times \frac{e^{-\frac{\|\beta\|_2^2}{2\tau^2}}}{(2\pi\tau^2)^{p/2}}\right]$$

$$= -\frac{n}{2}\log\left[2\pi\sigma^2\right] - \frac{1}{2\sigma^2}\|y - X\beta\|_2^2 - \frac{p}{2}\log\left[2\pi\tau^2\right] - \frac{1}{2\tau^2}\|\beta\|_2^2.$$

Alors: $$\underset{\beta\in\mathbb{R}^p}{\operatorname{argmax}}\ l[\beta|y] = \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}}\left\{\frac{n}{2}\log\left[2\pi\sigma^2\right] + \frac{\|y - X\beta\|_2^2}{2\sigma^2} + \frac{p}{2}\log\left(2\pi\tau^2\right) + \frac{1}{2\tau^2}\|\beta\|_2^2\right\}$$

$$= \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}}\left\{\frac{\|y - X\beta\|_2^2}{2\sigma^2} + \frac{1}{2\tau^2}\|\beta\|_2^2\right\}$$

$$= \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}}\left\{\|y - X\beta\|_2^2 + \frac{\sigma^2}{\tau^2}\|\beta\|_2^2\right\}$$

2) On pose $$g[\beta] = \frac{1}{(2\tau)^p} e^{-\frac{\|\beta\|_1}{\tau}}$$

$$\log\left[f_\beta[y]\, g[\beta]\right] = \log\left[\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\|y - X\beta\|_2^2}{2\sigma^2}} \times \frac{e^{-\frac{\|\beta\|_1}{\tau}}}{(2\tau)^p}\right]$$

$$= \frac{-n}{2}\log[2\pi\sigma^2] - \frac{1}{2\sigma^2}\|y - X\beta\|_2^2 - p\log[2\tau] - \frac{1}{\tau}\|\beta\|_1$$

$$\underset{\beta\in\mathbb{R}^p}{\operatorname{argmax}}\, \ell[\beta|y] = \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}}\left\{\frac{n}{2}\log(2\pi\sigma^2) + \frac{\|y - X\beta\|_2^2}{2\sigma^2} + p\log[2\tau] + \frac{\|\beta\|_1}{\tau}\right\}$$

$$= \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}}\left\{\frac{\|y - X\beta\|_2^2}{2\sigma^2} + \frac{\|\beta\|_1}{\tau}\right\}$$

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

3. En général, nous avons :

$$\hat{\beta}_{map} \in \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}}\left\{\|y - X\beta\|_2^2 - 2\sigma^2\log g[\beta]\right\}$$

Exercice (elasticnet vs LASSO) :

On suppose que $y \in \mathbb{R}^{n\times 1}$ et $X \in \mathbb{R}^{n\times(p+1)}$ et $\beta \in \mathbb{R}^{(p+1)\times 1}$.

On augmente X par :

$$\tilde{X} = \begin{pmatrix} X \\ \\ \gamma I_{p+1} \end{pmatrix} \in \mathbb{R}^{(n+p+1)\times(p+1)}$$

avec $\gamma > 0$.

$$\tilde{y} = \begin{pmatrix} y \\ \\ 0_{p+1} \end{pmatrix} \in \mathbb{R}^{(n+p+1)\times 1}$$

Nous avons donc :

$$\|\tilde{y} - \tilde{X}\beta\|_2^2 = \left\| \begin{pmatrix} y - X\beta \\ \\ \gamma\beta \end{pmatrix} \right\|_2^2 = \|y - X\beta\|_2^2 + \gamma^2 \|\beta\|_2^2$$

On considère la régression LASSO pour $(\tilde{y}, \tilde{X})$

$$\min_\beta \|\tilde{y} - \tilde{X}\beta\|_2^2 + \delta\|\beta\|_1 = \min_\beta \|y - X\beta\|_2^2 + \gamma\|\beta\|_2^2 + \delta\|\beta\|_1$$

On choisit $\gamma = \sqrt{\lambda\alpha}$ et $\delta = \lambda(1-\alpha)$, on obtient le problème elasticnet de départ.