BLANCHARD ITA205 TP3

Théau

OLS

- Par hypothèse $\tilde{\beta}$ est non biaisé donc

$\beta = \mathbb{E}(\tilde{\beta})$

$= \mathbb{E}(Hy) + \mathbb{E}(Dy)$

$= \mathbb{E}(\beta^*) + D\mathbb{E}(y)$, or β^* est aussi non biaisé.

De plus dans le cadre du modèle gaussien $y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$, $\sigma^2 > 0$.

D'où $DX\beta = 0$ on a vu alors $\underline{DX = 0}$.

* $Var(\tilde{\beta}) = Var(Cy) = \sigma^2 CC^T$

$= \sigma^2 (H+D)(H^T + D^T)$

$= \sigma^2 HH^T + \sigma^2 DD^T + \sigma^2 HD^T + \sigma^2 DH^T$

Or : $\circ\ DH^T = D X (X^TX)^{-1} = 0$

$\circ\ H^TD = (DH^T)^T = 0^T = 0$

$\circ\ Var(\beta^*) = \sigma^2 (X^TX)^{-1} X^T X (X^TX)^{-1} = \sigma^2 (X^TX)^{-1} = \sigma^2 HH^T$

D'où : $\underline{Var(\tilde{\beta}) = Var(\beta^*) + \sigma^2 DD^T}$

Ainsi $Var(\tilde{\beta}) - Var(\beta^*) = \sigma^2 DD^T$

On $\forall z \in \mathbb{R}^d \setminus \{0\}$, $z^T DD^T z = \|D^T z\|^2 > 0$ car $D \neq 0$

D'où $\underline{Var(\tilde{\beta}) > Var(\beta^*)}$

On a ici supposé qu'on se plaçait dans le cas du modèle gaussien tel que (X^TX) soit aussi inversible.

Ridge regression

On pose pour $\beta \in \mathbb{R}^d$, $f(\beta) = \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$

Alors $\beta_{ridge}^{\infty} = \underset{\beta}{\operatorname{argmin}}\, f(\beta)$.

On a alors $f'(\beta) = -2x_c^T(y_c - x_c\beta) + 2\lambda\beta$

$f''(\beta) = 2x^T x_c + 2\lambda \geq 0$.

Donc β^*_{ridge} est solution de $f'(\beta) = 0$.

D'où $\underline{\beta^*_{ridge} = (x_c^T x_c + \lambda I_m)^{-1} x_c^T y_c}$

- Alors $\mathbb{E}(\beta^*_{ridge}) = (x_c^T x_c + \lambda I_m)^{-1} x_c^T x_c \beta \neq \beta$.

The ridge estimator is biased.

- Soit (μ_i) les valeurs singulières de x_c : $D = diag(\mu_i)$.

On a $x_c^T x_c = VDU^T UDV^T = VD^T DV^T$ car U est orthogonale

D'où $\beta^*_{ridge} = (VD^T DV^T + \lambda VV^T)^{-1} x_c^T y_c$ car V est aussi orthogonale.

$= V(D^T D + \lambda I_m)^{-1} V^T x_c^T y_c$

$= V\, diag(\mu_i^2 + \lambda)^{-1} V^T x_c^T y_c$.

$= V\, diag\left(\frac{1}{\mu_i^2 + \lambda}\right) V^T x_c^T y_c$.

L'avantage d'utiliser la décomposition en valeurs singulières est qu'au lieu de devoir inverser une matrice compliquée, il suffit d'inverser des coefficients.

- $Var(\beta^*_{ridge}) = \sigma^2 V(D^T D + \lambda I_d)^{-1} V^T x_c^T x_c V(D^T D + \lambda I_d)^{-1} V^T$

$= \sigma^2 V(D^T D + \lambda I_d)^{-1} D^T D (D^T D + \lambda I_d)^{-1} V^T$

$= \sigma^2 V\, diag\left(\frac{1}{\mu_i^2 + \lambda}\right) diag(\mu_i^2)\, diag\left(\frac{1}{\mu_i^2 + \lambda}\right) V^T$

$= \sigma^2 V\, diag\left(\frac{\mu_i^2}{(\mu_i^2 + \lambda)^2}\right) V^T$

Et $Var(\beta^*_{ols}) = \sigma^2 V\, diag\left(\frac{1}{\mu_i^2}\right) V^T$

Donc $Var(\beta^*_{ols}) - Var(\beta^*_{ridge}) = \sigma^2 V diag\left(\frac{1}{\mu_i^2} - \frac{\mu_i^2}{(\mu_i^2+\lambda)^2}\right) V^T$

Or $\forall i,\ \frac{1}{\mu_i^2} - \frac{\mu_i^2}{(\mu_i^2+\lambda)^2} = \frac{1}{\mu_i^2}\left(1 - \underbrace{\frac{1}{(1+\lambda)^2}}_{<1}\right) \geq 0$

D'où le résultat $\underline{Var(\beta^*_{ridge}) < Var(\beta^*_{ols})}$

- On a $\mathbb{E}(\beta^*_{ridge}) = V diag\left(\frac{\mu_i^2}{\mu_i^2+\lambda}\right) V^T\beta$ et $Var(\beta^*_{ridge}) = \sigma^2 V diag\left(\frac{\mu_i^2}{(\mu_i^2+\lambda)^2}\right) V^T$

Donc si λ augmente alors le biais augmente mais la variance diminue et inversement si λ diminue le biais diminue mais la variance augmente.

- Si $x_c^T x = Id$ alors on a $\beta^*_{ols} = x_c^T y_c$ et

$\beta^*_{ridge_c} = ((1+\lambda)Id)^{-1} x_c^T y_c = \underline{\frac{\beta^*_{ols}}{1+\lambda}}$

Elastic net

On pose $g(\beta) = \|y_c - x_c\beta\|_2^2 + \lambda_2\|\beta\|_2^2 + \lambda_1\|\beta\|_1$

Selon le signe de β_i, le nombre gradient de $\beta_i \mapsto |\beta_i|$ vaut 1 ou -1

d'où on peut noter :

$g'(\beta) = -2x_c^T(y_c - x_c\beta) + 2\lambda_2\beta \pm \lambda_1$

Et $g''(\beta) = 2x_c^T x + 2\lambda_2 > 0$

Ainsi β^*_{ElNet} est solution de $g'(\beta) = 0$

Dou $\beta^*_{ElNet} = (x_c^T x_c + \lambda_2 Id)^{-1}\left(x_c^T y_c \pm \frac{\lambda_1}{2}\right)$

Donc avec $x_c^T x_c = Id$ on a bien

$$\underline{\beta^*_{ElNet} = \frac{\beta^*_{ols} \pm \frac{\lambda_1}{2}}{1+\lambda_2}}$$