TD 11 - mercredi

mercredi 21 avril 2021 08:53

Exercice 1

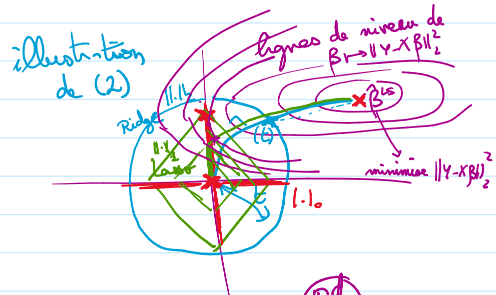
$$\hat\beta_1(\lambda) \in \underset{\beta\in\mathbb{R}^d}{\operatorname{argmin}} \{\| Y - X\beta\|_2^2 + \lambda h(\beta)\} \quad (1)$$

$$\hat\beta_2(t) \in \underset{\beta\in\mathbb{R}^d :\, h(\beta)\le t}{\operatorname{argmin}} \| Y - X\beta\|_2^2 \quad (2)$$

h peut être soit $\beta \mapsto \|\beta\|_2^2$
soit $\beta \mapsto \|\beta\|_1$
soit $\beta \mapsto |\beta|_0$

Soit $\lambda \geq 0$, posons $\boxed{t = h(\hat\beta_1(\lambda))}$.

illustration de (2)



1) Par définition, $h(\hat\beta_2(t)) \leq t = h(\hat\beta_1(\lambda))$

Et comme $h(\hat\beta_1(\lambda)) \leq t$,

$$\| Y - X\hat\beta_2(t)\|_2^2 \leq \| Y - X\hat\beta_1(\lambda)\|_2^2 .$$

2) Par conséquent, $\| Y - X\hat\beta_2(t)\|_2^2 + \lambda h(\hat\beta_2(t)) \leq \| Y - X\hat\beta_1(\lambda)\|_2^2 + \lambda h(\hat\beta_1(\lambda))$

or $\hat\beta_1(\lambda)$ est un minimiseur de $\beta \mapsto \| Y - X\beta\|_2^2 + \lambda h(\beta)$

$\mathbb{R}^d$

$\|Y - X\beta\|_2^2 = \|Y - X\hat\beta^{LS}\|_2^2 + \|X(\beta - \hat\beta^{LS})\|_2^2$

donc $\| Y - X\hat\beta_2(t)\|_2^2 + \lambda h(\hat\beta_2(t)) = \| Y - X\hat\beta_1(\lambda)\|_2^2 + \lambda h(\hat\beta_1(\lambda))$

donc $\hat\beta_2(t)$ est aussi un minimiseur de , donc est solution de (1).

3) D'après 2), $\| Y - X\hat\beta_2(t)\|_2^2 + \lambda h(\hat\beta_2(t)) = \| Y - X\hat\beta_1(\lambda)\|_2^2 + \lambda h(\hat\beta_1(\lambda))$

D'après 1), $h(\hat\beta_2(t)) \leq h(\hat\beta_1(\lambda))$

$\| Y - X\hat\beta_2(t)\|_2^2 \leq \| Y - X\hat\beta_1(\lambda)\|_2^2$

donc

$$\| Y - X\hat\beta_2(t)\|_2^2 = \| Y - X\hat\beta_1(\lambda)\|_2^2 \quad \text{et} \quad h(\hat\beta_2(t)) = h(\hat\beta_1(\lambda))$$

donc

$\hat\beta_1(\lambda)$ est une solution de (2).

Bilan : Les solutions de (1) sont solution de (2) et vice-versa, pourvu que t soit bien choisi.

Exercice 2

1) a) $\hat\beta^{LS}$ minimise $\beta \mapsto \| Y - X\beta\|_2^2 = (Y - X\beta)^T(Y - X\beta) = Y^TY - Y^T(X\beta) - (X\beta)^TY + (X\beta)^T(X\beta)$

$$= \|Y\|_2^2 - 2Y^TX\beta + \beta^TX^TX\beta$$

Le gradient de cette fonction en β est $-2Y^TX + 2\beta^TX^TX = -2(Y - X\beta)^TX$

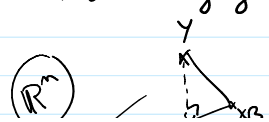
Comme $\hat\beta^{LS}$ minimise cette fonction et que cette fonction est C^1 sur $\mathbb{R}^d$, le gradient s'annule en $\hat\beta^{LS}$, c-à-d :

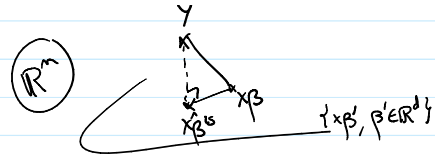
$$\boxed{(Y - X\hat\beta^{LS})^TX = 0.}$$

b) $\| Y - X\beta\|_2^2 = \| Y - X\hat\beta^{LS} + X\hat\beta^{LS} - X\beta\|_2^2 = \| Y - X\hat\beta^{LS}\|_2^2 + \| X\hat\beta^{LS} - X\beta\|_2^2 + \cancel{2 < Y - X\hat\beta^{LS}, X\hat\beta^{LS} - X\beta>}$

or $< Y - X\hat\beta^{LS}, X\hat\beta^{LS} - X\beta> = \underbrace{(Y - X\hat\beta^{LS})^TX}_{=0 \ (1a)}(\hat\beta^{LS} - \beta) = 0.$

On vient de remontrer une version du théorème de Pythagore : $X\hat\beta^{LS}$ est le projeté orthogonal de Y sur $\{X\beta, \beta\in\mathbb{R}^d\}$





c) * si $f:\mathbb{R}^n\to\mathbb{R}$ est convexe, $Y\in\mathbb{R}^n$ et $X\in\mathbb{R}^{n\times d}$, alors $\beta\mapsto f(Y-X\beta)$ est aussi convexe

$$\Big(\ f(Y-X(t\beta_1+(1-t)\beta_2)) = f(\ t(Y-X\beta_1)+(1-t)(Y-t\beta_2))\ \Big)$$

* Si $f:\mathbb{R}^d\to\mathbb{R}_+$ est convexe alors $\beta\mapsto f(\beta)^2$ est aussi convexe.

$$f(t\beta_1+(1-t)\beta_2)^2 \underset{f \text{ convexe, } x\mapsto x^2 \text{ croissante sur } \mathbb{R}_+}{\leq} (tf(\beta_1)+(1-t)f(\beta_2))^2 \underset{\text{convexité de } x\mapsto x^2}{\leq} t\,f(\beta_1)^2+(1-t)\,f(\beta_2)^2$$

* $\beta\mapsto\|\beta\|_2$ est convexe : $\|t\beta_1+(1-t)\beta_2\| \underset{\text{I.T.}}{\leq} \|t\beta_1\|+\|(1-t)\beta_2\| \underset{\text{norme}}{=} t\|\beta_1\|+(1-t)\|\beta_2\|$.

Donc $\beta\mapsto\|\beta\|_2^2$ est convexe
puis $\beta\mapsto\|Y-X\beta\|_2^2$ est convexe.

Indications : - Si f,g convexes, alors $f+g$ convexe
Si f strictement convexe et g convexe, alors $f+g$ strictement convexe) écrire la définition

- Si f continue, strictement convexe et $f(x)\underset{\|x\|\to+\infty}{\longrightarrow}+\infty$ $\quad f:\mathbb{R}^n\to\mathbb{R}$

$$\{x\in\mathbb{R}^n:\ f(x)\leq \underbrace{\inf_{x'\in\mathbb{R}^n} f(x')+1}_{\text{ou } 0 \text{ si } \inf_{x'\in\mathbb{R}^n} f(x')=-\infty}\}$$ est borné (car $f(x)\underset{\|x\|\to+\infty}{\longrightarrow}+\infty$)

c'est aussi un fermé (car f continue) donc c'est un compact.
Comme f continue, f atteint son minimum sur cet ensemble, et par définition de cet ensemble c'est un minimum global (sur $\mathbb{R}^n$).

Si ce minimum est atteint en x et x' avec $x\neq x'$, alors pour tout $t\in(0,1)$,
$$f(tx+(1-t)x') \underset{\text{strict. convexité}}{<} tf(x)+(1-t)f(x') = \inf_{x''\in\mathbb{R}^n} f(x'') \leq f(tx+(1-t)x')$$ impossible.

Donc le minimiseur est unique.

- Si $A\in\mathbb{R}^{d\times d}$ est inversible et $f:\mathbb{R}^d\to\mathbb{R}$ strictement convexe, alors $\beta\mapsto f(A\beta)$ est strictement convexe.

$\beta\mapsto\|\beta\|_2^2$ est strictement convexe.
Montrer que $\beta\mapsto\|Y-X\beta\|_2^2$ est strictement convexe revient à montrer que $\beta\mapsto\|X(\beta-\hat\beta^{LS})\|_2^2$ est strictement convexe (par 1.b)

donc à montrer que $\beta\mapsto\|X\beta\|_2^2$ est strictement convexe
$= \beta^T X^T X\beta$

Décomposition en valeurs singulières de X: $\underset{X\in\mathbb{R}^{n\times d}}{X} = \underset{n\times n}{U^T}\Sigma \underset{d\times d}{V}$ donc $X^TX = V^T\begin{pmatrix}\sigma_1^2 & & 0\\ & \ddots & \\ 0 & & \sigma_d^2\end{pmatrix}V$
(σ_i i-ème valeur singulière de X)

$\|X\beta\|^2 = \beta^TX^TX\beta = \beta^TV^T\begin{pmatrix}\sigma_1^2 & 0\\ 0 & \ddots\end{pmatrix}V\beta = \left\|\begin{pmatrix}\sigma_1 & 0\\ 0 & \ddots\end{pmatrix}V\beta\right\|^2$

$X \in \mathbb{R}^{n\times d}$ $\quad {}_n\boxed{X}^{\,d} \quad {}^{n\times n} \quad {}^{d\times d}$

$$\|X\beta\|_2^2 = \beta^T X^T X \beta = \beta^T V^T \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{pmatrix} V\beta = \left\| \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_d \end{pmatrix} V\beta \right\|_2^2$$

(σ_i i-ème valeur singulière de X)

inversible dès que $\sigma_i \neq 0$ pour tout i, c.à.d. X^TX inversible / définie positive

Donc $\beta \mapsto \|Y - X\beta\|_2^2$ strictement convexe dès que X^TX définie positive.

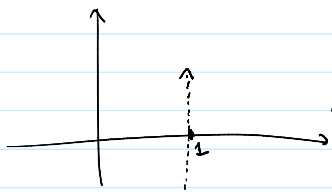
2) Comme $\beta \mapsto \|\beta\|_2^2$ est strictement convexe, $\beta \mapsto \lambda\|\beta\|_2^2$ est strictement convexe dès que $\lambda > 0$

et $\beta \mapsto \|Y - X\beta\|_2^2$ est convexe, donc $\beta \mapsto \|Y - X\beta\|_2^2 + \lambda\|\beta\|_2^2$ est strictement convexe (et continue), donc son minimum est atteint en un <u>unique point</u> : l'estimateur Ridge.

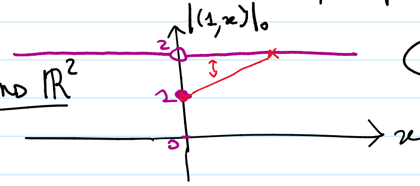
3) Lorsque X^TX est définie positive, $\beta \mapsto \|Y - X\beta\|_2^2$ est strictement convexe.
$\beta \mapsto \lambda\|\beta\|_1$ est convexe

donc $\beta \mapsto \|Y - X\beta\|_2^2 + \lambda\|\beta\|_1$ est strictement convexe (et continue) donc son minimum est atteint en un <u>unique point</u> : l'estimateur Lasso.

4)



<u>Dans $\mathbb{R}^2$</u>



pas convexe

donc <u>$\beta \mapsto |\beta|_0$ n'est pas convexe.</u>