TD N° 2 : Optimisation sous contraintes

Exercice 1. Résoudre le problème suivant :

$$\begin{aligned}\min_{(x_1,x_2,x_3)\in\mathbb{R}^3} & \tfrac{1}{2}\left[(x_1-2)^2+(x_2-2)^2+(x_3-2)^2\right] \\ \text{s.c.} \quad & x_1+x_2+x_3=1\end{aligned} \tag{1}$$

Exercice 2.

On prend $y \in \mathbb{R}^n, X \in \mathbb{R}^{n\times p}$ avec $\mathrm{rg}(X)=p$, $R\in\mathbb{R}^{q\times p}$ avec $\mathrm{rg}(R)=q$ et enfin $r\in\mathbb{R}^q$. On définit les moindres carrées sous contraintes de la manière suivante :

$$\begin{aligned}\widehat{\beta}^c \in \mathop{\arg\min}_{\beta\in\mathbb{R}^p} & \tfrac{1}{2}\left\|y-X\beta\right\|^2 \\ \text{s.c.} \quad & R\beta=r\end{aligned}$$

1) Montrer qu'avec $\hat{\beta}^{\mathrm{LS}}$ solution des moindres carrés (non contraints), on obtient :

$$\widehat{\beta}^c = \widehat{\beta}^{\mathrm{LS}} + \left(X^\top X\right)^{-1} R^\top \left[R\left(X^\top X\right)^{-1}R^\top\right]^{-1}(r-R\widehat{\beta}^{\mathrm{LS}})$$

2) Montrer de plus que sous l'hypothèse gaussienne de modèle de régression linéaire classique :

$$\frac{1}{q\widehat{\sigma^2}}(R(\widehat{\beta}^{\mathrm{LS}}-\beta^\star))^\top\left[R\left(X^\top X\right)^{-1}R^\top\right]^{-1}R(\widehat{\beta}^{\mathrm{LS}}-\beta^\star)\sim\mathcal{F}^q_{n-p}$$

où $\mathcal{F}^q_{n-p}$ est une loi de Fisher.

3) Montrer que $(R\widehat{\beta}^{\mathrm{LS}}-r)^\top\left[R\left(X^\top X\right)^{-1}R^\top\right]^{-1}(R\widehat{\beta}^{\mathrm{LS}}-r) = \left\|\hat{y}^{\mathrm{LS}}-\hat{y}^c\right\|^2$, et que $\left\|\hat{y}^{\mathrm{LS}}-\hat{y}^c\right\|^2 = \left\|y-\hat{y}^c\right\|^2 - \left\|\hat{y}^{\mathrm{LS}}-y\right\|^2$ où $\hat{y}^{\mathrm{LS}} = X\hat{\beta}^{\mathrm{LS}}$ et $\hat{y}^c = X\hat{\beta}^c$.

4) Prendre $X=[\mathbb{1}_{C_1},\dots,\mathbb{1}_{C_K}]$, et en déduire un test de l'hypothèse "$\mu_1=\cdots=\mu_K$" dans l'ANOVA à un facteur.

Exercice 3. Soit $X=\left[\frac{\mathbb{1}_{C_1}}{\sqrt{n_1}},\dots,\frac{\mathbb{1}_{C_K}}{\sqrt{n_K}}\right]$ (avec $\mathbf{x}_k=\frac{\mathbb{1}_{C_k}}{\sqrt{n_k}}$), avec $\hat{\mu}_k=\dfrac{1}{n_k}\sum_{i\in C_k} y_i$ et $n_k=\#\{i\in C_k, i\in [\![1,n]\!]\}$ avec $n_1+\cdots+n_K=n$, et on suppose que les C_k forment une partition de $[\![1,n]\!]$.

On propose l'estimateur suivant :

$$(\hat{\beta}_0,\hat{\beta}_1,\dots,\hat{\beta}_K)^\top = \mathop{\arg\min}_{\beta_0,\dots,\beta_K}\frac{1}{2}\left\|\mathbf{y}-\beta_0\mathbb{1}_n-\sum_{j=1}^K\beta_j\mathbf{x}_j\right\|^2+\frac{\lambda}{2}\sum_{j=1}^K\beta_j^2$$

1) Donner la valeur de $X^\top X$ et de $X^\top y$.
2) Donner une formule explicite pour l'estimateur Ridge, en fonction de y, $\hat{\mu}_1,\dots,\hat{\mu}_K$, et des $n_1,\dots,n_k$.
3) Comparer avec l'estimateur Ridge classique obtenu en forçant la contrainte $\beta_0=0$.

Exercice 4. On observe K classes $C_1,\dots,C_K$ et n observations d'un phénomène (*e.g.*, le rendement de la variété) sont consignées. On fait l'hypothèse que les classes C_k sont disjointes et forment une partition des observations : $\cup_{k=1}^K C_k = [\![1,n]\!]$ et $\forall (k,k') \in [\![1,K]\!], C_k \cap C_{k'} = \emptyset$. Enfin on suppose que la cardinalité de chaque classe C_k est n_k, et donc que $n = \sum_{k=1}^K n_k$.

Le modèle linéaire associé peut s'écrire de la façon suivante : on définit les α_k, coefficients qui correspondent au niveau d'influence de la k^{e} classe

$$y = \underbrace{\begin{bmatrix} \mathbb{1}_n & \mathbb{1}_{C_1} & \dots & \mathbb{1}_{C_K} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_K \end{bmatrix}}_{\beta} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_K \end{bmatrix} .$$

Donner alors l'estimateur des moindres carrés sous contraintes associé :

$$\min_{\beta \in \mathbb{R}^{K+1}} \frac{1}{2} \|y - X\beta\|^2 \quad ,$$
$$\text{t.q. } \beta = (\mu, \alpha_1, \dots, \alpha_K)^\top \text{ et } \sum_{k=1}^K c_k \alpha_k = 0 \quad ,$$

où le vecteur $c = (c_1, \dots, c_K)^\top \in \mathbb{R}^K$ est un vecteur encodant les contraintes choisies tel que $\sum_{k=1}^K c_k \neq 0$.

Exercice 5. Soit $X = \left[\frac{\mathbb{1}_{C_1}}{\sqrt{n_1}}, \dots, \frac{\mathbb{1}_{C_K}}{\sqrt{n_K}}\right]$ (avec $\mathbf{x}_k = \frac{\mathbb{1}_{C_j}}{\sqrt{n_k}}$), avec $\hat{\mu}_k = \frac{1}{n_k} \sum_{i \in C_k} y_i$ et $n_k = \#\{i \in C_k, i \in [\![1,n]\!]\}$ avec $n_1 + \dots + n_K = n$.

L'estimateur Lasso (sans pénalité sur la constante) est solution de

$$(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_K)^\top = \operatorname*{arg\,min}_{\beta_0, \dots, \beta_K} \frac{1}{2} \left\| \mathbf{y} - \beta_0 \mathbb{1}_n - \sum_{j=1}^K \beta_j \mathbf{x}_j \right\|_2^2 + \lambda \sum_{j=1}^K |\beta_j|$$

1) Donner la valeur de $X^\top X$ et de $X^\top y$
2) Donner une formule explicite pour l'estimateur Lasso, en fonction de y, $\hat{\mu}_1, \dots, \hat{\mu}_K$, et des $n_1, \dots, n_k$.