Modèles linéaires

Exercice 1: estimateurs linéaires sans biais et moindres carrés On souhaite estimer un paramètre dont la véritable valeur est notée $\overline{\boldsymbol{\theta}} \in \mathbb{R}^n$. On dispose pour cela d'observations $\mathbf{y} \in \mathbb{R}^m$ dont on suppose qu'elles sont données par le modèle

$$\mathbf{y} = \mathbf{A}\overline{\boldsymbol{\theta}} + \mathbf{b}$$

où $\mathbf{A} \in \mathbb{R}^{m\times n}$ $(m \geq n)$ est une matrice fixée de rang colonne plein ($\operatorname{rank} \mathbf{A} = n$) et $\mathbf{b}$ est un bruit aléatoire. On supposera $\mathbb{E}\{\mathbf{b}\} = \mathbf{0}$ et $\mathbb{C}\mathrm{ov}\{\mathbf{b}\} = \mathbb{E}\{\mathbf{b}\mathbf{b}^\top\} = \sigma^2\mathbf{Id}_m$. Enfin, par définition, pour des matrices symétriques $\mathbf{M}$ et $\mathbf{N}$, on notera $\mathbf{N} \preceq \mathbf{M}$ lorsque $\mathbf{M} - \mathbf{N}$ est semi-définie positive.

1. Soit l'estimateur des moindres carrés $\hat{\boldsymbol{\theta}}_{\mathrm{ls}} = \operatorname{Arg\,min}_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2$. Retrouver l'expression analytique de $\hat{\boldsymbol{\theta}}_{\mathrm{ls}}$ et montrer que c'est un estimateur linéaire (comme fonction du vecteur d'observations $\mathbf{y}$).
2. Calculer $\mathbb{E}\{\hat{\boldsymbol{\theta}}_{\mathrm{ls}}\}$ et montrer que $\hat{\boldsymbol{\theta}}_{\mathrm{ls}}$ est non biaisé.
3. Calculer la matrice de covariance de $\hat{\boldsymbol{\theta}}_{\mathrm{ls}}$.
4. Soit $\hat{\boldsymbol{\theta}}$ un autre estimateur. Montrer que si $\hat{\boldsymbol{\theta}}$ est linéaire et non biaisé, alors $\hat{\boldsymbol{\theta}} = \mathbf{B}\mathbf{y}$ où $\mathbf{B} \in \mathbb{R}^{n\times m}$ vérifie $\mathbf{BA} = \mathbf{Id}_n$. Calculer $\mathbb{C}\mathrm{ov}\{\hat{\boldsymbol{\theta}}\}$.
5. Montrer que si $\mathbf{BA} = \mathbf{Id}_n$, alors $\mathbf{BB}^\top \succeq (\mathbf{A}^\top\mathbf{A})^{-1}$.
6. Déduire de ce qui précède que si $\hat{\boldsymbol{\theta}}$ est un estimateur linéaire et non biaisé, alors :
   - $\mathbb{C}\mathrm{ov}\{\hat{\boldsymbol{\theta}}\} \succeq \mathbb{C}\mathrm{ov}\{\hat{\boldsymbol{\theta}}_{\mathrm{ls}}\}$,
   - $\mathbb{E}\{\|\hat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\|_2^2\} = \mathbb{V}\mathrm{ar}\{\hat{\boldsymbol{\theta}}\} \geq \mathbb{E}\{\|\hat{\boldsymbol{\theta}}_{\mathrm{ls}} - \overline{\boldsymbol{\theta}}\|_2^2\} = \mathbb{V}\mathrm{ar}\{\hat{\boldsymbol{\theta}}_{\mathrm{ls}}\}$,
   - pour tout $\mathbf{c} \in \mathbb{R}^n$, $\mathbb{E}\{(\mathbf{c}^\top\hat{\boldsymbol{\theta}} - \mathbf{c}^\top\overline{\boldsymbol{\theta}})^2\} \geq \mathbb{E}\{(\mathbf{c}^\top\hat{\boldsymbol{\theta}}_{\mathrm{ls}} - \mathbf{c}^\top\overline{\boldsymbol{\theta}})^2\}$.
7. Les propriétés précédentes sont connues sous le nom de «estimateur BLUE (Best Linear Unbiased Estimator)». Justifier que chercher l'estimateur BLUE revient à résoudre le problème d'optimisation suivant :

   $$\begin{aligned} &\text{min. } \mathbf{BB}^\top \quad \text{(par rapport à l'ordre } \preceq\text{)} \\ &\text{s.t. } \mathbf{BA} = \mathbf{Id}_n \end{aligned}$$

   Montrer qu'il s'agit d'un problème d'optimisation convexe (généralisé).
8. On suppose dorénavant que le bruit est gaussien $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{Id}_m)$.
   (a) Préciser la loi de $\hat{\boldsymbol{\theta}}_{\mathrm{ls}}$.
   (b) Calculer $\hat{\boldsymbol{\theta}}_{\mathrm{ml}}$ l'estimateur du maximum de vraisemblance.

Exercice 2: «least-squares» et «ridge» On souhaite estimer un paramètre dont la véritable valeur est notée $\overline{\boldsymbol{\theta}} \in \mathbb{R}^n$. On dispose pour cela d'observations $\mathbf{y} \in \mathbb{R}^m$ dont on suppose qu'elles sont données par le modèle

$$\mathbf{y} = \mathbf{A}\overline{\boldsymbol{\theta}} + \mathbf{b} \tag{1}$$

où $\mathbf{A} \in \mathbb{R}^{m\times n}$ $(m \geq n)$ est une matrice fixée et $\mathbf{b}$ est un bruit aléatoire. On supposera $\mathbb{E}\{\mathbf{b}\} = \mathbf{0}$ et $\mathbb{C}\mathrm{ov}\{\mathbf{b}\} = \mathbb{E}\{\mathbf{b}\mathbf{b}^\top\} = \sigma^2\mathbf{Id}_m$. Enfin, par définition, pour des matrices symétriques $\mathbf{M}$ et $\mathbf{N}$ on notera $\mathbf{N} \preceq \mathbf{M}$ lorsque $\mathbf{M} - \mathbf{N}$ est semi-définie positive.

1. De façon générale, soit $\hat{\boldsymbol{\theta}}$ un estimateur de $\overline{\boldsymbol{\theta}}$. On définit :
   - son biais : $\operatorname{biais}(\hat{\boldsymbol{\theta}}) = \mathbb{E}\{\hat{\boldsymbol{\theta}}\} - \overline{\boldsymbol{\theta}}$
   - son erreur quadratique moyenne : $\operatorname{eqm}(\hat{\boldsymbol{\theta}}) = \mathbb{E}\{\|\hat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\|_2^2\}$
   - sa variance : $\mathbb{V}\mathrm{ar}\{\hat{\boldsymbol{\theta}}\} = \mathbb{E}\{\|\hat{\boldsymbol{\theta}} - \mathbb{E}\{\hat{\boldsymbol{\theta}}\}\|_2^2\}$

Montrer que $\text{eqm}(\hat{\boldsymbol{\theta}}) = \mathbb{V}\text{ar}\{\hat{\boldsymbol{\theta}}\} + \|\text{biais}(\hat{\boldsymbol{\theta}})\|_2^2$. Est-il possible d'avoir un estimateur linéaire qui vérifie les propriétés suivantes (voir exercice sur l'estimateur des moindres carrés $\hat{\boldsymbol{\theta}}_{\text{ls}}$) :

(a) la variance soit plus faible que que celle de l'estimateur des moindres carrés ?

(b) l'erreur quadratique moyenne (eqm) soit plus faible que celle de l'estimateur des moindres carrés ?

2. On définit $\hat{\boldsymbol{\theta}}_{\text{ridge}} = \text{Arg}\min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_2^2$, où $\lambda > 0$ est un paramètre fixé.

   (a) Calculer l'expression analytique de $\hat{\boldsymbol{\theta}}_{\text{ridge}}$.

   (b) Vérifier directement que $(\mathbf{A}^\top\mathbf{A} + \lambda\mathbf{Id}_n)^{-1}\mathbf{A}^\top = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top + \lambda\mathbf{Id}_m)^{-1}$. En déduire une autre expression de $\hat{\boldsymbol{\theta}}_{\text{ridge}}$.

   (c) Noter que $\hat{\boldsymbol{\theta}}_{\text{ridge}}$ est solution du problème d'optimisation :

$$\begin{aligned} &\min.\ \frac{1}{2}(\|\mathbf{u}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_2^2) \\ &\text{s.t. } \mathbf{u} = \mathbf{A}\boldsymbol{\theta} - \mathbf{y} \end{aligned}$$

   Ecrire les conditions d'optimalité KKT de ce problème et montrer que, en cherchant les solutions de ces conditions, on peut trouver l'une ou l'autre des expressions précédentes pour $\hat{\boldsymbol{\theta}}_{\text{ridge}}$.

3. Calculer $\text{biais}\,\hat{\boldsymbol{\theta}}_{\text{ridge}}$ et $\text{eqm}\,\hat{\boldsymbol{\theta}}_{\text{ridge}}$ (en passant par $\mathbb{C}\text{ov}\{\hat{\boldsymbol{\theta}}_{\text{ridge}}\}$ et $\mathbb{V}\text{ar}\{\hat{\boldsymbol{\theta}}_{\text{ridge}}\}$).

4. On suppose ici $\mathbf{A}^\top\mathbf{A} = \mathbf{Id}_n$.

   (a) Donner une expression simplifiée de $\hat{\boldsymbol{\theta}}_{\text{ridge}}$.

   (b) Calculer $\text{eqm}\,\hat{\boldsymbol{\theta}}_{\text{ridge}}$.

   (c) Montrer qu'il existe $\lambda > 0$ tel que $\text{eqm}\,\hat{\boldsymbol{\theta}}_{\text{ridge}} \leq \text{eqm}\,\hat{\boldsymbol{\theta}}_{\text{ls}}$ puis calculer le paramètre λ qui minimise $\text{eqm}\,\hat{\boldsymbol{\theta}}_{\text{ridge}}$.

5. On suppose maintenant explicitement que $\overline{\boldsymbol{\theta}}$ est la réalisation d'un vecteur aléatoire $\boldsymbol{\theta}$ centré de matrice de covariance $\sigma_{\boldsymbol{\theta}}^2\mathbf{Id}_n$. Les observations sont toujours données par l'équation (1) et le bruit est indépendant de $\boldsymbol{\theta}$.

   On s'intéresse aux estimateurs linéaires du type $\hat{\boldsymbol{\theta}} = \mathbf{B}\mathbf{y}$ où $\mathbf{B}$ est une matrice.

   (a) Calculer $\text{eqm}\,\hat{\boldsymbol{\theta}}$ en fonction de $\mathbf{A}$, $\mathbf{B}$, σ^2, $\sigma_{\boldsymbol{\theta}}^2$.

   (b) On cherche l'estimateur qui minimise $\text{eqm}\,\hat{\boldsymbol{\theta}}$. Montrer qu'il est obtenu par un problème de minimisation par rapport à $\mathbf{B}$ d'une expression dépendant de $\mathbf{A}$, $\mathbf{B}$, $\frac{\sigma^2}{\sigma_{\boldsymbol{\theta}}^2}$. Calculer la dérivée de cette expression.

   (c) Montrer que l'estimateur linéaire qui minimise l'erreur quadratique moyenne est donné par $\hat{\boldsymbol{\theta}}_{\text{ridge}}$ avec une valeur de λ que l'on précisera.

   (d) Retrouver l'expression ci-dessus en travaillant dans l'espace des variables aléatoires de carré sommable et en cherchant la projection de $\boldsymbol{\theta}$ sur l'espace vectoriel engendré par $\mathbf{y}$.

Optimisation

Convex sets, convex functions

Exercice 3: For each of the following functions determine whether it is convex, concave, quasiconvex, or quasiconcave.

1. $f(x) = e^x - 1$ on $\mathbb{R}$.
2. $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$.
3. $f(x_1, x_2) = 1/(x_1 x_2)$ on $\mathbb{R}^2_{++}$.
4. $f(x_1, x_2) = x_1/x_2$ on $\mathbb{R}^2_{++}$.
5. $f(x_1, x_2) = x_1^2/x_2$ on $\mathbb{R}^2_{++}$.
6. $f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$, on $\mathbb{R}^2_{++}$, where $0 \leq \alpha \leq 1$.

Exercice 4:

1. Show that $X \mapsto \log\det X$ with domain $\mathbb{S}_{++}$ is concave.
2. Show that $X \mapsto \operatorname{tr} X^{-1}$ with domain $\mathbb{S}_{++}$ is convex.
3. Show that $X \mapsto (\det X)^{1/n}$ with domain $\mathbb{S}_{++}$ is concave.

Exercice 5: Let x be a real-valued random variable which takes values in $\{a_1, \ldots, a_n\}$ where $a_1 < a_2 < \cdots < a_n$, with $\mathbb{P}(x = a_i) = p_i$ for $i = 1, \ldots, n$. For each of the following functions of p (on the probability simplex $\{p \in \mathbb{R}^n_+ \,|\, \mathbf{1}^\top p = 1\}$), determine if the function is convex, concave (quasiconvex, quasiconcave).

1. $\mathbb{E}\{x\}$.
2. $\mathbb{P}(x \geq \alpha)$.
3. $\mathbb{P}(\alpha \leq x \leq \beta)$.
4. $\sum_{i=1}^n p_i \log p_i$ (the negative entropy of the distribution).
5. $\mathbb{V}\mathrm{ar}\{x\} = \mathbb{E}\{(x - \mathbb{E}x)^2\}$.
6. $\mathrm{quartile}(x) = \inf\{\beta \,|\, \mathbb{P}(x \leq \beta) \geq 0.25\}$.
7. The cardinality of the smallest set $\mathcal{A} \subseteq \{a_1, \ldots, a_n\}$ with probability $\geq 90\%$.
8.

LP

Exercice 6: Formulate the following problems as LPs. Explain in detail the relation between the optimal solution of each problem and the solution of its equivalent LP.

1. Minimize $\|Ax - b\|_\infty$.
2. Minimize $\|Ax - b\|_1$.
3. Minimize $\|Ax - b\|_1$ subject to $\|x\|_\infty \leq 1$.
4. Minimize $\|x\|_1$ subject to $\|Ax - b\|_\infty \leq 1$.
5. Minimize $\|Ax - b\|_1 + \|x\|_\infty$.
6. Minimize $\mathrm{dzl}_\alpha(Ax - b)$ where $\mathrm{dzl}_\alpha(u) = \sum_{i=1}^n \max(0, |u_i| - \alpha)$ (dead-zone linear, or α-insensitive loss).

In each problem, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are given.

Optimality conditions

Exercice 7: For a differentiable objective function f_0, using only the optimality condition : $\forall x$ feasible, $\nabla f_0(x^\star)^\top (x - x^\star) \geq 0$, derive optimality conditions in the following cases :

1. Unconstrained minimization : min. $f_0(x)$.
2. Equality constrained minimization : min. $f_0(x)$ s.t. $Ax = b$.
3. Minimization over nonnegative orthant : min. $f_0(x)$ s.t. $x \succeq 0$.

Exercice 8: Consider the optimization problem

$$\begin{aligned} \text{min. } & f_0(x_1, x_2) \\ \text{s.t. } & 2x_1 + x_2 \geq 1 \\ & x_1 + 3x_2 \geq 1 \\ & x_1 \geq 0, \quad x_2 \geq 0. \end{aligned}$$

Make a sketch of the feasible set. For each of the following objective functions, give the optimal set and the optimal value.

1. $f_0(x_1, x_2) = x_1 + x_2$.
2. $f_0(x_1, x_2) = -x_1 - x_2$.
3. $f_0(x_1, x_2) = x_1$.
4. $f_0(x_1, x_2) = \max\{x_1, x_2\}$.
5. $f_0(x_1, x_2) = x_1^2 + 9x_2^2$.

Exercice 9: Prove that $x^\star = (1, 1/2, -1)$ is optimal for the optimization problem :

$$\begin{aligned} \text{min. } & \frac{1}{2} x^\top P x + q^\top x + r \\ \text{s.t. } & -1 \leq x_i \leq 1, \quad i = 1, 2, 3, \end{aligned}$$

where :

$$P = \begin{pmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{pmatrix}, \quad q = \begin{pmatrix} -22.0 \\ -14.5 \\ 13.0 \end{pmatrix}, \quad r = 1.$$