Dérivation

Exercice 1: gradient, Hessien (exemples de calculs)

1. Retrouver le gradient $\nabla f(x)$ et le Hessien $\nabla^2 f(x)$ des fonctions :
   - $\mathbb{R}^n \to \mathbb{R}, x \mapsto f(x) = a^\top x$
   - $\mathbb{R}^n \to \mathbb{R}, x \mapsto f(x) = x^\top Ax$
2. Soit la fonction $f : \mathbb{S}_{++} \to \mathbb{R}, X \mapsto \log\det(X)$. Calculer son gradient. On pourra passer par un développement de $f(X+E)$ et on précisera le produit scalaire considéré.

Modèles linéaires

Exercice 2: 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 $\mathbf{B}\mathbf{A} = \mathbf{Id}_n$. Calculer $\mathbb{C}\mathrm{ov}\{\hat{\boldsymbol{\theta}}\}$.
5. Montrer que si $\mathbf{B}\mathbf{A} = \mathbf{Id}_n$, alors $\mathbf{B}\mathbf{B}^\top \succeq (\mathbf{A}^\top\mathbf{A})^{-1}$.
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 :

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

   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 3: «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 $\mathbb{C}\text{ov}\{\mathbf{b}\} = \mathbb{E}\{\mathbf{b}\mathbf{b}^\top\} = \sigma^2 \mathbf{Id}_m$. 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 : $\text{biais}\,(\hat{\boldsymbol{\theta}}) = \mathbb{E}\{\hat{\boldsymbol{\theta}}\} - \overline{\boldsymbol{\theta}}$
   - son erreur quadratique moyenne : $\text{eqm}\,(\hat{\boldsymbol{\theta}}) = \mathbb{E}\{\|\hat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\|_2^2\}$
   - sa variance : $\mathbb{V}\text{ar}\{\hat{\boldsymbol{\theta}}\} = \mathbb{E}\{\|\hat{\boldsymbol{\theta}} - \mathbb{E}\{\hat{\boldsymbol{\theta}}\}\|_2^2\}$

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

Analyse en composantes indépendantes

Exercice 4: (in)dépendance de variables aléatoires Soient X et Y deux variables aléatoires indépendantes. X suit une loi de Bernouilli et prend de façon équiprobable les valeurs $\{-1, 1\}$. On admet que la fonction de répartition de Y est continue et vérifie la symétrie $F_Y(-y) = 1 - F_Y(y)$. Soit Z la va $Z = XY$.

1. Calculer la fonction de répartition de (X, Z) et en déduire que X et Z sont indépendantes.
2. Montrer que, sauf cas particulier, Y et Z ne sont pas indépendantes.

Exercice 5: variables binaires : indépendance mutuelle et par paires Soient X et Y deux variables aléatoires qui suivent chacune une loi de Bernouilli et prennent de façon équiprobable les valeurs $\{-1, 1\}$.

1. Montrer que pour le couple (X, Y) la décorrélation est équivalente à l'indépendance, dans ce *cas particulier*.
2. On suppsose désormais que X et Y sont indépendantes. Soit $Z = XY$.
   (a) Montrer que l'on a indépendance par paires pour (X, Y, Z) (càd (X, Y), (X, Z) et (Y, Z) sont des couples indépendants).
   (b) Montrer que les variables aléatoires X, Y, Z ne sont pas mutuellement indépendantes.

Exercice 6: statistiques d'ordre supérieur et étude de quelques contrastes

1. On admet que pour des variables aléatoires centrées, on a les expressions suivantes des cumulants en fonction des moments :

$$\begin{aligned}
\mathbb{C}\text{um}\,\{a, b\} &= \mathbb{E}\{ab\} \\
\mathbb{C}\text{um}\,\{a, b, c\} &= \mathbb{E}\{abc\} \\
\mathbb{C}\text{um}\,\{a, b, c, d\} &= \mathbb{E}\{abcd\} - \mathbb{E}\{ab\}\mathbb{E}\{cd\} - \mathbb{E}\{ac\}\mathbb{E}\{bd\} - \mathbb{E}\{ad\}\mathbb{E}\{bc\}
\end{aligned}$$

   Vérifier qu'à partir de ces relations, on retrouve les propriétés des cumulants :
   — multilinéarité
   — nullité s'il existe un sous-ensemble de variables indépendantes des autres.

   Si a, b et c sont indépendantes, que peut-on dire de $\mathbb{C}\text{um}\,\{a, b, c\}$? Si a et b sont indépendantes et c quelconque, peut-on affirmer que $\mathbb{C}\text{um}\,\{a, b, c\} = 0$?
2. On note $\mathrm{C}_4\{a\} = \mathbb{C}\text{um}\,\{a, a, a, a\}$. Calculer $\mathrm{C}_4\{a\}$ dans les cas suivants :
   — a vaut $+1$ ou -1 de manière équiprobable.
   — a est une variable aléatoire centrée, de variance unité et distribuée uniformément sur un intervalle.
3. On considère un vecteur $\mathbf{s} = (s_1, \ldots, s_N)^\top$ de sources mutuellement indépendantes, $\mathbf{G} \in \mathbb{R}^{NN}$ une matrice orthogonale et $\mathbf{y} = \mathbf{G}\mathbf{s}$. On souhaite étudier les critères de séparation suivant en fonction de $\mathbf{G}$.

$$\mathcal{C}_1(\mathbf{y}) \triangleq \sum_{i=1}^{N} |\mathrm{C}_4\{y_i\}|, \qquad \mathcal{C}_2(\mathbf{y}) \triangleq -\sum_{i=1}^{N} \mathrm{C}_4\{y_i\}, \qquad \mathcal{C}_3(\mathbf{y}) \triangleq \sum_{i=1}^{N} (\mathrm{C}_4\{y_i\})^2$$

   On notera $\kappa_i \triangleq \mathrm{C}_4\{s_i\}$ pour tout i.
   (a) Développer $\mathrm{C}_4\{y_i\}$ en fonction des κ_i et des coefficients de la matrice $\mathbf{G}$.
   (b) Montrer que $\mathcal{C}_1(\mathbf{y}) \leq \sum_{i=1}^{N} |\kappa_i|$ (indication : utiliser l'orthogonalité de $\mathbf{G}$).
   (c) On suppose $\forall i, \kappa_i \neq 0$. Montrer qu'en cas d'égalité dans la majoration précédente, $\mathbf{G}$ n'a qu'un et un seul élément non nul sur chaque ligne et chaque colonne et que cet élément vaut $+1$ ou -1 (indication : utiliser l'orthogonalité de $\mathbf{G}$).

(d) Peut-on affaiblir l'hypothèse $\forall i, \kappa_i \neq 0$? Commenter.

4. Reprendre l'étude précédente pour $\mathcal{C}_2(\mathbf{y})$ en supposant que les $\kappa_i \neq 0$ sont négatifs. Préciser les hypothèses pour que $\mathcal{C}_2(\mathbf{y})$ soit un contraste.
5. Prouver que $\mathcal{C}_3(\mathbf{y})$ est un contraste sous les mêmes hypothèses que $\mathcal{C}_1(\mathbf{y})$ (indication : utiliser la convexité de $(.)^2$ afin d'obtenir une majoration).