Course Applied maths, Frédéric Richard

Master Mathématiques appliquées, statistique.

Parcours Data Sciences and CMB.

Evaluation 05/12/2024

Exercice 1.

1. L'application f est polynomiale, donc de classe $\mathcal{C}^\infty$.
2. Pour x et h dans $\mathbb{R}^p$, nous obtenons

$$f(x+h) - f(x) = \frac{1}{2}\langle Ax, h\rangle + \frac{1}{2}\langle Ah, x\rangle - \langle b, h\rangle + \frac{1}{2}\langle Ah, h\rangle,$$

en développant $\langle A(x+h), (x+h)\rangle$. Comme A est symétrique, nous avons encore

$$f(x+h) - f(x) = \langle Ax - b, h\rangle + \frac{1}{2}\langle Ah, h\rangle.$$

Posons $L(h) = \langle Ax - b, h\rangle$ et $\epsilon(h) = \frac{1}{2}\langle Ah, h\rangle$.

L'application L est linéaire et, comme elle est définie sur un espace de dimension finie, elle est continue.

Par ailleurs, A étant symétrique et définie positive, elle est diagonalisable et de valeurs propres strictement positives. De plus,

$$\epsilon(h) \leq c|h|^2,$$

où c est la plus grande valeur propre de A.

Ainsi, $\frac{\epsilon(h)}{|h|} \leq c|h|$, et

$$\lim_{|h|\to 0} \frac{\epsilon(h)}{|h|} = 0.$$

Par conséquent, l'application L est la différentielle df_x de f en x. Il s'agit d'une forme linéaire continue définie sur $\mathbb{R}^p$.

3. Ayant montré que

$$df_x(h) = \langle Ax - b, h\rangle,$$

nous pouvons immédiatement identifier le gradient de f en x:

$$\nabla f(x) = Ax - b,$$

qui est un élément de $\mathbb{R}^p$.

4. Un calcul rapide donne:

$$df_{x+k}(h) - df_x(h) = \langle Ak, h\rangle.$$

Or, l'application qui à k associe $\langle Ak, h\rangle$ est linéaire et continue. De ce fait, il s'agit de la différentielle de df_x. Par conséquent, la différentielle à l'ordre 2 de f en x est la forme bilinéaire $d^2 f_x$ qui à tout $(h, k) \in \mathbb{R}^p \times \mathbb{R}^p$ associe

$$d^2 f_x(h, k) = \langle Ak, h\rangle.$$

5. La forme de cette différentielle à l'ordre 2 de f permet d'identifier directement la matrice hessienne de f en x:

$$\nabla^2 f(x) = A.$$

6. Nous avons vu que

$$f(x+h) - f(x) = \langle \nabla f(x), h\rangle + \frac{1}{2}\langle Ah, h\rangle.$$

De plus, comme A est définie positive, nous avons
$\langle Ah, h\rangle > 0$ dès lors que $h \neq 0$.

Ainsi, lorsque $h \neq 0$,

$$f(x+h) - f(x) > \langle \nabla f(x), h\rangle.$$

7. De par la caractérisation des fonctions strictement convexes vue en cours, cette dernière égalité implique que f est strictement convexe. Par définition, cela revient à dire que, pour tout $t \in]0, 1[$ et tous $x, y \in \mathbb{R}^p$,

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y).$$

Exercice 2.

1. D'après les hypothèses du modèle, les variables observées Y_i sont indépendantes et de loi normale $\mathcal{N}\left(\exp(-\sum_{j=1}^p \theta_j x_{ij}), \sigma^2\right)$. Ainsi la vraisemblance du modèle s'écrit

$$\begin{aligned}\mathcal{L}(\theta; y_1, \cdots, y_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(y_i - \exp\left(-\sum_{j=1}^p \theta_j x_{ij}\right)\right)^2\right) \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n\left(y_i - \exp\left(-\sum_{j=1}^p \theta_j x_{ij}\right)\right)^2\right)\end{aligned}$$

2. Maximiser la vraisemblance revient à minimiser la log-vraisemblance négative, qui, à une constante près vaut $g(\theta)$.

3. L'application g est différentiable car il s'agit d'une somme d'applications, qui sont des composées d'applications différentiables.
4. Calculons les dérivées partielles de g. Pour cela, posons

$$h_i(\theta) = \exp\left(-\sum_{j=1}^{p}\theta_j x_{ij}\right)$$

et commençons par calculer des dérivées partielles de

$$g_i(\theta) = \left(y_i - h_i(\theta)\right)^2.$$

En utilisant les formules de dérivation de fonctions composées, nous avons

$$\frac{\partial g_i(\theta)}{\partial \theta_k} = -2\left(y_i - h_i(\theta)\right)\frac{\partial h_i(\theta)}{\partial \theta_k}$$

et

$$\frac{\partial h_i(\theta)}{\partial \theta_k} = -x_{ik}\exp\left(-\sum_{j=1}^{p}\theta_j x_{ij}\right) = -x_{ik}h_i(\theta).$$

Ainsi,

$$\frac{\partial g(\theta)}{\partial \theta_k} = \sum_{i=1}^{n} x_{ik}(y_i - h_i(\theta))h_i(\theta).$$

En outre,

$$\nabla g(x) = \begin{pmatrix} \frac{\partial g(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial g(\theta)}{\partial \theta_p} \end{pmatrix} = \sum_{i=1}^{n} x_i(y_i - h_i(\theta))h_i(\theta).$$

5. A tout $h \in \mathbb{R}^p$, la différentielle dg_x de g en x associe

$$dg_x(h) = \langle \nabla g(x), h\rangle.$$

Plus spécifiquement,

$$dg_x(h) = \sum_{i=1}^{n}\langle x_i, h\rangle(y_i - h_i(\theta))h_i(\theta).$$

6. Nous avons

$$
\begin{aligned}
\frac{\partial^2 g(\theta)}{\partial\theta_l \partial\theta_k} &= \frac{\partial}{\partial\theta_l}\frac{\partial g(\theta)}{\partial\theta_k} \\
&= \frac{\partial\left[\sum_{i=1}^n x_{ik}(y_i - h_i(\theta))h_i(\theta)\right]}{\partial\theta_l}, \\
&= \sum_{i=1}^n x_{ik}\frac{\partial\left[(y_i - h_i(\theta))h_i(\theta)\right]}{\partial\theta_l}, \\
&= \sum_{i=1}^n x_{ik}\left[\frac{\partial(y_i - h_i(\theta))}{\partial\theta_l}h_i(\theta) + (y_i - h_i(\theta))\frac{\partial h_i(\theta)}{\partial\theta_l}\right] \\
&= \sum_{i=1}^n x_{ik}\left[-\frac{\partial h_i(\theta)}{\partial\theta_l}h_i(\theta) + (y_i - h_i(\theta))\frac{\partial h_i(\theta)}{\partial\theta_l}\right] \\
&= \sum_{i=1}^n x_{ik}\left[x_{il}h_i^2(\theta) - (y_i - h_i(\theta))x_{il}h_i(\theta)\right] \\
&= \sum_{i=1}^n x_{ik}x_{il}(2h_i^2(\theta) - y_i).
\end{aligned}
$$

En outre, nous remarquons que la matrice hessienne de g en x vaut

$$
\nabla^2 g(x) = \sum_{i=1}^n x_i x_i^T (2h_i^2(\theta) - y_i).
$$

7. Le développement de Taylor-Young de g en θ s'écrit

$$
g(\theta+\kappa) = g(\theta) + \langle\nabla g(\theta), \kappa\rangle + \frac{1}{2}\langle\nabla^2 g(\theta)\kappa, \kappa\rangle + |\kappa|^2\epsilon(\kappa),
$$

où $\epsilon(\kappa)$ tend vers 0 lorsque κ tend vers 0.