TD 01 : Introduction à l'optimisation

Outils d'optimisation pour les sciences des données et de la décision, M2 MIAGE

15 septembre 2023

Exercice 1 : Moindres carrés linéaires

On considère un jeu de données de la forme $\{(\boldsymbol{a}_i, b_i)\}_{i=1}^m$, où chaque $\boldsymbol{a}_i$ est un vecteur de $\mathbb{R}^n$ et chaque b_i appartient à $\mathbb{R}$. On cherche un modèle linéaire qui explique les données, que l'on obtient en considérant le problème :

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimiser}}\, f(\boldsymbol{x}) := \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2 = \frac{1}{2}\sum_{i=1}^{m}(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}-b_i)^2, \tag{1}$$

où $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ et $\boldsymbol{b} \in \mathbb{R}^m$ concatènent les données, c'est-à-dire que

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}} \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

Ce problème est un des plus classiques en analyse de données; sa fonction objectif est de classe $\mathcal{C}^2$, et on peut montrer qu'il possède toujours au moins une solution.

a) Supposons que $\boldsymbol{x}^*$ vérifie $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$ (c'est donc une solution du système linéaire $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$). Justifier que $\boldsymbol{x}^*$ est alors un minimum global du problème.

b) Le gradient de f en $\boldsymbol{x} \in \mathbb{R}^n$ est donné par $\nabla f(\boldsymbol{x}) = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})$. Si $\boldsymbol{x}^*$ est un minimum local de f, que vaut $\nabla f(\boldsymbol{x}^*)$?

c) La matrice hessienne de f en $\boldsymbol{x} \in \mathbb{R}^n$ est donnée par $\nabla^2 f(\boldsymbol{x}) = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$. Elle est donc constante et définie par les données du problème.

   i) On a toujours $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \succeq \boldsymbol{0}$. Quelle propriété sur f cela implique-t-il ?

   ii) On suppose que $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \succeq \mu \boldsymbol{I}_n$ avec $\mu > 0$. Dans ce cas, que peut-on dire de $\nabla^2 f(\boldsymbol{x})$ pour tout $\boldsymbol{x}$? Qu'en déduit-on sur l'ensemble des solutions du problème (1) ?

Exercice 2 : Fonction convexe

Soit la fonction $q : \mathbb{R}^n \to \mathbb{R}$ définie par $q(\boldsymbol{x}) = \frac{1}{4}\|\boldsymbol{x}\|^4$. Cette fonction est de classe $\mathcal{C}^2$, et pour tout $\boldsymbol{x} \in \mathbb{R}^n$, on a :

$$\nabla q(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 \boldsymbol{x}, \qquad \nabla^2 q(\boldsymbol{x}) = 2\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} + \|\boldsymbol{x}\|^2 \boldsymbol{I}_n.$$

a) En utilisant sa matrice hessienne, montrer que la fonction q est convexe. Quelle conséquence cela a-t-il sur ses minima ?

b) Montrer que le vecteur nul $\mathbf{0}_{\mathbb{R}^n}$ est un minimum local. Satisfait-il la condition suffisante à l'ordre 2 ?

c) En fonction de la réponse à la question précédente, la fonction peut-elle alors être fortement convexe ?

Exercice 3 : Fonctions quasi-convexes

Une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est dite **quasi-convexe** si

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,\ \forall t \in [0,1], \quad f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \leq \max\{f(\boldsymbol{x}), f(\boldsymbol{y})\}. \tag{2}$$

Toute fonction convexe est quasi-convexe, mais la réciproque est fausse.

On s'intéresse ici aux solutions du problème

$$\underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimiser}}\ f(\boldsymbol{x}), \tag{3}$$

où l'on suppose que f est quasi-convexe et de classe $\mathcal{C}^2$.

a) Donner les conditions d'optimalité nécessaires à l'ordre 1 et à l'ordre 2 pour le problème (3).

b) Comme f est quasi-convexe, on a la propriété suivante :

$$\forall \boldsymbol{x} \in \mathbb{R}^n,\ \forall \boldsymbol{v} \in \mathbb{R}^d, \quad \boldsymbol{v}^{\mathrm{T}} \nabla f(\boldsymbol{x}) = 0 \Rightarrow \boldsymbol{v}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}) \boldsymbol{v} \geq 0. \tag{4}$$

Soit $\boldsymbol{x}^*$ un point stationnaire d'ordre 1. Justifier que $\boldsymbol{x}^*$ est aussi un point stationnaire d'ordre 2.

Solutions des exercices

Solutions de l'exercice 1

a) Si $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$, alors on a
$$f(\boldsymbol{x}^*) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}\|^2 = \frac{1}{2}\|\mathbf{0}\|^2 = 0.$$
Or la fonction f est toujours positive ou nulle; on a ainsi
$$\forall \boldsymbol{x} \in \mathbb{R}^n, f(\boldsymbol{x}) \geq 0 = f(\boldsymbol{x}^*).$$
Cette propriété correspond à la définition d'un minimum global, d'où l'on conclut que $\boldsymbol{x}^*$ est bien un minimum global du problème.

b) Si $\boldsymbol{x}^*$ est un minimum local du problème (1) et donc de f, alors on a $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$. C'est la condition d'optimalité nécessaire à l'ordre 1.

i) Si $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \succeq \mathbf{0}$, alors on a $\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0}$ pour tout $\boldsymbol{x}$: c'est une caractérisation de la convexité pour une fonction de classe $\mathcal{C}^2$, et l'on en conclut donc que f est convexe.

ii) Comme dans la question précédente, le fait que $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \succeq \mu \boldsymbol{I}_n$ signifie que $\nabla^2 f(\boldsymbol{x}) \succeq \mu \boldsymbol{I}_n$ pour tout $\boldsymbol{x} \in \mathbb{R}^n$. C'est une caractérisation de la convexité forte, d'où l'on conclut que f est μ-fortement convexe. Par conséquent, la solution du problème (on sait qu'il en existe au moins une d'après l'énoncé) est unique.

Solutions de l'exercice 2

a) Pour tous $\boldsymbol{x} \in \mathbb{R}^n$ et $\boldsymbol{v} \in \mathbb{R}^n$, on a en utilisant la linéarite des produits scalaires et produits matrice-vecteur :
$$\begin{array}{rcl}
\boldsymbol{v}^{\mathrm{T}}\nabla^2 q(\boldsymbol{x})\boldsymbol{v} & = & \boldsymbol{v}^{\mathrm{T}}(2\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} + \|\boldsymbol{x}\|^2 \boldsymbol{I}_n)\boldsymbol{v} \\
& = & \boldsymbol{v}^{\mathrm{T}}(2\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v} + \|\boldsymbol{x}\|^2 \boldsymbol{v}) \\
& = & 2\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v} + \|\boldsymbol{x}\|^2 \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} \\
& = & 2(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v})^2 + \|\boldsymbol{x}\|^2 \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} \\
& = & 2(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v})^2 + \|\boldsymbol{x}\|^2 \|\boldsymbol{v}\|^2 \\
& \geq & 0.
\end{array}$$
Par conséquent, pour tout $\boldsymbol{x} \in \mathbb{R}^n$, la matrice hessienne $\nabla^2 q(\boldsymbol{x})$ est semi-définie positive : on a $\nabla^2 q(\boldsymbol{x}) \succeq \mathbf{0}$. On en déduit que la fonction q est convexe, et donc que tous ses minima locaux sont globaux (elle ne possède ainsi que des minima globaux).

b) Puisque q est convexe, il y a équivalence entre minimum local et minimum global. Or, on a
$$q(\boldsymbol{x}) = \tfrac{1}{4}\|\boldsymbol{x}\|^4 \geq 0 = q(\mathbf{0}_{\mathbb{R}^n}).$$
Le vecteur nul $\mathbf{0}_{\mathbb{R}^n}$ est donc un minimum global de q. Pour satisfaire la condition suffisante d'optimalité à l'ordre 2, il faudrait avoir $\nabla^2 q(\boldsymbol{x}) \succ \mathbf{0}$; or, on trouve en remplaçant dans l'expression donnée dans l'énoncé que
$$\nabla^2 q(\mathbf{0}_{\mathbb{R}^x}) = \mathbf{0},$$

qui n'est pas définie positive mais uniquement semi-définie positive : par conséquent, le vecteur nul ne vérifie pas la condition suffisante d'optimalité. *Remarque : cela ne contredit pas le fait que $\mathbf{0}$ est un minimum global.*

c) Si la fonction était fortement convexe, on aurait $\nabla^2 q(\boldsymbol{x}) \succeq \mu \boldsymbol{I}_n \succ \mathbf{0}$ pour tout $\boldsymbol{x}$, et donc en particulier pour le vecteur nul. Ce n'est pas le cas, et on en conclut donc que cette fonction n'est pas fortement convexe.

Solutions de l'exercice 3

a) *Il s'agit d'une question de cours.* La condition nécessaire d'optimalité à l'ordre 1 s'énonce comme suit : si $\boldsymbol{x}^* \in \mathbb{R}^n$ est un minimum local de f, alors $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$. La condition nécessaire d'optimalité à l'ordre 2 est plus précise encore : si $\boldsymbol{x}^* \in \mathbb{R}^n$ est un minimum local de f, alors

$$\nabla f(\boldsymbol{x}^*) = \mathbf{0} \quad \text{et} \quad \nabla^2 f(\boldsymbol{x}^*) \succeq \mathbf{0}.$$

b) Puisque $\boldsymbol{x}^*$ est un point stationnaire d'ordre 1, il vérifie la condition d'optimalité nécessaire à l'ordre 1 : on a donc $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$. Par conséquent, on a

$$\forall \boldsymbol{v} \in \mathbb{R}^d, \quad \boldsymbol{v}^{\mathrm{T}} \nabla f(\boldsymbol{x}^*) = \boldsymbol{v}^{\mathrm{T}} \mathbf{0} = 0.$$

La première partie de l'implication (4) est donc vraie pour $\boldsymbol{x}^*$ et pour tout vecteur $\boldsymbol{v}$. On en déduit donc que la seconde partie de l'implication l'est aussi, c'est-à-dire que l'on a :

$$\boldsymbol{v}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{v} \geq 0 \ \forall \boldsymbol{v} \in \mathbb{R}^n,$$

qui correspond à $\nabla^2 f(\boldsymbol{x}^*) \succeq \mathbf{0}$. Par conséquent, $\boldsymbol{x}^*$ vérifie la condition nécessaire d'optimalité à l'ordre 2 : c'est donc bien un point stationnaire d'ordre 2.