

TD2 - Optimisation pour l'apprentissage automatique

Exercice 1 : Moindres carrés linéaires

On considère un ensemble de données $\{(x_i, y_i)\}_{i=1}^n$, où $x_i \in \mathbb{R}^d$ et $y_i \in \mathbb{R}$ pour tout $i = 1, \dots, n$. Nous cherchons un modèle linéaire qui ajuste au mieux les données, ce qui se formule comme le problème d'optimisation suivant :

$$\min_{w \in \mathbb{R}^d} f(w) := \frac{1}{2} \|Xw - y\|^2 = \frac{1}{2} \sum_{i=1}^n (x_i^T w - y_i)^2,$$

où $X \in \mathbb{R}^{n \times d}$ et $y \in \mathbb{R}^n$ sont donnés par

$$X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

Ce problème est parmi les plus classiques en analyse de données. La fonction objectif est de classe C^2 , ce qui garantit au moins une solution.

- a) Si $w^* \in \mathbb{R}^d$ satisfait $Xw^* = y$, justifiez que w^* est un minimum global de la fonction objectif.
- b) Le gradient de f en tout $w \in \mathbb{R}^d$ est donné par $\nabla f(w) = X^T(Xw - y)$. Si w^* est un minimum local de f , quelle est la valeur de $\nabla f(w^*)$?
- c) La matrice Hессienne de f en tout $w \in \mathbb{R}^d$ est $\nabla^2 f(w) = X^T X$. i) En utilisant le fait que $X^T X \succeq 0$, quelle propriété cela implique-t-il pour f ? ii) Supposons que $X^T X \succeq \mu I_d$ avec $\mu > 0$. Que peut-on en dire concernant les solutions de ce problème d'optimisation ?

Exercice 2 : Fonction convexe

Soit $q : \mathbb{R}^d \rightarrow \mathbb{R}$ définie par $q(w) = \frac{1}{4}\|w\|^4$. Cette fonction est de classe C^2 , et pour tout $w \in \mathbb{R}^d$, nous avons

$$\nabla q(w) = \|w\|^2 w, \quad \nabla^2 q(w) = 2ww^T + \|w\|^2 I_d.$$

- a) Montrez que la fonction q est convexe en utilisant l'expression de la matrice Hessianne. Que cela implique-t-il sur ses minima locaux ?

- b) Montrez que le vecteur nul $0 \in \mathbb{R}^d$ est un minimum local de q . Satisfait-il les conditions suffisantes d'optimalité du second ordre ?
- c) En se basant sur la réponse précédente, la fonction q peut-elle être fortement convexe ?
- d) Justifiez que la fonction possède un unique minimum global.

Exercice 3 : Fonctions quasiconvexes

Une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est dite quasiconvexe si

$$\forall w, v \in \mathbb{R}^d, \forall t \in [0, 1], \quad f(tw + (1 - t)v) \leq \max\{f(w), f(v)\}.$$

Toute fonction convexe est quasiconvexe, mais la réciproque n'est pas vraie.

Soit f une fonction quasiconvexe de classe C^2 . Considérons le problème suivant :

$$\min_{w \in \mathbb{R}^d} f(w).$$

a) Écrivez les conditions d'optimalité du premier et du second ordre pour ce problème.

b) Puisque f est quasiconvexe, on peut montrer que

$$\forall w \in \mathbb{R}^d, \forall v \in \mathbb{R}^d, \quad v^T \nabla f(w) = 0 \implies v^T \nabla^2 f(w) v \geq 0.$$

Justifiez que si w^* est un point stationnaire du premier ordre, il est aussi un point stationnaire du second ordre.