Tutorial 1: Basics of optimization

Optimization for data science, M2 MIAGE ID/ID Apprentissage

September 29, 2025

Exercise 1: Linear least squares

We consider a dataset $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, wherein $\boldsymbol{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$ for every $i = 1, \ldots, n$. We seek a linear model that best fits the data, which we formulate as the following optimization problem:

$$\underset{\boldsymbol{w}\in\mathbb{R}^d}{\text{minimize}}\, f(\boldsymbol{w}) := \frac{1}{2\,n}\|\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\|^2 = \frac{1}{2\,n}\sum_{i=1}^{n}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{w} - y_i)^2, \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ and $\boldsymbol{y} \in \mathbb{R}^n$ are given by

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}} \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

This problem is among the most classical in data analysis. Its objective function is $\mathcal{C}^2$, and the problem (1) always has at least one solution.

a) Let $\boldsymbol{w}^* \in \mathbb{R}^d$ satisfy $\boldsymbol{X}\boldsymbol{w}^* = \boldsymbol{y}$ (hence $\boldsymbol{w}^*$ is a solution of the linear system $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$). Justify then that $\boldsymbol{w}^*$ is a global minimum of the objective function.

b) The gradient of f at any $\boldsymbol{w} \in \mathbb{R}^d$ is given by $\nabla f(\boldsymbol{w}) = \frac{1}{n}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y})$. If $\boldsymbol{w}^*$ satisfies $\boldsymbol{X}\boldsymbol{w}^* = \boldsymbol{y}$ as in question a), what is the value of $\nabla f(\boldsymbol{w}^*)$?

c) The Hessian matrix of f at $\boldsymbol{w} \in \mathbb{R}^d$ is given by $\nabla^2 f(\boldsymbol{w}) = \frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$. Note that it is constant with respect to $\boldsymbol{w}$, and that it only depends on the data matrix $\boldsymbol{X}$.

   i) By construction, we have $\frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \succeq \mathbf{0}$. What property on f does this imply?

   ii) Suppose that $\frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \succeq \mu\boldsymbol{I}_d$ with $\mu > 0$. Given $\boldsymbol{w} \in \mathbb{R}^d$, what can we say about $\nabla^2 f(\boldsymbol{w})$ in that case? What information does this provide about the set of solutions of problem (1)?

Exercise 2: Convex function

Let $q : \mathbb{R}^d \to \mathbb{R}$ be defined as $q(\boldsymbol{w}) = \frac{1}{4}\|\boldsymbol{w}\|^4$. This function is $\mathcal{C}^2$, and for every $\boldsymbol{w} \in \mathbb{R}^d$, we have

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

a) Using the expression of the Hessian matrix of q, show that the function q is convex. What does it imply on its local minima?

b) Show that the zero vector $\mathbf{0}_{\mathbb{R}^d}$ is a local minimum of q. Does it satisfy the second-order sufficient condition?

c) Given the answer to the previous question, can the function q be strongly convex?

d) Justify that the function has a single global minimum.

Exercise 3: Quasiconvex functions

A function $f : \mathbb{R}^d \to \mathbb{R}$ is called **quasiconvex** if

$$\forall \boldsymbol{w}, \boldsymbol{v} \in \mathbb{R}^d,\ \forall t \in [0,1], \quad f(t\boldsymbol{w} + (1-t)\boldsymbol{v}) \le \max\{f(\boldsymbol{w}), f(\boldsymbol{v})\}. \tag{2}$$

Any convex function is quasiconvex, but the converse is not true.

Let f be a quasiconvex, $\mathcal{C}^2$ function. We consider:

$$\operatorname*{minimize}_{\boldsymbol{w}\in\mathbb{R}^d} f(\boldsymbol{w}). \tag{3}$$

a) Write the first- and second-order optimality conditions for problem (3).

b) Since f is quasiconvex, it can be shown that

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

Let $\boldsymbol{w}^*$ be a first-order stationary point. Justify that $\boldsymbol{w}^*$ is also a second-order stationary point.