### **Session 9: Convex functions**

Optimization and Computational Linear Algebra for Data Science

#### Léo Miolane

#### **Contents**

- 1. Recap of the videos
- 2. Convex sets and convex functions
- 3. Convex functions and derivatives
- 4. Jensen's inequality

### **Optimization**

In machine learning, we often have to minimize functions

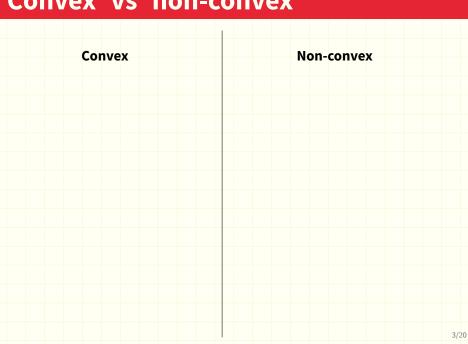
$$f(\theta) = \operatorname{Loss}(\operatorname{data}, \operatorname{model}_{\theta})$$
 with respect to  $\theta \in \mathbb{R}^n$ .

- For n = 1, 2, one could plot f to find the minimizer.
- This is intractable for larger dimension.

#### We will

- focus on convex cost functions f.
- study gradient descent algorithms to minimize *f*.

### Convex vs non-convex



### **Gradient/Hessian**

For  $f: \mathbb{R}^n \to \mathbb{R}$ :

. J . 22 / 22

Gradient at 
$$x \in \mathbb{R}^n$$
:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix} \in \mathbb{R}^n$$

Hessian at  $x \in \mathbb{R}^n$ :

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix} \in \mathbb{R}^{n \times n}$$

### Taylor's formulas

Let  $x \in \mathbb{R}^n$ . Heuristically, for  $h \in \mathbb{R}^n$  "small", we have

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

### Taylor's formulas

Let  $x \in \mathbb{R}^n$ . Heuristically, for  $h \in \mathbb{R}^n$  "small", we have

$$f(x+h) \simeq f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} h^{\mathsf{T}} H_f(x) h.$$

# **Convex sets**

Convex sets 7/20

#### Convex set

#### Definition

A set  $S \subset \mathbb{R}^n$  is called a convex set if for all  $x, y \in C$  and all  $\alpha \in [0,1]$ ,

$$\alpha x + (1 - \alpha)y \in C.$$

#### **Exercise**

- 1. Show that any subspace S of  $\mathbb{R}^n$  is convex.
- 2. Let  $\|\cdot\|$  be a (arbitrary) norm and  $r\geq 0$ . Show that the "ball" of radius r:

$$B(r) = \{ x \in \mathbb{R}^n \mid ||x|| \le r \}$$

is convex.

## **Convex functions**

Convex functions 10/20

### **Convex / concave functions**

#### Definition

A function  $f:\mathbb{R}^n o \mathbb{R}$  is convex if for all  $x,y \in \mathbb{R}^n$  and all  $\alpha \in [0,1]$ ,

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y). \tag{1}$$

#### **Convex / concave functions**

#### Definition

A function  $f:\mathbb{R}^n \to \mathbb{R}$  is convex if for all  $x,y \in \mathbb{R}^n$  and all  $\alpha \in [0,1]$ ,

$$f(\alpha x + (1 - \alpha)y) \le \alpha f(x) + (1 - \alpha)f(y). \tag{1}$$

- We say that f is *strictly convex* is there is strict inequality in (1) whenever  $x \neq y$  and  $\alpha \in (0, 1)$ .
- A function f is called concave if -f is convex.

Convex functions 11/20

#### **Exercise**

- 1. Show that any linear map  $f:\mathbb{R}^n \to \mathbb{R}$  is convex and concave.
- 2. Show that a norm  $\|\cdot\|$  is convex.
- 3. Show that the sum of two convex functions is also a convex function.

Convex functions 12/20

### **Convex functions vs their tangents**

#### Proposition

A differentiable function  $f:\mathbb{R}^n \to \mathbb{R}$  is convex if and only if for all  $x,y\in\mathbb{R}^n$ 

$$f(y) \ge f(x) + \langle \nabla f(x), (y - x) \rangle.$$

### **Proof**

15/20

### **Proof**

15/20

### Proof

15/20

#### Minimizers of a convex function

#### Corollary

Let  $f:\mathbb{R}^n \to \mathbb{R}$  be a differentiable convex function and  $x \in \mathbb{R}^n$ . Then

$$x \ \ \text{is a minimizer of} \ \ f \quad \Longleftrightarrow \quad \nabla f(x) = 0.$$

#### **Hessian of convex function**

#### Proposition

Let  $f: \mathbb{R}^n \to \mathbb{R}$  be a twice-differentiable function. Then f is convex if and only if for all  $x \in \mathbb{R}^n$ ,  $H_f(x)$  is positive semi-definite.

#### **Hessian of convex function**

#### Proposition

Let  $f: \mathbb{R}^n \to \mathbb{R}$  be a twice-differentiable function. Then f is convex if and only if for all  $x \in \mathbb{R}^n$ ,  $H_f(x)$  is positive semi-definite.

# Jensen's inequality

Jensen's inequality

### Jensen's inequality

#### Theorem

Let  $f: \mathbb{R}^n \to \mathbb{R}$  be a convex function. Then for all  $x_1, \dots, x_k \in \mathbb{R}^n$  and all  $\alpha_1, \dots, \alpha_k \ge 0$  such that  $\sum_{i=1}^k \alpha_i = 1$  we have

$$f\left(\sum_{i=1}^{k} \alpha_i x_i\right) \le \sum_{i=1}^{k} \alpha_i f(x_i).$$

More generally, if X is a random variable that takes value in  $\mathbb{R}^n$  we have

$$f(\mathbb{E}[X]) \le \mathbb{E}[f(X)].$$

Jensen's inequality 19/20

# **Example: entropy**

Jensen's inequality

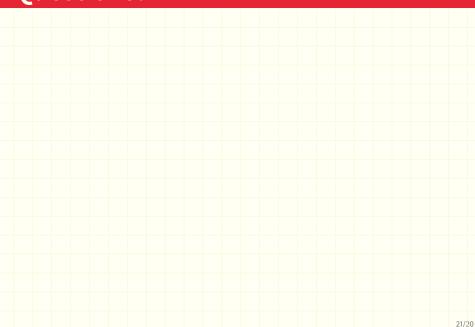
20/20

# **Example: entropy**

Jensen's inequality

20/20

# **Questions?**



# **Questions?**

