HMMA 307 : Advanced Linear Modeling

Chapter 2 : Linear regression and optimization

Delage Cindy, Hanna Bacave and Ruoyu Wang
https://github.com/hannabacave/MLA Université de

Montpellier

Table of Contents

Table of Contents

Convexity reminder

Theorem

Consider a function $f : \begin{array}{ccc} \mathbb{R}^d & \longrightarrow & \mathbb{R}^d \\ x & \longmapsto & f(x) \end{array}$, if f is convex and $\mathcal{C}^1$, then

$$x^\star \in \arg\min_{x \in \mathbb{R}^d} f(x) \Leftrightarrow \nabla f(x^\star) = 0$$

<u>Rem</u>: f convex and $\mathcal{C}^1 \Leftrightarrow \forall x_1 \in \mathbb{R}^d, \forall x_2 \in \mathbb{R}^d$,

$$f(x) \geq f(x_1) + < \nabla f(x_1), x_1 - x_2 >$$

Table of Contents

Model implementation

Target

The target is to reach optimum of the function

$$f_0 : \begin{array}{ccc} \mathbb{R}^d & \longrightarrow & \mathbb{R}^d \\ x & \longmapsto & f_0(x) \end{array}$$ under constraints.

We apply n_1 inegality constraints and n_2 egality constraints at f_0, like :

- $f_i(x) \;\leq\; 0 \;\; \forall i \in \{1, \cdots, n_1\}$;
- $h_j(x) \;=\; 0 \;\; \forall j \in \{1, \cdots, n_2\}$.

To achieve the goal, we need to suppose that :

- f_0 and $\forall i, \forall j,\; f_i$, h_j are $\mathcal{C}^1$;
- f_0 and $\forall i, f_i$ are convex.

Model implementation

Definition

We define the **feasability set** as
$\mathcal{F} = \{x \in \mathbb{R}^d \ : \ \forall i \in [\![1, n_1]\!], f_i(x) \leq 0, \ \forall j \in [\![1, n_2[\![, \ h_j(x) = 0\}$

We get the following problem **constrained**: $\min\limits_{x \in \mathcal{F}} f_0\left(x\right)$

Definition

We define the **primal value** (**optimal value**) of the problem by

$$p^\star = \min_{x \in \mathcal{F}} f_0(x)$$

Definition

We call **Lagrangian** (or **Lagrangian multiplier**) the function such that $\forall\, x \in \mathbb{R}^d, \lambda \in \mathbb{R}^{n_1}$ and $\nu \in \mathbb{R}^{n_2}$:

$$\mathcal{L}(x, \lambda, \nu) = f_0\,(x) + \sum_{i=1}^{n_1} \lambda_i\, f_i\,(x) + \sum_{j=1}^{n_2} \nu_j\, h_j\,(x)$$

Definition

$$\begin{aligned} g: \quad \mathbb{R}^{n_1} * \mathbb{R}^{n_2} &\to \mathbb{R} \\ (\lambda, \nu) &\mapsto \min_{x \in \mathbb{R}^d} \mathcal{L}(x, \lambda, \nu) \end{aligned}$$

is the *dual function* of the problem.

<u>Rem</u>:

- g is a concave function (minimum of affine functions)
- $\forall \lambda \geq 0 \, (\lambda_1 \geq 0, \ldots, \lambda_{n_1} \geq 0)$ and $\forall \, x \in \mathcal{F}$ we have :

$$\begin{aligned} \mathcal{L}(x, \lambda, \nu) &= f_0(x) + \underbrace{\sum_{i=1}^{n_1} \lambda_i \, f_i(x)}_{\leq 0} + \underbrace{\sum_{j=1}^{n_2} \nu_j \, h_j(x)}_{=0} \\ &\leq f_0(x). \end{aligned}$$

So we have :

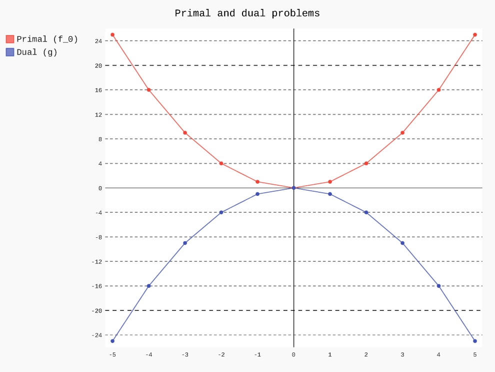
$$\forall \{\min_{x \in \mathbb{R}^d} \mathcal{L}(x, \lambda, \nu)\}_{g(\lambda,\nu)} \leq f_0(x)$$

$$g(\lambda, \nu) \leq \min_{x \in \mathcal{F}} f_0\left(x\right) = p^*.$$

Definition

The **dual problem** consists to find $d^* = \max_{(\lambda,\nu) \in \mathbb{R}^{n_1} * \mathbb{R}^{n_2}} g(\lambda, \nu)$ such as $\lambda \geq 0$.

Primal/Dual



Primal and dual problems
Primal (f_0)
Dual (g)
24
20
16
12
8
4
0
-4
-8
-12
-16
-20
-24
-5
-4
-3
-2
-1
0
1
2
3
4
5

Remark

$\forall \lambda \geq 0, \forall x,$

$$g(\lambda, \nu) \overbrace{\leq}^{\text{def. of } d^*} d^* \overbrace{\leq}^{\text{weak duality}} p^* \overbrace{\leq}^{\text{def. of } p^*} f_0(x)$$

We call the *strong duality* when $d^* = p^*$.

Theorem

If $\forall i \in [\![1, n_1]\!], f_i$ are convex; $\forall j \in [\![1, n_2]\!], h_j$ are affine. If $\exists \tilde{x} \in \mathbb{R}^d$ such as:

$$f_i(\tilde{x}) < 0, \forall i \in [\![1, n_1]\!]$$
$$h_j(\tilde{x}) = 0, \forall j \in [\![1, n_2]\!]$$

the strong duality is satisfied and

$$d^* = p^*.$$

Consequence of strong duality

If $x^* \in \mathbb{R}^d$ is the solution of primal problem $f_0(x) \in \mathbb{R}^d$ and $(\lambda^*, \nu^*) \in \mathbb{R}_+^{n_1} \times \mathbb{R}^{n_2}$ is the solution of dual problem $g(\lambda^*, \nu^*) = d^*$, we will have:

$$\begin{aligned} f(x^*) &= p^* = d^* = g(\lambda^*, \nu^*) \\ &= \min_{x \in \mathbb{R}^d} (f_0(x) + \sum_{i=1}^{n_1} \lambda_i f_i(x) + \sum_{j=1}^{n_2} \nu_j h_j(x)) \\ &\leq f_0(x^*) + \sum_{i=1}^{n_1} \underbrace{\lambda_i^*}_{\geq 0} \underbrace{f_i(x^*)}_{\leq 0} + \underbrace{\sum_{j=1}^{n_2} \nu_j^* h_j(x^*)}_{=0}. \end{aligned}$$

We deduce that:

$$\sum_{i=1}^{n_1} \lambda_i^* f_i(x^*) = 0 \Longrightarrow \qquad \underbrace{\forall i, \lambda_i^* f_i(x^*) = 0.}_{\textbf{The complementarity problem}}$$

So, we obtain:

- If $\lambda^* > 0$, then $f_i(x^*) = 0$. (*constraints saturation*)
- If $f_i(x^*) < 0$, then $\lambda^* = 0$.

Remark: The first order condition

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \nu^*) = 0$$

$$\Longleftrightarrow$$

$$\nabla_x f_0(x^*) + \sum_{i=1}^{n_1} \lambda_i^* \nabla_x f_i(x^*) + \sum_{j=1}^{n_2} \nu_j^* \nabla_x h_j(x^*) = 0.$$

Example

If we have $A \in \mathbb{R}^{n\times p},\ b \in \mathbb{R}^n, C \in \mathbb{R}^{r\times p}, d \in \mathbb{R}^r$, we need to resolve the least square problem:

$$\begin{aligned} \min_{x\in\mathbb{R}^p} \quad & \frac{1}{2}\left\|Ax-b\right\|^2 \\ \text{s.t.} \quad & Cx-d=0. \end{aligned}$$

We write the Langragian as:

$$\mathcal{L}(x,\nu) = \frac{1}{2}\left\|Ax-b\right\|^2 + \nu^\top(Cx-d).$$

We need to solve $\nabla_x\mathcal{L} = 0$:

$$\nabla_x\mathcal{L} = A^\top(Ax-b) + C^\top\nu = 0 \Longleftrightarrow A^\top Ax^* = A^\top b - C^\top\nu^*.$$

We obtain a linear system:

$$\begin{cases} Cx^* = d & \textit{(feasibility)} \\ A^\top Ax^* = A^\top b - c^\top\nu^* \end{cases}.$$

Table of Contents

Optimization of hovercraft trajectory - Presentation

We are in command of a hovercraft which must pass through k waypoints at certain times given. Our objective is to hit the waypoints at the prescribed times while minimizing fuel use. To do this, we need to introduce some notations :

- k is the number of waypoints ;
- $t = 0, 1, \ldots, T$ is the discretize time ;
- x_t is the hovercraft position at t time ;
- v_t is the velocity at the time t ;
- u_t is the thrust of hovercraft at the time t ;
- w_i is the waypoint.

Optimization of hovercraft trajectory - Resolution

First Model : hiting the waypoints exactly

We want to reach

$$\min_{x_t, v_t, u_t} \sum_{t=0}^{T} \|u_t\|^2,$$

under the constraints :

- $\forall t = 0, \ldots, T, \begin{cases} x_{t+1} = x_t + v_t \\ v_{t+1} = v_t + u_t \end{cases}$
- $x_0 = v_0 = 0$;
- $x_{t_i} = w_i \ \forall i = 1, \ldots, k.$

Optimization of hovercraft trajectory - Resolution

Second Model : Passing near waypoints

We want to reach

$$\min_{x_t, v_t, u_t} \sum_{t=0}^{T} \|u_t\|^2 + \lambda \sum_{t=0}^{T} \|x_{t_i} - w_i\|^2,$$

under the constraints :

- $\forall t = 0, \ldots, T,$

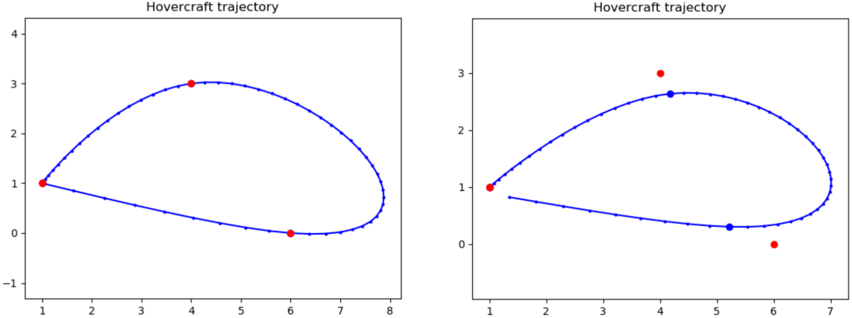
$$\begin{cases} x_{t+1} = x_t + v_t \\ v_{t+1} = v_t + u_t \end{cases} \quad (1)$$

- $x_0 = v_0 = 0$;
- $x_{t_0} = w_0$.

Here λ controls the tradeoff between making u small and hitting all the waypoints.

Optimization of hovercraft trajectory - Plot

Figure: Resolution thanks to the First Model (left) and the Second Model (right).



Moving Average reminder

Model

Moving average model is :

$$\forall t \in \{1, \ldots, T\},\ y_t = w_1 u_t + w_2 u_{t-1} + \cdots + w_k u_{t-k+1}$$

where :

- $(u_t)_{t \in 1 \ldots, T}$ is the time serie of input date ;
- $(y_t)_{t \in 1 \ldots, T}$ is the time serie of output date ;
- k is the size for which each output is a weighted combination of k previous inputs ;
- $(w_i)_{i \in 1 \ldots, k}$ is the weight of each input.

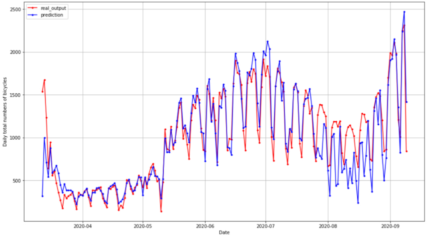
Moving Average on the dataset "Données comptages Totem"

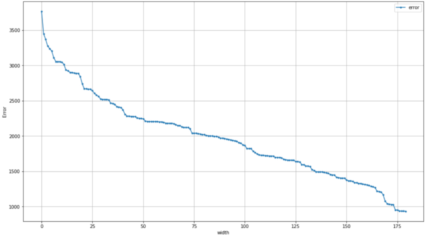
Thanks to this dataframe, we want to modelize the frequency, using of moving average, of passage of cyclists in front of the totem pole located at Place Albert 1er.
Our process to do this, is :

1. Count the number of cyclists passed per day ;
2. Modelize the moving average to modelize the frequency of passage of cyclists ;
3. Calculate error of prediction.

Moving Average on the dataset "Données comptages Totem" - Plot

Figure: Moving average to predict frequency of passage of cyclists (left) and prediction error (right).





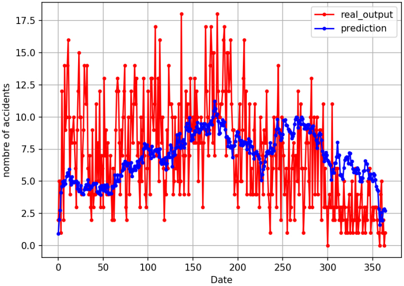
Moving Average on the dataset "Accidents vélos"

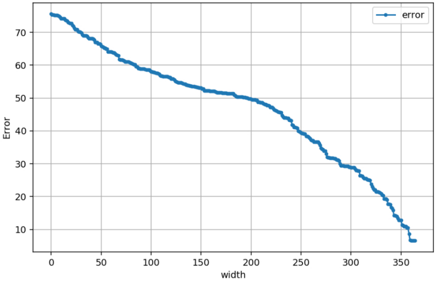
With the data "Accidents vélos", we want to predict, thanks to moving average calculate over the years 2005 to 2017, the frequency of bicycle accidents in 2018.
Our process to do this is :

1. Count the number of cyclists passed per day over the years of 2005 to 2017 (or 2016) ;
2. Modelize the moving average to predict the frequency of bike accidents in 2018 (or 2017) ;
3. Calculate error of prediction.

Moving Average on the dataset "Accidents vélos" - Plot

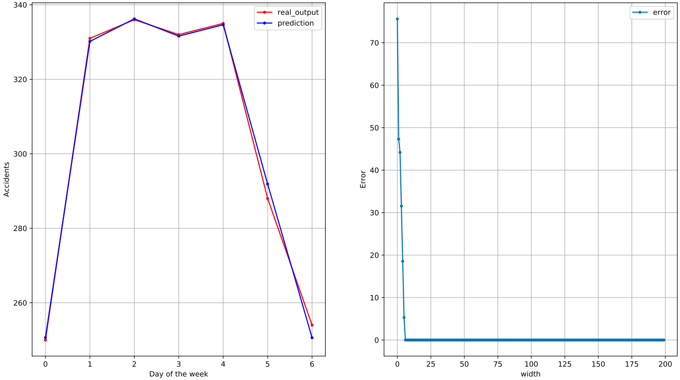
Figure: Moving average to predict frequency of bike accidents by day in 2018 (left) and prediction error (right).





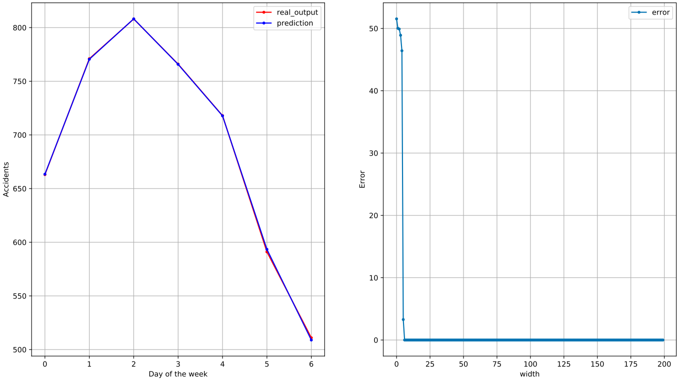
Moving Average on the dataset "Accidents vélos" - Plot

Figure: Moving average to predict frequency of bike accidents in 2018 by days of the week (left) and prediction error (right).



Moving Average on the dataset "Accidents vélos" - Plot

Figure: Moving average to predict frequency of bike accidents in 2017 by days of the week (left) and prediction error (right).



Bibliography

[1] Joseph Salmon, *Modéle linéaire avancé : Optimisation*, 2019, http://josephsalmon.eu/enseignement/Montpellier/HMMA307/Optimisation.pdf
[2] Laurent Lessard *Introduction to optimization*, 2017,