CO327 (2022Spring) Assignment 6

Lecturer: Andersen Ang

May 20, 2022

- Assignment deadline: July-3 23:55.
- Submit your electronic copy (in a single PDF) to the dropbox in Waterloo LEARN.

1 Computation (10 points)

Let $|a|$ denotes the absolute value of a, which is defined as

$$|a| = \begin{cases} a & \text{if } a >= 0, \\ -a & \text{if } a < 0. \end{cases}$$

Let $\max\{a,b\}$ denotes the maximum of the pair (a,b), which is defined as

$$\max\{a,b\} = \begin{cases} a & \text{if } a >= b, \\ b & \text{if } a < b. \end{cases}$$

Now, find the scalar x that is the minimizer of the following problem

$$\min_x \Big| 3 - \max\{2, x\} \Big|. \tag{P_0}$$

Your tasks

1. Introduce a new variable on P_0 to remove the max operator in the cost function, convert problem P to a new problem P_1, write down P_1.
2. Introduce a new variable on P_1 to remove the absolute value operator in the cost function, convert problem P_1 to a new problem P_2, write down P_2.
3. Turn problem P_2 into a LP in canonical form, denoted as P_3.
4. Solve P_3

Remember to show your working.

2 Robust curve fitting (20 points)

Introduction This part refers to `Q2_main.m` and `Q2.mat` in the data file.

A task in data analysis is to find the pattern(s) in data. In this question you will perform a simple data analysis known as "Regression". More specifically, "Least Sqaures" and "Robust regression'.

About the problem You are given n points (x_i, y_i), $i = 1, 2, \ldots, n$ which you believe they are generated under a model $y = mx + c$, where m, c are unknown parameters to be determined from the data. However, the data points are corrupted by noise and therefore the points are not exactly lying on a straight line.

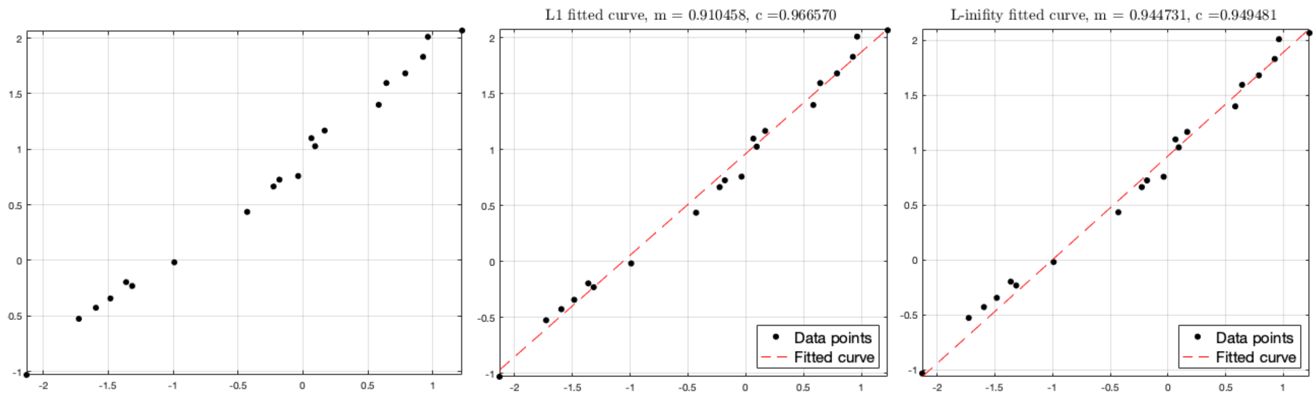

Figure 1: The input data and what you should get for different types of fitting.

For the data points, you propose the following model $y_i = mx_i + c + \epsilon_i$, $i = 1, 2, \ldots, n$ where ϵ refers to the noise. Refer to the lecture on curve fitting, a way to find (m, c) is to minimize the norm $\mathbf{e} = [\epsilon_1, \epsilon_2, \ldots, \epsilon_n]$. Mathematically, we have the following so-called "linear model":

$$\mathbf{Ax} = \mathbf{b} + \mathbf{e}, \quad \text{where} \quad \mathbf{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c \\ m \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Minimizing the norm of $\mathbf{e}$ is equivalent to minimizing $\|\mathbf{Ax} - \mathbf{b}\|_p$ for $p \in \{1, \infty\}$. That is,

$$\min_{x} \|\mathbf{e}\|_p \overset{\mathbf{e}=\mathbf{Ax}-\mathbf{b}}{=} \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_p = \min_{c,m} \left\| \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} c \\ m \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|_p. \tag{1}$$

Your task Write 2 MATLAB codes that solves (1) for $p = 1, \infty$. After you obtain (m, c), plot the result. For details, see `Q2_main.m` in the data file.

Hints

- Turn problem (1) for $p = 1, \infty$ as a LP and then solve these LP using `linprog`.
- When turning problem (1) for $p = 1, \infty$ as a LP, use the canonical form.
- Look at assignment 5 Q2 for hints when you constructing the constants $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}$ in MATLAB.
- If your solution is correct, your plot will look exactly the same as the one in Figure 1.

END of assignment.