

**2<sup>st</sup> Assignment: Mixed-Integer Linear Programming**

Generate random data, consider  $n=500$  observations ( $i=1,\dots,500$ ), from a predefined linear regression model with  $m=20$  variables ( $j=1,\dots,20$ ). Assume that the regression coefficients are integers so that  $-5 \leq \beta_j \leq 5$ . Assume also independent normal residuals.

$$Y = \beta'X + \epsilon$$

- a) The classical least squares approach is to find the values of vector  $\beta = (\beta_1, \dots, \beta_j, \dots, \beta_m)^T$  so that:

$$\min_{\beta} \sum_{i=1}^n (y_i - \beta'x_i)^2$$

where  $x_i = (x_{1i}, \dots, x_{2i}, \dots, x_{mi})^T$  for  $i = 1, \dots, n$ .

Estimate the value of the regression coefficients by using the analytical solution for the least squares estimation problem. Tip:

$$\beta_{ls} = (X^T X)^{-1} X^T Y$$

- b) As an alternative, the least absolute value approach seeks to find the values of  $\beta$  by solving the following problem:

$$\min_{\beta} \sum_{i=1}^n |y_i - \beta'x_i|$$

Propose and implement in Pyomo an equivalent linear formulation for this problem. Compare the resulting  $\beta$  coefficients with the ones obtained in a).

- c) Now assume that we want to impose the condition that only  $k$  factors (variables) affect the dependent variable  $Y$ . Extend the formulation in b) to a MILP (mixed integer linear optimization problem) to model the additional condition that up to  $k$  out of the  $\beta_j$  coefficients have nonzero values.
- d) Solve the problem in c) in Pyomo for  $k=1, \dots, 20$  and represent the behavior of the objective function with respect to  $k$ .
- e) Another possibility to find  $\beta$  is the robust linear regression problem that exploits the robustness of the error median against outliers. It can be formulated as:

$$\min_{\beta} \text{median}(|y_1 - \beta'x_1|, \dots, |y_n - \beta'x_n|)$$

Formulate this model as a MILP and implement and solve it in Pyomo. Due to its computational complexity, consider as an input  $n=30$  random observations ( $i=1,\dots,30$ ), from a linear regression model with  $m=5$  variables ( $j=1,\dots,5$ ).

Upload to campus global a pdf file with a resume of the implemented approaches and the solution. Additionally, upload as a separated file the Jupyter notebook (or python files) used for this assignment.