

IEOR 4004: Programming assignment 1

1. **(Linear regression)** This problem is based on the linear regression example in Lecture 1. It requires you to fit a linear function to data provided in *dataLR.txt*. The dataset contains $N = 10000$ rows and 11 columns (comma separated). The first column in every row i gives the label (y_i) and the remaining 10 columns are features (x_i). Formulate each of the following problems:

(a) find (w, b) that minimizes $\max_{i \in 1, \dots, N} |y_i - w^T x_i - b|$, and

(b) find (w, b) that minimizes $\sum_{i=1}^N |y_i - w^T x_i - b|$

as linear programs (LP), and solve using Gurobi.

2. **(Using duality)** Write duals of the LPs formulated for part (a) and part(b) of Question 1 and solve them using Gurobi. Verify that your dual optimal objective value matches the primal optimal objective value. Also, compare the time taken to solve the primal vs dual for part (b).

In your submission, write the primal and dual LP formulations, the optimal solution and objective values found using Gurobi, and submit the code for solving *both* primal and dual LPs.

Submit your written solutions as well as the code (printed as a single pdf). ONLINE SUBMISSIONS ON gradescope ONLY.

You have following options to complete this assignment:

Using Matlab: We have provided LR.mat file for Matlab users. Using `sload('LR.mat')` will populate the data into an array x and an array y , x being the 10000×10 array of features and y being the 10000 dimensional array of labels. After loading the data, you will need to write Matlab code to build LP model (into c, A, b etc.) and solve it using Gurobi as done in the Matlab examples provided on courseworks (e.g. look at `cashFlow.m`).

Using Python: Note: In `read.py`, we have provided a code snippet that can read the *dataLR.txt* file in order to populate arrays x and y , x being the 10000×10 array of features and y being the 10000 dimensional array of labels. After reading the data, you will need to build the LP model and solve it using Gurobi. Note that because of large number of variables and constraints, you cannot write the LP model file (*.lp) manually and then use `mygurobi.py` to solve the LP in this case. You have two options to build the LP model:

- Using dense matrix format: The python script *myMatrixLpSolver.py* provides a function *lp.optimize* that takes LP model in matrix notation (similar to Matlab) and solves it (it also automatically writes the lp file). For example, look at *cashFlowMatrix.py* example on courseworks.
- Constructing the lp model file directly: You can write python code to directly construct the lp model, instead of using dense matrix format. This method is more efficient and preferred in practice, however, it requires slightly more programming experience. Here is an example : http://www.gurobi.com/documentation/6.5/examples/diet_py.html, and here is a video tutorial explaining this example and some other examples <http://www.gurobi.com/resources/seminars-and-videos/modeling-with-the-gurobi-python-interface>.

Important: Please download all the files/scripts provided with this homework in your local directory where you are working.