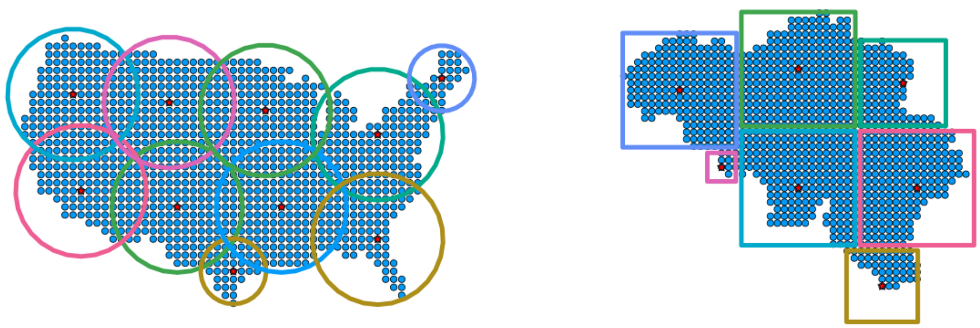
CSCI 5654, Spring 2023: Assignment 4

Assigned date: Tu 3/21/2023
Due date: Mo 4/10/2023

Instructions: Please upload your HW as a PDF file. IPs and MILPs can be solved using any tool of your choice but we recommend Julia, Python (pulp/cvxopt/gurobi), Matlab or GLPK.

Part 1: Optimal cover and partition

We consider the problem of optimal placement and sizing of facilities. More precisely, suppose that you need to decide where to place a set of fire stations or warehouses across the country and decide which range they should cover, while ensuring that the whole country is covered and minimizing the cost. See Figure 1 for illustrations.



(a) Optimal cover of the USA with circular regions. (b) Optimal partition of Belgium with square regions.

Figure 1: Examples of optimal cover and partition.

We will see, step by step, how to solve this problem using Integer Programming.

Problem 1 (7 points): Optimal cover and partition with circular regions

Let $D = \{(x_i, y_i)\}_{i=1}^N$ be a set of N points in $\mathbb{R}^2$ (e.g., the blue dots in Figure 1). Let us focus on circular regions for the moment. More specifically, the regions will be disks centered at one of the points in D, and their radius is assumed to take value in a finite set of nonnegative numbers $\{r_k\}_{k=1}^K$. Thus, there is a finite set of possible disks (indexed, by $i \in \{1, \dots, N\}$ for their center and $k \in \{1, \dots, K\}$ for their radius) and the goal is to cover the whole set D with some of these disks, while minimizing some cost function (depending on their radius; see later).

To decide which disks are taken, we associate a binary variable x_{ik} to each $i \in \{1, \dots, N\}$ and $k \in \{1, \dots, K\}$, with the interpretation that the disk with center (x_i, y_i) and radius r_k is taken if and only if $x_{ik} = 1$.

(a) Write the linear conditions on the binary variables $\{x_{ik}\}_{i,k=1}^{N,K}$ expressing that all points in D is covered by at least one taken disk (i.e., the taken disks form a cover of D).

(b) Write the linear conditions on the binary variables $\{x_{ik}\}_{i,k=1}^{N,K}$ expressing that all points in D is covered by one and only one taken disk (i.e., the taken disks form a partition of D).

Consider the following cost for building a facility: there is a fixed cost $c_f \geq 0$ for building the facility (i.e., does not apply if the facility is not built), and a variable cost $c_v r$ proportional to the radius r of the range covered by the facility, wherein $c_v \geq 0$.

(c) Express the cost function to minimize as a linear function of the binary variables $\{x_{ik}\}_{i,k=1}^{N,K}$.

Problem 2 (3 points): Implementation

Consider the sets of points given in the files `usa.txt` and `belgium.txt`.[1] For normalization, let the horizontal and vertical distance between neighboring points be equal to 1. For the set of values for the radii of the regions, let them be in $\{0, 1, 2, 3, 4, 5, 6\}$. Consider the cost function given by $1 + \frac{1}{2}r$, for each built facility with range radius r.

(a) Solve the problem of optimal cover with circular regions for the USA.

(b) Solve the problem of optimal partition with square regions[2] for Belgium.

For each problem, you can use the numerical solver of your choice.[3] Report the optimal cost, as well as the optimal solution in the form of plots similar to those in Figure 1. Upload your source codes as files `hw4_usa.*` and `hw4_belgium.*` respectively.

Part 2: Linear regression with outliers

We consider the problem of Linear Regression in the presence of outliers. See Figure 2 for an illustration. Therefore, we consider a *saturated* loss function $\ell(e) = \min\{|e|, \epsilon\}$, for some user-defined parameter $\epsilon > 0$. Given a set $D = \{(x_i, y_i)\}_{i=1}^N$ of N data points in $\mathbb{R}^2$ (e.g., the green dots in Figure 2), our goal is to find the parameters $a, b \in \mathbb{R}$ of a straight line $y = ax + b$ that minimizes that total error loss $\sum_{i=1}^N \ell(e_i)$, wherein $e_i = y_i - ax_i - b$.

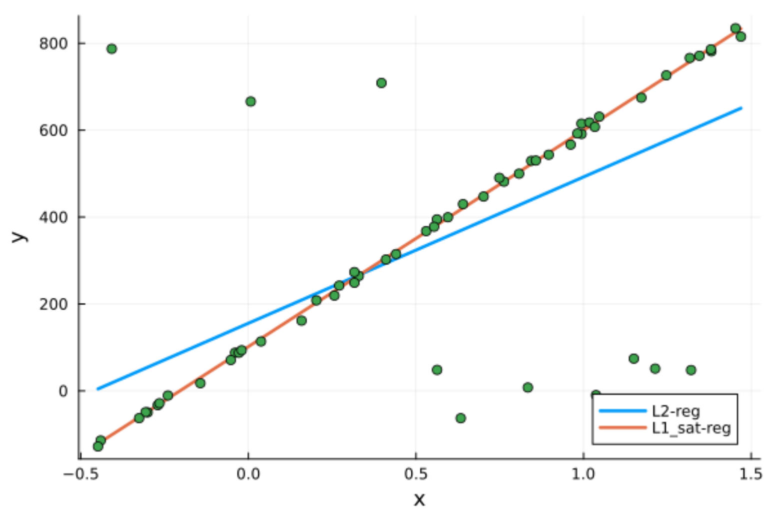


Figure 2: Linear Regressions with L^2 loss and saturated L^1 loss.

We will see, step by step, how to solve this problem using Mixed-Integer Linear Programming.

Problem 3 (7 points): MILP formulation

(a) Let u and v be two variables of an optimization program. Let $\epsilon > 0$ and $M \geq 0$ be two given constants, and assume that $0 \leq v \leq M$. Using an auxiliary binary variable β, write linear conditions on u, v and β expressing that $u \geq \min\{v, \epsilon\}$. Prove carefully the correctness of your condition (i.e., show that it is equivalent to $u \geq \min\{v, \epsilon\}$).

[1] Each set of points has the form of a boolean matrix with entry being 1 if and only if the point is taken. When plotting the points, you should obtain something as in Figure 1.

[2] The radius of a square region is defined as the half of its side length.

[3] I recommend Gurobi though.

(b) For each $i \in \{1, \ldots, N\}$, assume that $-M \leq e_i \leq M$ and let t_i be a variable. Write linear conditions expressing that for each $i \in \{1, \ldots, N\}$, $t_i \geq \ell(e_i)$. (Hint: you may need to introduce one or several auxiliary binary variables β_i's.)

(c) Provide a Mixed-Integer Linear Program solving the problem of Linear Regression with error loss $\sum_{i=1}^{N} \ell(e_i)$, as described above. You may assume that a constant M satisfying the conditions in (b) is given. In particular, describe precisely what are all the variables of your problem and whether they are continuous, integral or binary.

Problem 4 (3 points): Implementation

Consider the sets of points given in the file `linreg.txt`.[4] Let $\epsilon = 50$. Solve the problem of Linear Regression with error loss $\sum_{i=1}^{N} \ell(e_i)$, as described above, for this data set. As value for M, you can take the largest distance between any two y_i's, namely, $M = \max_{i,j} y_i - y_j$.

For each problem, you can use the numerical solver of your choice.[5] Report the optimal cost, as well as the optimal solution in the form of a plot similar to the one in Figure 2. Upload your source code as a file `hw4_linreg.*`.

[4] The first column corresponds to x_i's and the second column to y_i's. When plotting the points, you should obtain something as in Figure 2.

[5] I recommend Gurobi though.