Assignment 2: Constrained Optimization and the KKT Conditions

Jacob Puthipiroj

March 30, 2019

KKT Conditions for Linear Programming

A linear program can be expressed in canonical form as:

$$\min_x c^T x \qquad \text{subject to} \qquad Ax \leq b$$

for matrices $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^{m \times 1}$ and $c \in \mathbb{R}^{n \times 1}$.

The Lagrangian would be

$$\mathcal{L}(x, \lambda) = c^T x - \lambda(Ax - b)$$

Where λ is a vector of n values. The KKT Conditions for the LP are as follows:

- Primal feasibility: $g(x) = Ax - b \leq 0$
- Dual feasibility: $\lambda \geq 0$
- Complementary Slackness: $\lambda^T(Ax - b) = 0$
- Lagrangian Stationarity: $\nabla c^T x + \lambda^T \nabla (Ax - b) = 0$

Here, the complementary slackness KKT condition indicates that either $\lambda^T = 0$, in which case the Lagrange multiplier is active, and the solution for that dimensional lies within the interior of the feasible area, or $Ax - b = 0 \implies Ax = b$, in which case the constraint is active for said dimension, and optimal solution lies on the boundary.

For the optimal solution to lie completely within the interior of the feasible region would therefore be to require that $\lambda = 0$ (or even more precisely $\vec{\lambda} = 0$). However by the KKT condition of Langrangian Stationarity, this would also imply $\nabla c^T x = 0$, meaning that $c = 0$. The problem would then only be the trivial case. If the problem is not trivial, then there is some boundary line on which the optimal solution lies.

Expressing L_1 and L_∞ Regression Problems as Linear Programs

We have a set of points $(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$. Ideally we would have a line of the form $y = \theta_1 x + \theta_2$ that would allow us to perfectly line the set of all points, so that

$$\begin{matrix} \theta_1 x_1 + \theta_2 = y_1 \\ \theta_1 x_2 + \theta_2 = y_2 \\ \vdots \\ \theta_1 x_n + \theta_2 = y_n \end{matrix} \iff \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \iff X\Theta = Y \iff Y - X\Theta = 0$$

This is not always possible, so we perform regression by finding $\Theta = [\theta_1 \ \theta_2]^T$ such that $\|Y - X\Theta\|_1$ or $\|Y - X\Theta\|_\infty$ is minimized, depending on which type of regression we wish to perform. Both can be expressed as linear programming problems, as follows:

In the L_1 case, we have

$$\min_\Theta \|Y - X\Theta\|_1, \quad \text{where} \quad \|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

By definition of the L_1 norm $\|x\|_1 = \sum_{i=1}^n |x_i|$, and by introducing N we have

$$\begin{aligned} \underset{t}{\text{minimize}} \quad & \sum_{i=1}^n t_i \\ \text{subject to} \quad & |Y_i - X_i\Theta| \le t_i \quad \forall i \\ & t_i \ge 0 \quad \forall i \end{aligned}$$

We can also clean up the notation by getting rid of the absolute values.

$$\begin{aligned} \underset{t}{\text{minimize}} \quad & \sum_{i=1}^n t_i \\ \text{subject to} \quad & -t_i \le Y_i - X_i\Theta \le t_i \quad \forall i \\ & t_i \ge 0 \quad \forall i \end{aligned}$$

In the L_∞ case, we have

$$\min_\Theta \|Y - X\Theta\|_\infty, \quad \text{where} \quad \|x\|_\infty = \max\{|x_1|, |x_2|, \cdots, |x_n|\}$$

Again, we introduce the decision variable $t \in \mathbb{R}$ as

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & |Y_i - X_i\Theta| \le t, \quad \forall i \\ & t \ge 0. \end{aligned}$$

which can also be rewritten by introducing the $\mathbf{1}^T t$ vector and getting rid of the absolute values as

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & -\mathbf{1}^T t \preccurlyeq Y - X\Theta \preccurlyeq \mathbf{1}^T t, \\ & t \ge 0. \end{aligned}$$

Solving l_1 and l_∞ regression problems using CVXPY

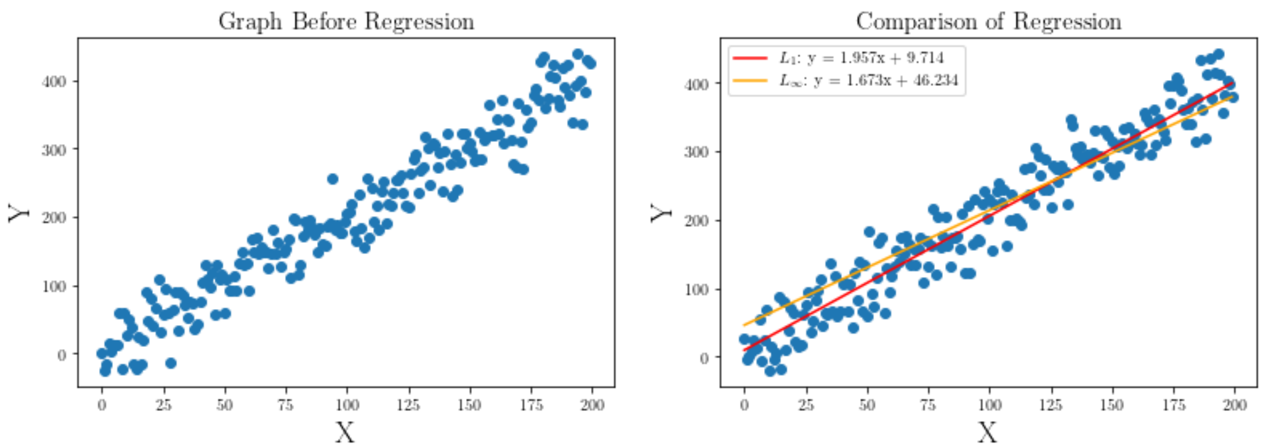

Figure 1: Using different types of regression produces entirely different regression lines. In general, after running the regression multiple times, L_1 regression was generally found to produce a line of greater slope that L_∞. As a result, the intercept of the L_∞ regression was usually higher than the intercept of L_1 regression.

In this section, we use CVXPY to perform both L_1 and L_∞ regression on a given dataset. The code used to generate the data, as well as the plots are available here.