

15.097 - Homework 1

Colin Pawlowski

March 2, 2016

1 Compressed Sensing

Assume \mathbf{A} is an $m \times n$ matrix, $n > m$. Consider the problem

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_0 \\ \text{s.t.} \quad & \mathbf{Ax} = \mathbf{b}. \end{aligned} \tag{1}$$

Using MIO, we reformulate this problem as

$$\begin{aligned} \min \quad & \sum_{i=1}^n z_i \\ \text{s.t.} \quad & \mathbf{Ax} = \mathbf{b}, \\ & -Mz_i \leq x_i \leq Mz_i \quad i = 1, \dots, n, \\ & z_i \in \{0, 1\} \quad i = 1, \dots, n. \end{aligned} \tag{2}$$

Next, consider the problem

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_1 \\ \text{s.t.} \quad & \mathbf{Ax} = \mathbf{b}. \end{aligned} \tag{3}$$

Using linear optimization, we can reformulate this problem as

$$\begin{aligned} \min \quad & \sum_{i=1}^n y_i \\ \text{s.t.} \quad & \mathbf{Ax} = \mathbf{b}, \\ & -y_i \leq x_i \leq y_i \quad i = 1, \dots, n. \end{aligned} \tag{4}$$

We implemented both optimization problems 2 and 4 in JuMP. For the simulated experiments, we generated a random 20×20 matrix \mathbf{M} with entries $a_{ij} \sim N(0, 1)$ *i.i.d.*. For a fixed triplet (k, m, n) , we solved problems 2 and 4, where \mathbf{A} is the upper $m \times n$ submatrix of \mathbf{M} , $\mathbf{x}_0 \in \{0, 1\}^n$ is the vector with the first k components 1 and the rest 0, and $\mathbf{b} = \mathbf{A}\mathbf{x}_0$. We repeated this experiment for all possible combinations $k \leq m \leq n = 20$, and tracked the number of successfully recovered components of \mathbf{x}_0 recovered by each method.

We found that the MIO formulation always found a solution with at most k nonzero components, and in some cases found solutions which were even more sparse than \mathbf{x}_0 due to numerical approximations. On the other hand, the LO formulation correctly recovered the \mathbf{x}_0 solution sometimes, but in many cases LO yielded solutions with greater than k nonzero components. For the LO formulation, we plot the results below in a phase diagram with axes m/n and k/m , indicating whether or not the correct sparsity pattern was recovered at each data point.

2 Algorithmic Framework for Regression using MIO

3 First Order Method

Here, we derive a first order method following the notes from Lecture 2 - Best Subset Selection. Consider the problem

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & g(\boldsymbol{\beta}) := \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \Gamma\|\boldsymbol{\beta}\|_1 \\ \text{s.t.} \quad & \|\boldsymbol{\beta}\|_0 \leq k. \end{aligned} \tag{5}$$

Since $g(\boldsymbol{\beta})$ is convex and $\|\nabla g(\boldsymbol{\beta}) - \nabla g(\boldsymbol{\beta}_0)\| \leq \ell\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|$, it follows that for all $L \geq \ell$

$$g(\boldsymbol{\beta}) \leq Q(\boldsymbol{\beta}) := g(\boldsymbol{\beta}_0) + \nabla g(\boldsymbol{\beta}_0)^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \frac{L}{2}\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^2 + \Gamma\|\boldsymbol{\beta}\|_1. \tag{6}$$

To find feasible solutions, we solve the following problem

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & Q(\boldsymbol{\beta}) \\ \text{s.t.} \quad & \|\boldsymbol{\beta}\|_0 \leq k. \end{aligned} \tag{7}$$

This is equivalent to

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & \frac{L}{2} \left\| \boldsymbol{\beta} - \left(\boldsymbol{\beta}_0 - \frac{1}{L} \nabla g(\boldsymbol{\beta}_0) \right) \right\|_2^2 - \frac{1}{2L} \|\nabla g(\boldsymbol{\beta}_0)\|_2^2 + \Gamma \|\boldsymbol{\beta}\|_1 \\ \text{s.t.} \quad & \|\boldsymbol{\beta}\|_0 \leq k, \end{aligned} \tag{8}$$

which reduces to the following plus a constant term:

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & \frac{L}{2} \|\boldsymbol{\beta} - \mathbf{u}\|_2^2 + \Gamma \|\boldsymbol{\beta}\|_1 \\ \text{s.t.} \quad & \|\boldsymbol{\beta}\|_0 \leq k. \end{aligned} \tag{9}$$

For the vector $\mathbf{u} \in \mathbb{R}^p$, let $(1), (2), \dots, (p)$ be the indices of the order statistics $|u_{(1)}| \geq |u_{(2)}| \geq \dots \geq |u_{(p)}|$. At the optimal solution $\boldsymbol{\beta}^*$ to problem 9, we have $|\beta_{(1)}^*| \geq |\beta_{(2)}^*| \geq \dots \geq |\beta_{(p)}^*|$, which implies that $|\beta_{(k+1)}^*| = |\beta_{(k+2)}^*| = \dots = |\beta_{(p)}^*| = 0$. For $i \leq k$, $\beta_{(i)}^*$ is the optimal solution to the following unconstrained single variable problem:

$$\min_{\beta_{(i)}} \frac{L}{2} (\beta_{(i)} - u_{(i)})^2 + \Gamma |\beta_{(i)}|. \tag{10}$$

Problem 10 has closed form solution

$$\beta_{(i)}^* = \begin{cases} u_{(i)} - \frac{\Gamma}{L} \text{sign}(u_{(i)}), & \text{if } |u_{(i)}| \geq \frac{\Gamma}{L}, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Thus, the optimal solution to problem 9 is $\beta^* = \mathbf{H}_k(\mathbf{u})$, where

$$(\mathbf{H}_k(\mathbf{u}))_i = \begin{cases} u_{(i)} - \frac{\Gamma}{L} \text{sign}(u_{(i)}), & \text{if } |u_{(i)}| \geq \frac{\Gamma}{L} \text{ and } i \leq k, \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

Using this update iteratively to determine the β_i 's, we obtain the following first order method:

Algorithm 1

Input: $g(\beta), L, \epsilon$.

Output: A first order stationary solution β^* .

1. Initialize with $\beta_1 \in \mathbb{R}^p$ such that $\|\beta_1\|_0 \leq k$.
2. For $m \geq 1$

$$\beta_{m+1} \leftarrow \mathbf{H}_k(\beta_0 - \frac{1}{L} \nabla g(\beta_0))$$

3. Repeat Step 2, until $g(\beta_m) - g(\beta_{m+1}) \leq \epsilon$.