Homework 1

Ying Daisy Zhuo

March 2, 2016

1 Problem One

I implemented the problem in JuMP. The formulations are as follows: Original problem (L_0):

$$
\begin{aligned}
\max \quad & \sum_{i=1}^{n} z_i \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\
& |x_i| \leq M(1 - z_i) \quad \forall i = 1, 2, \ldots, n \\
& z_i \in \{0, 1\} \quad \forall i = 1, 2, \ldots, n
\end{aligned} \tag{1}
$$

L_1 relaxation problem:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n} x_i^+ + x_i^- \\
\text{s.t.} \quad & \mathbf{A}(\mathbf{x}^+ - \mathbf{x}^-) = \mathbf{b} \\
& x_i^+ \geq 0 \quad \forall i = 1, 2, \ldots, n \\
& x_i^- \geq 0 \quad \forall i = 1, 2, \ldots, n
\end{aligned} \tag{2}
$$

For fixed $n = 100$, we solve the problem for various $m = 1, 10, \ldots, 100$ and $k = 1, 5, \ldots, m$. If the L_1 relaxation solves the problem and gets the same sparse solution as the solution to the true L_0 problem, then we indicate it as "recovered". Figure 1 depicts whether the problem is recovered based on the ratio between m and n on the x-axis, and the ratio between k and m on the y-axis.

Observe that along the diagonal line, it separates the region into two areas. In the one to the lower right (with high m/n and low k/m), L_1 can recover the true solution; in the one to the upper left (with low m/n and high k/m), L_1 is unable to recover the true solution.

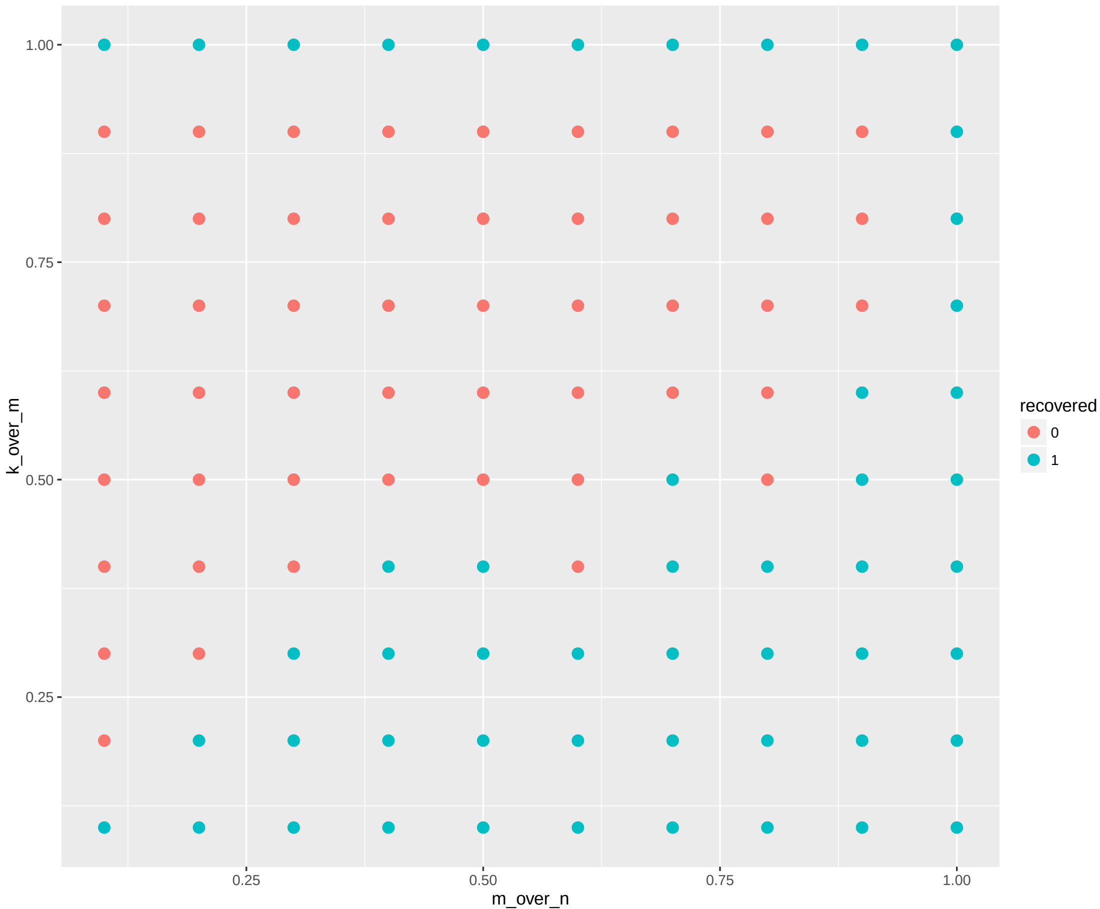

Figure 1: Plot of whether the sparse solution is recovered in the L_1 relaxation problem based on m/n and k/m.

2 Problem Two

In this problem, we are asked to formulate regression problems with some additional properties using MIO. In particular, some desirable properties include:

1. Pairwise multi-collinearity
2. Group sparsity
3. Sparsity
4. Robustness
5. Non-linear transformation
6. Statistical significance

To do so, we need some pre-processing. First, we expand the dataset to include non-linear transformations. For this example, we include x^2 and $\log(x)$ for each of the data dimensions. In addition, we compute the correlation matrix and obtain pairs $\{i, j\}$ of covariates that have correlation greater than a threshold, say 0.8, and save the set of pairs $\mathcal{HC}$.

We can now implement the first five items as the following problem:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n} \frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\| + \Gamma\|\boldsymbol{\beta}\|_1 & \\
\text{s.t.} \quad & -Mz_d \le \boldsymbol{\beta}_d \le Mz_d & \forall d = 1, 2, \ldots, D \\
& \sum_{d=1}^{D} z_d \le k & \\
& z_i + z_j \le 1 & \forall \{i, j\} \in \mathcal{HC} \\
& \sum_{i \in \mathcal{T}_m} z_i \le 1 & \forall m \\
& z_d \in \{0, 1\} & \forall d = 1, 2, \ldots, D
\end{aligned}
\tag{3}
$$

The implementation of statistical significance is done in a iterative fashion. At we solve the initial Problem 5, we run a simple linear model with all the variables selected and see if all variables are statistically significant. If so, do not add any constraint and terminate. Otherwise, add the following constraint: $\sum z_i \le \mathit{current\_count} - 1$ and solve the problem. Iterate until termination.

The parameters Γ and k can be tuned using cross validation. Using the *lpga2009* dataset, we obtain an optimal solution with $\beta_1 1 = -11.395$ and the rest being zero. This give an out-of-sample R^2 of 0.80.

3 Problem Three

Using the same technique from class, we find the following majorization of the first term:

$$g(\boldsymbol{\beta}) \leq Q(\boldsymbol{\beta}) = g(\boldsymbol{\beta}_0) + \nabla g(\boldsymbol{\beta}_0)^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \frac{L}{2} \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2, \quad \forall L \geq l, \tag{4}$$

where l is the Lipschitz constant on the first order derivative of g. We can solve the following problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} \frac{L}{2} \|\boldsymbol{\beta} - \mathbf{u}\|^2 + \Gamma \|\boldsymbol{\beta}\|_1 - \frac{1}{2L} \|\nabla g(\boldsymbol{\beta}_0)\|^2 \\ \text{s.t.} \quad & \|\boldsymbol{\beta}\| \leq k, \end{aligned} \tag{5}$$

where $u_i = \boldsymbol{\beta}_0 - \frac{1}{L} \nabla g(\boldsymbol{\beta}_0)$. Notice that the last term in the objective is a constant, and the rest of the objective term is separable. Therefore, for each i,

$$Q_i = \min \quad \frac{L}{2} (\beta_i - u_i)^2 + \Gamma |\beta_i|^2 \tag{6}$$

To find the optimal solution to the above problem, we look at the following scenarios:

1. When $u_i \geq \Gamma/L$, optimality is achieved at $\beta_i = \Gamma/L - u_i$.
2. When $u_i \leq -\Gamma/L$, optimality is achieved at $\beta_i = \Gamma/L + u_i$.
3. When $|u_i| \leq \Gamma/L$, $beta_i = 0$ achieves optimality.

We are now able to compute the objective Q_i for each i. We then sort the Q_i's and pick the k lowest values.