ST790-003, Homework 7

Due Tuesday, Apr 21, 2015 @ 11:59PM

Sparse Regression

This homework explores different algorithms for solving lasso penalized linear regression. The goal is to (1) know the limit of convex optimization softwares and appreciate the scalability of first order methods, (2) understand coordinate descent and (accelerated) proximal gradient methods and appreciate their implementation details, and (3) fine tune your coding ability to deal with big data. This is a *solo* homework. Discussion with fellow students is allowed but you have to write your code and report independently.

1. Implement lasso penalized linear regression using three methods: (a) a convex optimization solver (you already did this in HW5), (b) coordinate descent (CD), and (c) accelerated proximal gradient method (FISTA). You can use any language of your choice. But keep in mind this homework is scored mostly by the correctness and efficiency of your implementation.

2. We are going to generate a fake data set to test your algorithms. First form a square matrix $\boldsymbol{H} \in \mathbf{R}^{n \times n}$ of form

$$\boldsymbol{H} = \begin{pmatrix} n^{-1/2} & 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{12} & \cdots & 1/\sqrt{n(n-1)} \\ n^{-1/2} & -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{12} & \cdots & 1/\sqrt{n(n-1)} \\ n^{-1/2} & 0 & -2/\sqrt{6} & 1/\sqrt{12} & \cdots & 1/\sqrt{n(n-1)} \\ n^{-1/2} & 0 & 0 & -3/\sqrt{12} & \cdots & 1/\sqrt{n(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ n^{-1/2} & 0 & 0 & \cdots & \cdots & -(n-1)/\sqrt{n(n-1)} \end{pmatrix}.$$

   Our design matrix $\boldsymbol{X} \in \mathbf{R}^{n \times p}$ will be the first p columns of $\boldsymbol{H}$. The true regression coefficients will be $\boldsymbol{\beta} = (1, 2, 3, 4, 5, 0, \ldots, 0)^T$. Response vector is generated from model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

   where $\boldsymbol{\epsilon}$ are iid standard normal errors. Set random seed at 2015790003 and generate a specific data set $(\boldsymbol{X}, \boldsymbol{y})$ with $n = 10,000$ and $p = 1,000$, which is to be used in Q4.

3. Derive an analytical solution to the lasso problem

$$\hat{\boldsymbol{\beta}}_{\text{lasso}}(\lambda) = \arg\min \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda \sum_{j=2}^{p} |\beta_j|$$

   for data $(\boldsymbol{X}, \boldsymbol{y})$ of the specific form in Q2. That is to express $\hat{\beta}_{\text{lasso}}(\lambda)$ in terms of $\boldsymbol{y}$, $\boldsymbol{X}$, and λ. We will use this analytical lasso solution to gauge the correctness of your algorithms.

4. Test your algorithms on the data set generated in Q2. For ease of grading, use the grid points $\lambda = 0, 0.6, 1.2, \ldots, 6$ and a tolerance of 10^{-4} in the relative change of objective values for the CD and proximal gradient method.

For each of the three methods (convex optimization solver, coordinate descent, and proximal gradient method), please

(a) plot solution path,

(b) compare your solution to the analytical solution derived in Q3 and report the worst error $\max_i \|\hat{\boldsymbol{\beta}}(\lambda_i) - \hat{\boldsymbol{\beta}}_{\text{lasso}}(\lambda_i)\|_2$ along the solution path, and

(c) report the run time for the whole solution path on the teaching server.