Coordinate Descent for SLOPE

[1,*]

[1] *Department of Statistics, Lund University*
[*] *Corresponding author:*

February 15, 2022

Abstract

1 Introduction

Sorted L-One Penalized Estimation (SLOPE) [Bog+13; Bog+15] is a type of sparse regression represented by the convex optimization problem

$$\text{minimize}_{\beta \in \mathbb{R}^p} f(\beta; \lambda) = g(\beta) + J(\beta; \lambda) \tag{1}$$

where we take $g(\beta)$ to be smooth and twice differentiable and

$$J(\beta; \lambda) = \sum_{j=1}^{p} \lambda_j |\beta_{(j)}| \tag{2}$$

is the *sorted ℓ_1 norm*, defined such that

$$|\beta_{(1)}| \geq |\beta_{(2)}| \geq \cdots \geq |\beta_{(p)}|$$

and

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq 0.$$

2 Theory

2.1 Directional Derivatives

2.1.1 The Sorted ℓ_1 Norm

Theorem 2.1. *Let $h_0 \in \big(0, \min_{i,j \in \{i \mid \beta_i \neq 0\}} \big||\beta_i| - |\beta_j|\big|\big]$ and define $(\cdot)^*$ to be the permutation such that*

$$|\beta + h_0 v|_{(1)^*} \geq |\beta + h_0 v|_{(2)^*} \geq \cdots \geq |\beta + h_0 v|_{(p)^*}.$$

The directional derivative for the sorted ℓ_1 norm, $J(\beta, \lambda)$, is

$$D_v J(\beta; \lambda) = \sum_i \sum_{j \in \mathcal{B}_i} \lambda_j v_{(j)^*} \operatorname{sign}(\beta_{(j)^*} + h_0 v_{(j)^*})$$

where

$$\mathcal{B}_i = \{j \mid |\beta_j| = |\beta_i|\}.$$

JL

We need something that doesn't define clusters multiple times.

Proof. The directional derivative for the sorted ℓ_1 norm and a direction v with $\|v\| = 1$ is

$$\begin{aligned} D_v J(\beta, \lambda) &= \lim_{h \searrow 0} \frac{J(\beta + hv; \lambda) - J(\beta; \lambda)}{h} \\ &= \lim_{h \searrow 0} \frac{\sum_{j=1}^{p} \lambda_j \big(|\beta + vh|_{(j)} - |\beta|_{(j)}\big)}{h} \\ &= \lim_{h \searrow 0} \frac{\sum_i \sum_{j \in \mathcal{B}_i} \lambda_j \big(|\beta + vh|_{(j)} - |\beta|_{(j)}\big)}{h} \end{aligned} \tag{3}$$

First, assume that that is an i such that $\operatorname{card} \mathcal{B}_i \neq 0$ and $\beta_i = 0$. Then

$$\begin{aligned} \sum_{j \in \mathcal{B}_i} \frac{\lambda_j \big(|\beta + vh|_{(j)} - |\beta|_{(j)}\big)}{h} &= \sum_{j \in \mathcal{B}_i} \lambda_j \operatorname{sign}(v)_{(j)} v_{(j)} \\ &= \sum_{j \in \mathcal{B}_i} \lambda_j \operatorname{sign}(\beta + hv)_{(j)^*} v_{(j)^*}, \end{aligned}$$

where the last equality follows from the fact that $\operatorname{sign}(\beta + hv) = \operatorname{sign}(\beta)$ and $(i) = (i)^*$ for all i whenever $h > 0$ and $\beta \neq 0$.

Next, for each i such that $\operatorname{card} \mathcal{B}_i \neq 0$ and $\beta_i \neq 0$, observe that $\operatorname{sign}(\beta + hv) = \operatorname{sign}(\beta)$ and $(i)^* = (i)$ whenever $0 < h < h_0$, recalling the construction of h_0. Therefore, we have

$$\sum_{j \in \mathcal{B}_i} \frac{\lambda_j \big(|\beta + hv|_{(j)} - |\beta|_{(j)}\big)}{h} = \sum_{j \in \mathcal{B}_i} \lambda_j \operatorname{sign}(\beta + vh)_{(j)^*} v_{(j)^*}.$$

From this, we see that (3) reduces to

$$\lim_{h \searrow 0} \sum_i \sum_{j \in \mathcal{B}_i} \lambda_j \operatorname{sign}(\beta + vh)_{(j)^*} v_{(j)^*} = \sum_i \sum_{j \in \mathcal{B}_i} \lambda_j \operatorname{sign}(\beta + vh_0)_{(j)^*} v_{(j)^*}.$$

□

3 Experiments

4 Discussion

References

[1] Małgorzata Bogdan et al. "SLOPE – Adaptive Variable Selection via Convex Optimization". In: *The annals of applied statistics* 9.3 (2015), pp. 1103–1140. ISSN: 1932-6157. DOI: 10.1214/15-AOAS842. pmid: 26709357. URL: https://projecteuclid.org/euclid.aoas/1446488733 (visited on 12/17/2018).

[2] Małgorzata Bogdan et al. *Statistical Estimation and Testing via the Sorted L1 Norm*. Oct. 29, 2013. arXiv: 1310.1969 [math, stat]. URL: http://arxiv.org/abs/1310.1969 (visited on 04/16/2020).

A Proofs