The Fastest $\ell_{\infty,1}$ Prox in the West

Béjar, Benjamín and Dokmanić, Ivan and Vidal, René

汇报人: Apple Zhang

Shenzhen University

2022年9月13日

目录

1. Our Problem
2. Preliminary
3. Methodology
4. Experiments

目录

1. Our Problem
2. Preliminary
3. Methodology
4. Experiments

Sparse regression with $L_{\infty,1}$ constraints

求解以下的回归模型[**?**]:

$$\min_{\boldsymbol{W}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{W}\|_2^2, \quad \text{s.t. } \|\boldsymbol{W}\|_{\infty,1} \le \tau. \tag{1}$$

Recall $L_{p,q}$-norm

$L_{p,q}$ 范数的定义

$$\|\boldsymbol{W}\|_{p,q} = \left(\sum_{i=1}^{m} \|\boldsymbol{w}^{(i)}\|_p^q \right)^{1/q} \tag{2}$$

我们常用的 $L_{2,1}$ 范数:

$$\|\boldsymbol{W}\|_{2,1} = \sum_{i=1}^{m} \|\boldsymbol{w}^{(i)}\|_2 \tag{3}$$

本文将用到的 $L_{\infty,1}$、$L_{1,\infty}$ 范数:

$$\|\boldsymbol{W}\|_{\infty,1} = \sum_{i=1}^{m} \|\boldsymbol{w}^{(i)}\|_\infty = \sum_{i=1}^{m} \max_{j=1,2,\cdots,c} |w_{ij}|, \tag{4}$$

$$\|\boldsymbol{W}\|_{1,\infty} = \max_{i=1,2,\cdots,m} \|\boldsymbol{w}^{(i)}\|_1. \tag{5}$$

目录

1 Our Problem

2 Preliminary

3 Methodology

4 Experiments

KKT conditions for convex problems

对于以下优化问题，其中 $f_i(\boldsymbol{x}),\ i=0,1,\cdots,m$ 均为凸函数.

$$\min_{\boldsymbol{x}} f_0(\boldsymbol{x}),\quad \text{s.t. } f_i(\boldsymbol{x}) \le 0\ (i=1,2,\cdots,m),\quad \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}. \tag{6}$$

拉格朗日函数

$$L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \alpha_i f_i(\boldsymbol{x}) + \boldsymbol{\beta}^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}).$$

其 KKT 条件为

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*), \tag{7}$$

$$f_i(\boldsymbol{x}^*) \le 0,\quad \boldsymbol{A}\boldsymbol{x}^*=\boldsymbol{b},\quad \alpha_i^* \ge 0,\quad \alpha_i^* f_i(\boldsymbol{x}^*) = 0. \tag{8}$$

当然，如果 $f_i(\boldsymbol{x}),\ i=0,1,\cdots,m$ 全部可微，第一个条件就写成

$$\nabla_{\boldsymbol{x}} L = \nabla_{\boldsymbol{x}} f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \alpha_i \nabla_{\boldsymbol{x}} f_i(\boldsymbol{x}) - \boldsymbol{A}^\top \boldsymbol{\beta} = \boldsymbol{0}. \tag{9}$$

Proximal operator

对于可微凸函数 $f(\cdot)$，其近端算子 (proximal operator) 定义为

$$\operatorname{prox}_f(\boldsymbol{x}) = \arg\min_{\boldsymbol{z}} f(\boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2$$

实际含义: 使得 $f(\boldsymbol{z})$ 最小化，但尽量靠近 $\boldsymbol{x}$ 的最优解.

Projection: a special case of proximal operator

对于非空凸集 $\mathcal{C}$, 向量 $\boldsymbol{x}$ 到该集合的投影 (projection) 为

$$\pi_{\mathcal{C}}(\boldsymbol{x}) = \arg\min_{\boldsymbol{z}\in\mathcal{C}} \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2. \tag{10}$$

事实上，投影就是特定的近端算子:

$$\pi_{\mathcal{C}}(\boldsymbol{x}) = \mathrm{prox}_{\mathbb{I}(\cdot\in\mathcal{C})}(\boldsymbol{x}) = \arg\min_{\boldsymbol{z}} \mathbb{I}(\boldsymbol{z}\in\mathcal{C}) + \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2. \tag{11}$$

其中 $\mathbb{I}(\cdot\in\mathcal{C})$ 为示性函数 (indicator function):

$$\mathbb{I}(\boldsymbol{z}\in\mathcal{C}) = \begin{cases} 0, & \boldsymbol{z}\in\mathcal{C}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{12}$$

Example: projection onto of L_1-ball

考虑 L_1 球的投影问题:

$$\pi_{\|\cdot\|_1 \leq \lambda}(\boldsymbol{x}) = \arg\min_{\boldsymbol{z}} \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2, \quad \text{s.t. } \|\boldsymbol{z}\|_1 \leq \lambda. \tag{13}$$

拉格朗日函数

$$L(\boldsymbol{z},\alpha) = \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2 + \alpha(\|\boldsymbol{z}\|_1 - \lambda) \tag{14}$$

$$= \left[\frac{1}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2 + \alpha\|\boldsymbol{z}\|_1\right] - \alpha\lambda. \tag{15}$$

根据 **KKT** 条件可以得到

$$\boldsymbol{z} = \arg\min_{\boldsymbol{z}} L(\boldsymbol{z},\alpha) = \operatorname{prox}_{\alpha\|\cdot\|_1}(\boldsymbol{x}), \tag{16}$$

$$\alpha \geq 0, \quad \alpha(\|\boldsymbol{z}\|_1 - \lambda) = 0. \tag{17}$$

Example: projection onto of L_1-ball

分类讨论:

- 如果 $\|\boldsymbol{z}\|_1 < \lambda$, 则必有 $\alpha = 0$, 将推出 $\boldsymbol{z} = \mathrm{prox}_{0\|\cdot\|_1}(\boldsymbol{x}) = \boldsymbol{x}$.
- 如果 $\|\boldsymbol{z}\|_1 = \lambda$, 因 $\boldsymbol{z} = \mathrm{prox}_{\alpha\|\cdot\|_1}(\boldsymbol{x})$, 即 $z_i = \mathrm{sign}(x_i)[|x_i| - \alpha]_+$, 则

$$\|\boldsymbol{z}\|_1 = \sum_{i=1}^{n}[|x_i| - \alpha]_+ = \lambda. \tag{18}$$

根据以上方程求出 α，即可求出 $\boldsymbol{z}$.

Example: projection onto of L_1-ball

Algorithm 1 $O(n\log n)$ Algorithm for $\pi_{\|\cdot\|_1\le\lambda}$

Require: $\boldsymbol{x}\in\mathbb{R}^n, \lambda>0$.

Ensure: $\boldsymbol{z}=\pi_{\|\cdot\|_1\le\lambda}(\boldsymbol{x})$.

1: **if** $\|\boldsymbol{x}\|_1\le\lambda$ **then**

2: $\quad\boldsymbol{z}\leftarrow\boldsymbol{x}$.

3: **else**

4: $\quad\boldsymbol{u}\leftarrow\mathrm{sort}(\boldsymbol{x}, \texttt{`descend'})$,

5: $\quad\rho\leftarrow\max\{j=1,2,\cdots,n\mid u_j-(\sum_{r=1}^{j}u_r-\lambda)/j>0\}$,

6: $\quad\alpha\leftarrow(\sum_{r=1}^{\rho}u_i-\lambda)/\rho$,

7: $\quad z_i\leftarrow\mathrm{sign}(x_i)[u_i-\alpha]_+,\quad\forall i=1,2,\cdots,n$.

8: **end if**

目录

1 Our Problem

2 Preliminary

3 Methodology

4 Experiments

Projected gradient descent

$$\min_{\boldsymbol{W}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{W}\|_2^2, \quad \text{s.t. } \|\boldsymbol{W}\|_{\infty,1} \le \tau. \tag{19}$$

迭代更新公式:

$$\boldsymbol{V} \leftarrow \boldsymbol{W} - \eta \nabla_{\boldsymbol{W}} J(\boldsymbol{W}), \tag{20}$$

$$\boldsymbol{W} \leftarrow \pi_{\|\cdot\|_{\infty,1} \le \tau}(\boldsymbol{V}). \tag{21}$$

问题在于 $\pi_{\|\cdot\|_{\infty,1} \le \tau}(\boldsymbol{V})$ 的计算.

Projection and proximal operator

引理

对任意 $\tau > 0$,

$$\mathrm{prox}_{\tau\|\cdot\|_{1,\infty}}(\boldsymbol{V}) + \pi_{\|\cdot\|_{\infty,1}\leq\tau}(\boldsymbol{V}) = \boldsymbol{V} \tag{22}$$

因此问题可以转化为求解 $\mathrm{prox}_{\tau\|\cdot\|_{1,\infty}}(\boldsymbol{V})$，即

$$\mathrm{prox}_{\tau\|\cdot\|_{1,\infty}}(\boldsymbol{V}) = \arg\min_{\boldsymbol{W}} \tau\|\boldsymbol{W}\|_{1,\infty} + \frac{1}{2}\|\boldsymbol{W} - \boldsymbol{V}\|_2^2. \tag{23}$$

回忆一下:

$$\|\boldsymbol{W}\|_{1,\infty} = \max_{i=1,2,\cdots,m} \|\boldsymbol{w}^{(i)}\|_1. \tag{24}$$

目录

1 Our Problem

2 Preliminary

3 Methodology

4 Experiments

$L_{\infty,1}$ v.s. $L_{2,1}$?!

Datasets	RFS ($L_{2,1}$)	$L_{\infty,1}$-regression	SVM
DNA	91.69±0.97	93.60±0.59	90.51±1.31
Binaryalpha	55.27±2.56	57.53±1.37	62.81±1.81
USPS	88.10±0.51	88.26±0.38	92.10±0.59

谢谢!