Learning with Sparsity

Zhenwei Lin

日期：2020 年 8 月 16 日

摘　要

本文主要为处罚函数的介绍，处罚函数本质上就是对其的先验认识

关键词:

1　Group LASSO(成块收缩为 0)

$$\min\left\{\frac{1}{2N}\sum_{i=1}^{N}(y_i - z_i^T\theta)^2 + \lambda\sum_{j=1}^{J}\left\|\theta_j\right\|_2\right\}$$

1.1　Algorithm: Block coordinate Descent

Given $\theta_1, \theta_2, \cdots, \theta_J$,assume $\theta_2, \cdots, \theta_J$ are known, denote $r_1 = y - \sum_{i=2}^{J} z_j^T\theta_j$

$$\min_{\theta_1}\left\{\frac{1}{2N}\sum_{i=1}^{N}(r_{1i} - z_i^T\theta_1)^2 + \lambda\left\|\theta_1\right\|_2\right\}$$

引理 1.1. Suppose $f(\theta) = \|\theta\|_2, \quad \theta \in R^d$

1. if $\theta \neq 0, \quad \nabla f(\theta) = \frac{\theta}{\|\theta\|_2}$
2. if $\theta = 0, \partial f(0) = \{v \mid \|v\|_2 \leq 1\}$

证明.

$$f(\theta) = \|\theta\|_2 = (\sum_{i=1}\beta_i^2)^{\frac{1}{2}} \Rightarrow \frac{\partial f}{\partial \beta_1} = \frac{1}{2}\frac{1}{\|\theta\|_2}2\beta_1$$

$$\forall x, f(x) \geq f(0) + (v, x) \Rightarrow \|x\|_2 \geq (v, x)$$

因为对于任意 v 都成立，所以对其利用 C-S 不等式取到最大值仍然成立，故要求 $\|v\|_2 \leq 1$ □

在这个引理的基础上，对其进行计算最小值。

when $\theta_1 \neq 0$

$$l(\theta) = \frac{1}{2N}\left\|r_1 - z_1\theta_1\right\|^2 + \lambda\left\|\theta_1\right\|_2$$

$$\nabla l(\theta) = -\frac{1}{N}z_1^T(r_1 - z_1\theta_1)^2 + \frac{2\lambda\theta_1}{\|\theta_1\|^2} = 0$$

$$\Rightarrow (z_1^T z_1 + \frac{N\lambda}{\|\theta_1\|_2^2})\theta_1 = z_1^T r_1$$

$$\Rightarrow \theta_1 = (z_1^T z_1 + \frac{N\lambda I}{\|\theta_1\|_2^2})^{-1} z_1^T r_1$$

when $\theta_1 = 0$

$$0 \in \partial l(0) = -\frac{1}{N} z_1^T (r_1 - z_1\theta_1)^T + \lambda v$$

$$\Rightarrow v = \frac{1}{N\lambda} z_1^T r_1 \overset{\|v\|_2 \le 1}{\Rightarrow} \left\| z_1^T r_1 \right\|_2 \le N\lambda$$

因此有

$$\theta_1 = \begin{cases} 0 & \left\| z_1^T r_1 \right\|_2 \le N\lambda \\ N(z_1^T z_1 + \frac{N\lambda I}{\|\theta_1\|_2})^{-1} z_1^T r_1 & otherwise \end{cases}$$

$$\theta_1^1 = (z_1^T z_1 + \frac{N\lambda T}{\left\|\theta_1^0\right\|_2})^{-1} z_1^T r_1$$

1.2 Algorithm2: Proximal GD

$$\theta_{t+1} = Prox_{g/\beta}(\theta_t)$$

$$prox(y) = argmin\left\{ \frac{1}{N} \|y - \theta\|^2 + \lambda \sum_{j=1}^{J} \left\|\theta_j\right\|_2 \right\}$$

当好解的时候，即可以分解的时候可以采用这种方法。

$$min_{\theta_1} \left\{ \frac{1}{2N} \|y - \theta_1\|^2 + \lambda \|\theta_1\|_2 \right\}$$

when $\theta_1 \neq 0$

$$\frac{1}{N}(y_1 - \theta_1) + \frac{\lambda\theta_1}{\|\theta_1\|_2} = 0$$

$$\Rightarrow y_1 = (1 + \frac{N\lambda}{\|\theta_1\|_2})\theta_1 \Rightarrow \theta_1 = \frac{\|\theta_1\|_2}{N\lambda + \|\theta_1\|_2} y_1$$

$$\|y_1\|_2 = (1 + \frac{N\lambda}{\|\theta_1\|_2}) \|\theta_1\|_2 = \|\theta_1\|_2 + N\lambda$$

$$\Rightarrow \theta_1 = \frac{\|y_1\|_2 - N\lambda}{\|y_1\|_2} y_1 = (1 - \frac{N\lambda}{\|y_1\|_2}) y_1$$

when $\theta_1 = 0$

$$-\frac{1}{N} y_1 + \lambda v = 0 v \in \partial \|0\|_2, v = \frac{y_1}{N\lambda}, \|y\|_2 \le N\lambda$$

$$\theta_1 = \begin{cases} 0 & \|y_1\|_2 \le N\lambda \\ (1 - \frac{N\lambda}{\|y_1\|_2}) y_1 & otherwise \end{cases}$$