

## 一、问题一

### 1. 问题描述

考虑线性测量  $\mathbf{b} = A\mathbf{x} + \mathbf{e}$ ，其中  $\mathbf{b}$  为50维的测量值， $A$  为  $50 \times 100$  维的测量矩阵， $\mathbf{x}$  为100维的未知稀疏向量且稀疏度为5， $\mathbf{e}$  为50维的测量噪声。从  $\mathbf{b}$  与  $A$  中恢复  $\mathbf{x}$  的一范数规范化，最小二乘模型如下：

$$\min \left( \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + p \|\mathbf{x}\|_1 \right)$$

其中  $p$  为非负的正则化参数。

请设计下述算法求解该问题：

1. 邻近点梯度下降法
2. 交替方向乘子法
3. 次梯度法

在实验中，设  $\mathbf{x}$  的真值中的非零元素服从均值为0方差为1的高斯分布， $A$  中的元素服从均值为0方差为1的高斯分布， $\mathbf{e}$  中的元素服从均值为0方差为0.1的高斯分布。对于每种算法，请给出每步计算结果与真值的距离以及每步计算结果与最优解的距离。此外，请讨论正则化参数  $p$  对计算结果的影响。

### 2. 算法设计

#### (i) 邻近点梯度下降法

设

$$\begin{cases} s(\mathbf{x}) := \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 \\ r(\mathbf{x}) := p \|\mathbf{x}\|_1 \end{cases}$$

其中， $A \in \mathbb{R}^m \times \mathbb{R}^n$ ， $\mathbf{x}, \mathbf{b}, \mathbf{e} \in \mathbb{R}^n$ ，则原式

$$f(\mathbf{x}) := \left( \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + p \|\mathbf{x}\|_1 \right) = s(\mathbf{x}) + r(\mathbf{x})$$

先求  $r(\mathbf{x})$  的邻近点投影

$$\text{prox } \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left( p \|\mathbf{x}\|_1 + \frac{1}{2\alpha} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \right) \quad (1)$$

对1式右侧展开有

$$\arg \min_{\mathbf{x}} \left( p \sum_{i=1}^n |x_i| + \frac{1}{2\alpha} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \right) \quad (2)$$

注意到(2)式对于下标 $i$ 相互独立，故要求2的最小值，等价于对每一个下标 $i$ 求最小值后求和，即

$$\arg \min_{x_i} \left( p|x_i| + \frac{1}{2\alpha} (x_i - \hat{x}_i)^2 \right), \forall i \quad (3)$$

由不可微函数的极值判断条件有

$$0 \in \partial_{x_i} p|x_i| + \frac{1}{\alpha} (x_i - \hat{x}_i) \quad (4)$$

对 $x_i$ 进行分类讨论

- 若 $x_i > 0$ ，则 $|x_i|$ 对于 $|x_i|$ 可微，即 $\partial_{x_i} |x_i| = 1$ ，有

$$p + \frac{1}{\alpha} (x_i - \hat{x}_i) = 0$$

整理得

$$x_i = \hat{x}_i - \alpha p$$

由于 $x_i > 0$ ，故 $\hat{x}_i - \alpha p > 0$ ，即 $\hat{x}_i > \alpha p$

- 若 $x_i < 0$ ，则 $|x_i|$ 对于 $|x_i|$ 可微，即 $\partial_{x_i} |x_i| = -1$ ，有

$$-p + \frac{1}{\alpha} (x_i - \hat{x}_i) = 0$$

整理得

$$x_i = \hat{x}_i + \alpha p$$

由于 $x_i < 0$ ，故 $\hat{x}_i + \alpha p < 0$ ，即 $\hat{x}_i < -\alpha p$

- 若 $x_i = 0$ ，则 $|x_i|$ 对于 $|x_i|$ 不可微，需要求次梯度， $\partial_{x_i} |x_i| = [-1, 1]$ ，即

$$0 \in \left[ -p - \frac{\hat{x}_i}{\alpha}, p - \frac{\hat{x}_i}{\alpha} \right]$$

那么，需要满足

$$\begin{cases} p - \frac{\hat{x}_i}{\alpha} \geq 0 \\ -p - \frac{\hat{x}_i}{\alpha} \leq 0 \end{cases}$$

推得

$$\hat{x}_i \in [-\alpha p, \alpha p]$$

综上，有

$$x_i = \begin{cases} \hat{x}_i + \alpha p & \hat{x}_i < -\alpha p \\ 0 & \hat{x}_i \in [-\alpha p, \alpha p] \\ \hat{x}_i - \alpha p & \hat{x}_i > \alpha p \end{cases} \quad (5)$$

可以得到图1的软门限图

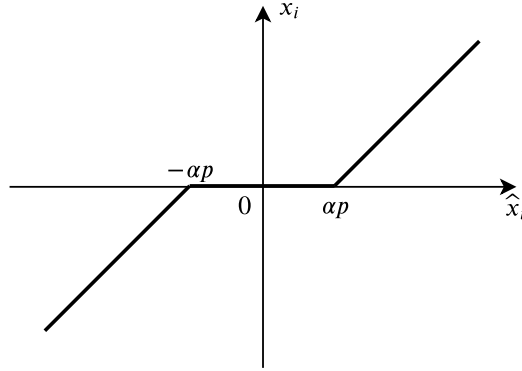


图 1: 关于  $x_i$  与  $\hat{x}_i$  的软门限图

进一步，得到邻近点梯度下降法的迭代式如下

$$\begin{cases} \mathbf{x}^{(k+\frac{1}{2})} = \mathbf{x}^{(k)} - \alpha \nabla s(\mathbf{x}^{(k)}) = \mathbf{x}^{(k)} - \alpha A^T (A\mathbf{x}^{(k)} - \mathbf{b}) \\ \mathbf{x}^{(k+1)} = \text{prox } \mathbf{x}^{(k+\frac{1}{2})} = \arg \min_{\mathbf{x}} \left( p \|\mathbf{x}\|_1 + \frac{1}{2\alpha} \left\| \mathbf{x} - \mathbf{x}^{(k+\frac{1}{2})} \right\|_2^2 \right) \end{cases} \quad (6)$$

其中， $x^{(k+\frac{1}{2})}$ 可直接计算， $x^{(k+1)}$ 可由(5)式求得。

## (ii) 交替方向乘子法

对原问题进行变形，等价于下述约束问题

$$\begin{aligned} \min \quad & \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + p \|\mathbf{y}\|_1 \\ \text{s.t.} \quad & \mathbf{x} - \mathbf{y} = 0 \end{aligned}$$

构造增广拉格朗日函数

$$L_c(\mathbf{x}, \mathbf{y}, \mathbf{v}) = \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + p \|\mathbf{y}\|_1 + \mathbf{v}^T (\mathbf{x} - \mathbf{y}) + \frac{c}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 \quad (7)$$

可以得到交替方向乘子法的迭代格式

$$\begin{cases} \mathbf{x}^{(k+1)} = \arg \min_{\mathbf{x}} L_c(\mathbf{x}, \mathbf{y}^{(k)}, \mathbf{v}^{(k)}) \\ \mathbf{y}^{(k+1)} = \arg \min_{\mathbf{y}} L_c(\mathbf{x}^{(k+1)}, \mathbf{y}, \mathbf{v}^{(k)}) \\ \mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} + c(\mathbf{x}^{(k+1)} - \mathbf{y}^{(k+1)}) \end{cases} \quad (8)$$

对 $\mathbf{x}^{(k+1)}$ 展开并配方，并将非主元项忽略，可求得(8)与下面的式子等价

$$\begin{cases} \mathbf{x}^{(k+1)} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \frac{c}{2} \left\| \mathbf{x} - \mathbf{y}^{(k)} + \frac{\mathbf{v}^{(k)}}{c} \right\|_2^2 & (9a) \\ \mathbf{y}^{(k+1)} = \arg \min_{\mathbf{y}} p \|\mathbf{y}\|_1 + \frac{c}{2} \left\| \mathbf{x}^{(k+1)} - \mathbf{y} + \frac{\mathbf{v}^{(k)}}{c} \right\|_2^2 & (9b) \\ \mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} + c(\mathbf{x}^{(k+1)} - \mathbf{y}^{(k+1)}) & (9c) \end{cases}$$

对于(9a)式，可直接通过求梯度的方法得到显式解，即

$$\mathbf{x}^{(k+1)} = (A^T A + cI)^{-1} (A^T \mathbf{b} + c\mathbf{y}^{(k)} - \mathbf{v}^{(k)})$$

对于(9b)式，由于涉及一范数，故需要求次微分，类似(5)式的方法，设 $z_i = x_i^{(k+1)} + \frac{v_i^{(k)}}{c}$ ，可得到

$$y_i = \begin{cases} z_i - \frac{p}{c} & z_i > \frac{p}{c} \\ 0 & z_i \in \left[-\frac{p}{c}, \frac{p}{c}\right] \\ z_i + \frac{p}{c} & z_i < -\frac{p}{c} \end{cases}$$

进而可以通过(9a)(9b)(9c)式迭代求解。

### (iii) 次梯度法

对原式直接求次梯度有

$$\partial f(\mathbf{x}) = A^T (\mathbf{Ax} - \mathbf{b}) + p \partial \|\mathbf{x}\|_1 \quad (10)$$

其中

$$(\partial \|\mathbf{x}\|_1)_i = \partial |x_i| = \begin{cases} 1 & x_i > 0 \\ [-1, 1] & x_i = 0 \\ -1 & x_i < 0 \end{cases}$$

进而可以直接得到次梯度法的迭代格式

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha \partial f(\mathbf{x}^{(k)}) \quad (11)$$

### 3. 数值实验

代码如下，也以附在附件中。

### 4. 结果分析

## 二、问题二

### 1. 问题描述

请设计下述算法，求解MNIST数据集上的Logistic Regression问题：

1. 梯度下降法

2. 随机梯度法

对于每种算法，请给出每步计算结果与最优解的距离以及每步计算结果在测试集上所对应的分类精度。此外，请讨论随机梯度法中Mini Batch大小对计算结果的影响。

可参考：<http://deeplearning.net/tutorial/logreg.html>

### 2. 算法设计

(i) 梯度下降法

(ii) 随机梯度法

### 3. 数值实验

代码如下，也以附在附件中。

### 4. 结果分析