李健宁
机器学习中的优化问题
2025 年 6 月 3 日

Homework 14

Question 1.

应用题目中要求的几种方法，最大迭代次数 1000 次，学习率 $\eta = 0.01$，其他参数采取 Pytorch 中相应优化器的默认参数，实验结果如图1。

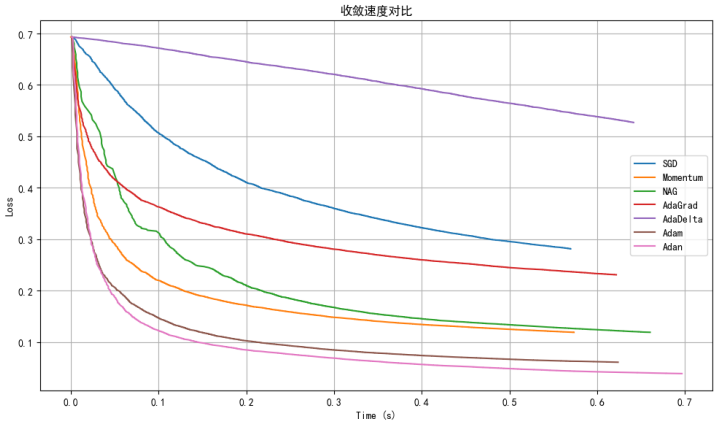

图 1. 收敛速度对比

Question 2.

在 RCD(1) 方法中，我们对每个坐标 i 维护一个自适应步长缩放因子 β_i，其估计采用指数滑动平均方式，具体更新如下

$$\beta_i^{(t)} = \rho \cdot \beta_i^{(t-1)} + (1-\rho) \cdot \left(\nabla_i f(w^{(t-1)})\right)^2 \tag{1}$$

其中 $\rho \in [0,1)$ 是滑动平均系数，这里取 $\rho = 0.9$；$\nabla_i f(w^{(t-1)})$ 表示在第 $t-1$ 次迭代中，第 i 个分量的梯度；$\beta_i^{(t)}$ 表示当前估计的第 i 个分量的尺度。

最大迭代次数 1000 次，学习率 $\eta = 0.01$，其他参数采取 Pytorch 中相应优化器的默认参数，实验结果如图2。

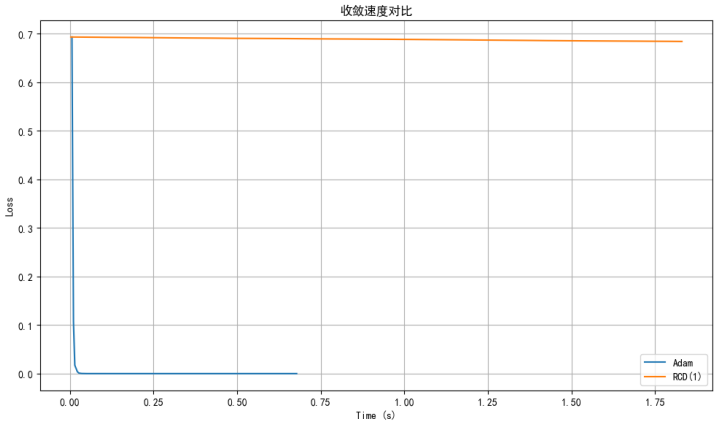

图 2. 收敛速度对比

Question 3.

为引入 ADMM 结构，我们将变量 $\boldsymbol{w}$ 分裂为两个变量 $\boldsymbol{w}$ 与 $\boldsymbol{z}$，得到等价问题

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{z}} \quad & \frac{1}{N}\sum_{i=1}^{N}\log\left(1+\exp\left(-y_i\cdot \boldsymbol{x}_i^\top \boldsymbol{w}\right)\right) \\ \text{s.t.} \quad & A_1 w + A_2 z = b, \end{aligned}$$

$$A_1 = \begin{bmatrix} I_d \\ \mathbf{0}^\top \end{bmatrix} \in \mathbb{R}^{(d+1)\times d},\ A_2 = \begin{bmatrix} -I_d \\ \mathbf{1}^\top \end{bmatrix} \in \mathbb{R}^{(d+1)\times d},\ b = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} \in \mathbb{R}^{(d+1)}$$

每次迭代的具体更新方式，首先随机选取样本索引 $i_k \in \{1,2,\ldots,N\}$，然后根据以下公式依次更新

$$w^{k+1} = \arg\min_{w}\left\{\nabla f_{i_k}(w^k)^\top (w-w^k) + \frac{\rho}{2}\left\|A_1 w + A_2 z^k - b + \frac{1}{\rho}u^k\right\|^2 + \frac{1}{2\eta_k}\left\|w-w^k\right\|^2\right\} \tag{2}$$

$$z^{k+1} = \arg\min_{z}\left\{I_{\{\mathbf{1}^\top z=1\}}(z) + \frac{\rho}{2}\left\|A_1 w^{k+1} + A_2 z - b + \frac{1}{\rho}u^k\right\|^2\right\} \tag{3}$$

$$u^{k+1} = u^k + \rho\left(A_1 w^{k+1} + A_2 z^{k+1} - b\right) \tag{4}$$

参数选取上，惩罚参数 $\rho = 0.1$，初始学习率 $\eta_0 = 0.1$, $\eta_k = \frac{\eta_0}{\sqrt{k+1}}$，最大迭代次数 10000 次，实验结果如图3,4。

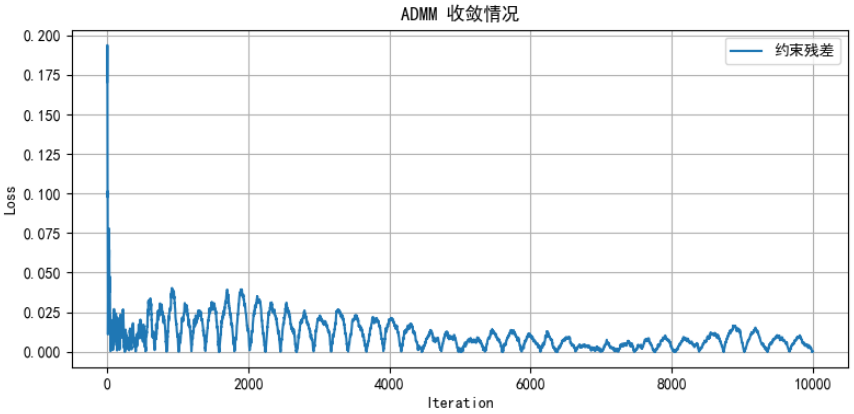

图 3. 约束残差 $\|A_1 w + A_2 z - b\|$

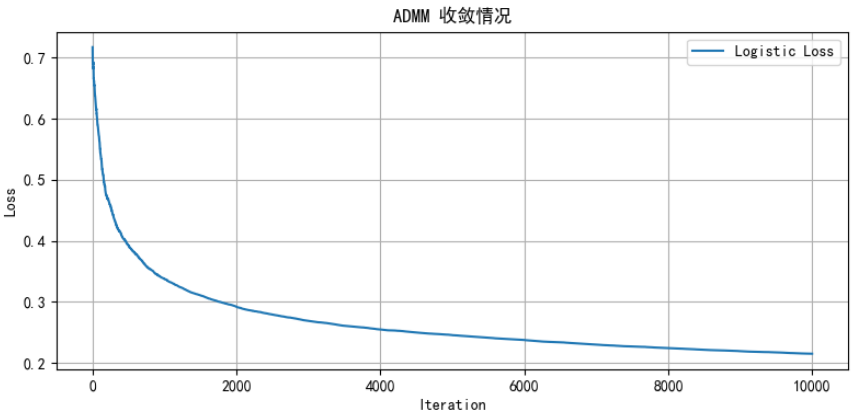

图 4. Logistice Loss（目标函数值）