最优化算法实验报告

陈文轩

更新：June 8, 2025

1 基本模型与优化目标

1.1 任务与目标

- 任务：对二分类数据 $\{(a_i, b_i)\}_{i=1}^m$，其中 $b_i \in \{-1, +1\}$，学习稀疏权重向量 $x \in \mathbb{R}^n$。
- 目标函数：

$$\min_x \ell(x) = \frac{1}{m}\sum_{i=1}^m \log\left(1 + \exp\left(-b_i a_i^\top x\right)\right) + \lambda\|x\|_2^2 + \mu\|x\|_1,$$

  - 第一项是逻辑回归损失函数。
  - 第二项 $\lambda\|x\|_2^2$ 提供 L^2 正则化，帮助数值稳定并减少共线性。
  - 第三项 $\mu\|x\|_1$ 提供 L^1 正则化，促进稀疏性。
  - $\ell(x)$ 是一个凸函数。

1.2 ADMM 变量拆分与增广拉格朗日函数

为了处理非光滑的 L^1 项，引入辅助变量 z，并添加约束 $x = z$：

$$\min_{x,z} \frac{1}{m}\sum_{i=1}^m \log\left(1 + \exp\left(-b_i a_i^\top x\right)\right) + \lambda\|x\|_2^2 + \mu\|z\|_1, \quad \text{s.t. } x = z,$$

增广拉格朗日函数为：

$$\mathcal{L}_\rho(x, z, y) = \frac{1}{m}\sum_{i=1}^m \log\left(1 + \exp\left(-b_i a_i^\top x\right)\right) + \lambda\|x\|_2^2 + \mu\|z\|_1 + \frac{\rho}{2}\|x - z\|_2^2 + y^\top(x - z),$$

令 $v_k = z_k - \dfrac{y_k}{\rho}$ 即可得到 ADMM 的各个子问题：

1. x-子问题：

$$x_{k+1} = \arg\min_x \left(\frac{1}{m}\sum_{i=1}^m \log\left(1 + \exp\left(-b_i a_i^\top x\right)\right) + \lambda\|x\|_2^2 + \frac{\rho}{2}\|x - v_k\|^2\right)$$

2. z-子问题：

$$z_{k+1} = \arg\min_{z} \left(\mu \|z\|_1 + \frac{\rho}{2} \|v_{k+1} - z\|^2 \right)$$

3. 对偶更新：

$$y_{k+1} = y_k + \rho(x_{k+1} - z_{k+1})$$

1.3 子问题求解

1.3.1 x-子问题

x-子问题是一个光滑的凸优化问题，可以使用牛顿法等方法求解：

- 梯度：

$$\nabla f(x) = \frac{1}{m} A^\top \left(\sigma(-b \odot (Ax)) - b \right) + 2\lambda x + \rho(x - v_k),$$

  其中 $\sigma(t) = \dfrac{1}{1 + e^{-t}}$。
- Hessian 矩阵：

$$H \coloneqq \nabla^2 f(x) = \frac{1}{m} A^\top D A + (2\lambda + \rho) I, D_{ij} = \delta_{ij} \sigma(-b_i a_i^\top x)(1 - \sigma(-b_i a_i^\top x))$$

- 方向：$p_k = -H^{-1} \nabla f(x)$，若出现奇异或方向上升，则退化为负梯度步。
- 步长：Armijo 条件 $f(x + t p_k) \le f(x) + c_1 t \nabla f^\top p_k$，程序中取 $c_1 = 0.01, \beta = 0.5$。
- 终止条件：$\|t p_k\|_2 < 10^{-7}$ 或迭代次数满 `max_newton_iter`。

上述细节完全对应 `x_update_objective_and_grad_hess` 与 `solve_x_subproblem` 函数。

1.3.2 z-子问题

z-子问题有解析解

$$z_{k+1} = S_{\frac{\mu}{\rho}} \left(x_{k+1} + \frac{y_k}{\rho} \right),$$

其中软阈值算子定义为

$$S_\kappa(u) = \operatorname{sgn}(u) \max\{|u| - \kappa, 0\}$$

实现见 `soft_threshold` 函数。

1.4 收敛判据

原始残差 $r_k^{\text{pri}} = \|x_k - z_k\|_2$，对偶残差 $r_k^{\text{dual}} = \rho \|z_k - z_{k-1}\|_2$，当 $\max\{r_k^{\text{pri}}, r_k^{\text{dual}}\} < \text{tol\_abs} = 10^{-6}$ 时认为收敛。

2 实验结果

$\lambda = \dfrac{1}{2m}, \mu = 0.01$ 时 ADMM 两个最优条件与迭代步数的关系图以及误差与迭代步数的关系图如下所示：

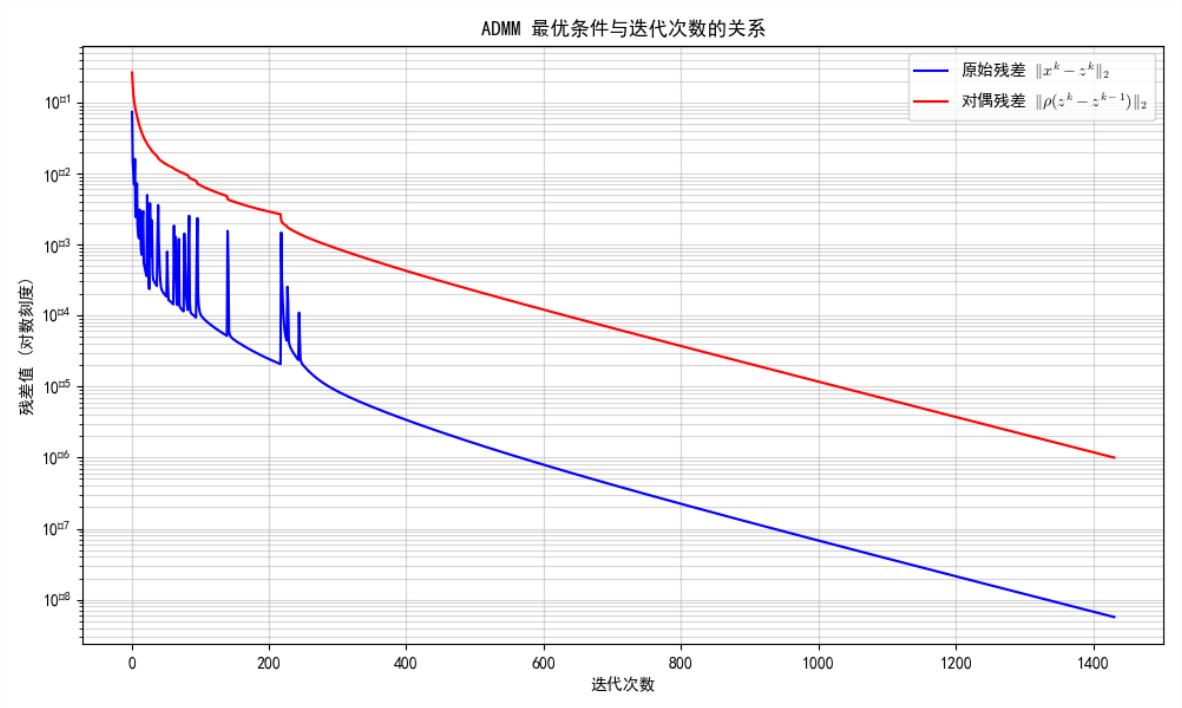

图 1: ADMM 原始残差与对偶残差

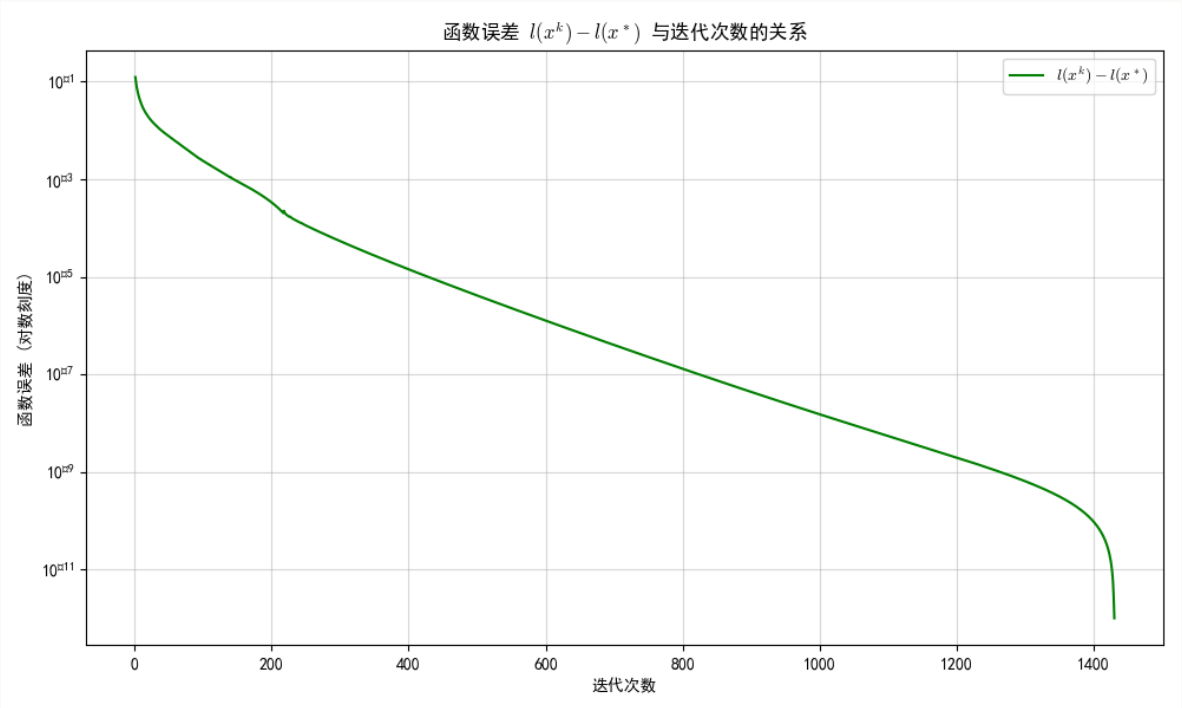

图 2: ADMM 目标函数误差

迭代步数，稀疏度与 μ 的关系如下表所示：

μ	迭代步数	稀疏度	非零元数量
0.001	4570	0.6829	39
0.01	1430	0.8862	14
0.05	645	0.9675	4
0.1	204	0.9675	4

表 1: ADMM 迭代步数与稀疏度

可以发现，随着 μ 的增大，迭代步数减少，稀疏度增加，非零元数量减少。$\mu = 0.05$ 时，模型非常稀疏，仅有 4 个非零元。

3 参数选取

参数	作用	默认值	说明
λ	L^2 正则化系数	$\frac{1}{2m}$	控制模型复杂度，防止过拟合
μ	L^1 正则化系数	$0.001, 0.01, 0.05$	促进稀疏性，选择特征
ρ	ADMM 增广拉格朗日参数	1	控制原始和对偶残差的平衡
`max_admm_iter`	ADMM 最大迭代次数	10000	过大仅保证极端情况。
`max_newton_iter`	牛顿法最大迭代次数	20	若单次步长极小/收敛慢，可增大；但超 50 回报递减
`tol_abs`	收敛判据绝对容忍度	10^{-6}	原始和对偶残差均小于此值时认为收敛
`bt_c1,bt_beta`	Armijo 条件参数	$0.01, 0.5$	通常保持默认即可；若目标函数震荡可减小 `bt_beta`

表 2: ADMM 参数选取