

Report

PB20020480 王润泽

Machine Learning

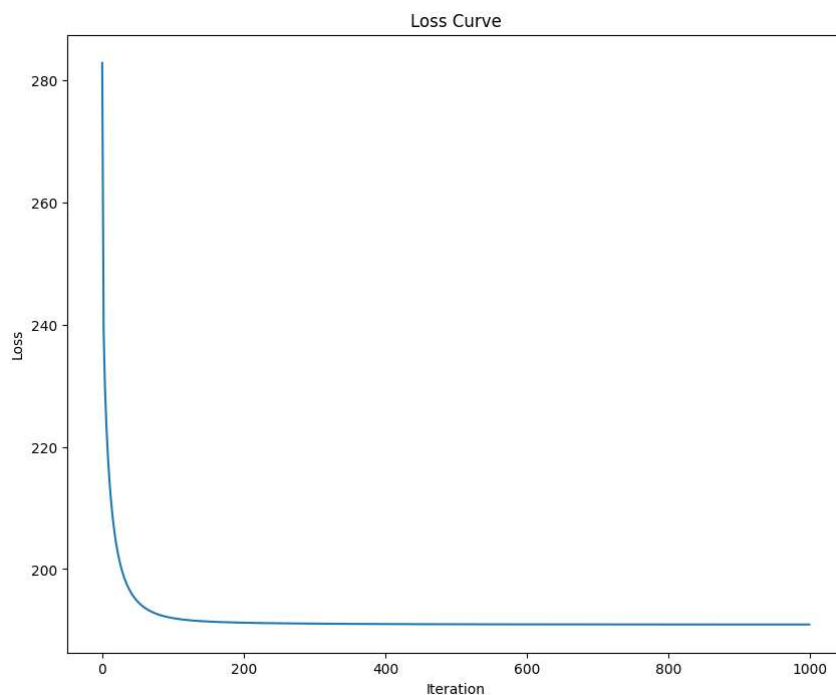
1. 损失曲线

数据集分割: 数据集大小511 训练集大小408(80%) 随机分割

参数:

- `penalty = l2`
- `gamma = 0.5`
- `fit_intercept = False`
- `lr = 0.005`
- `tol = 1e-7`
- `max_iter = 1e3`

得到: $accuracy = 0.825$



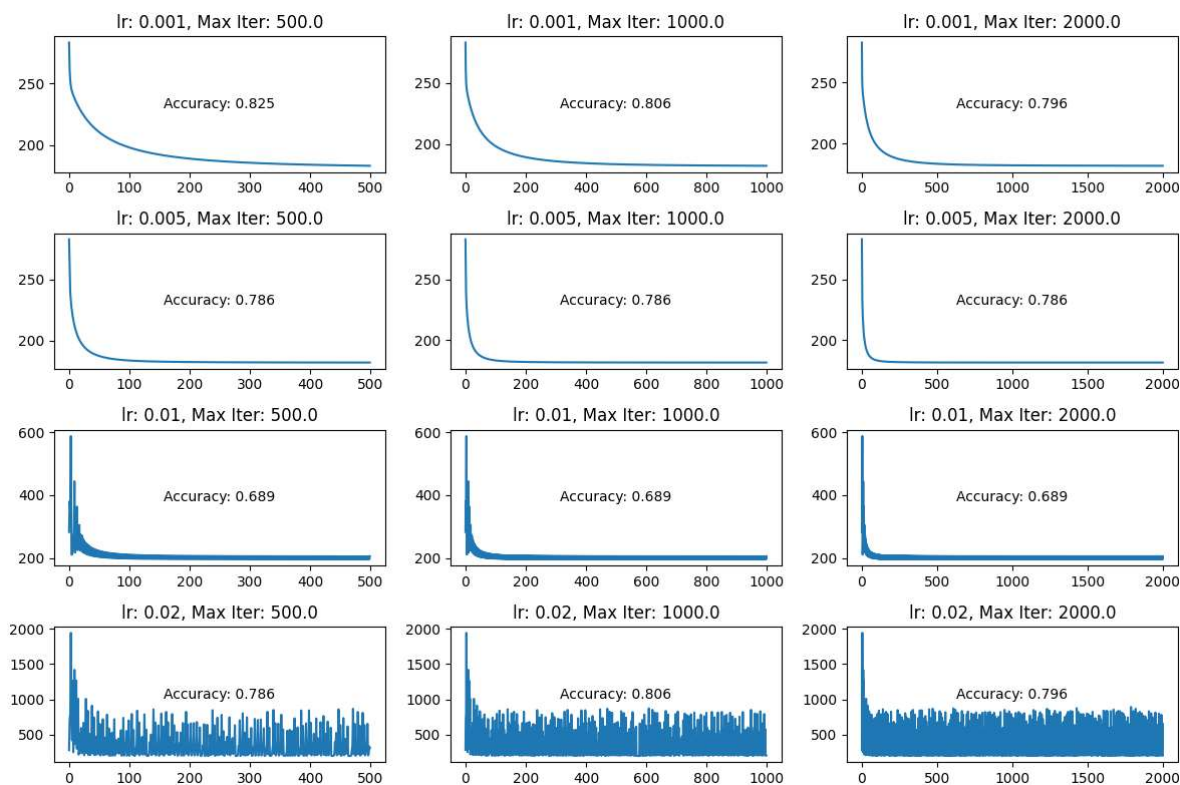
2. 参数影响

以下数据集都是原始数据集 8:2划分下进行的测试。

2.1 探究学习率与迭代次数的影响

该实验在以下参数下进行:

- `penalty = l2` , `gamma = 0` , `fit_intercept = False` , `tol = 1e-7`

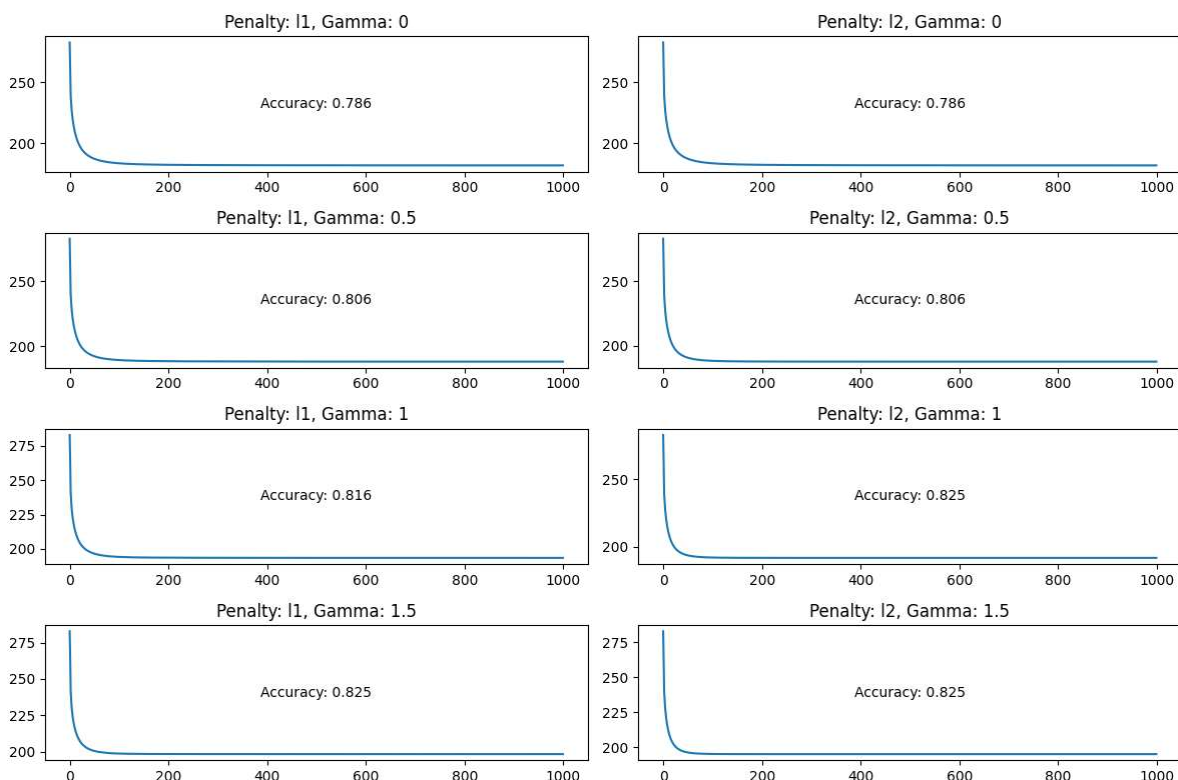


可以发现，lr在0.005时曲线都较为稳定，而0.01开始已经出现了明显的震荡。当学习率进一步增大时，过拟合也变得更加明显

2.2 探究正则化参数的影响

该实验在以下参数下进行：

- `lr = 0.005` , `tol = 1e-7` , `max_iter = 1e3`

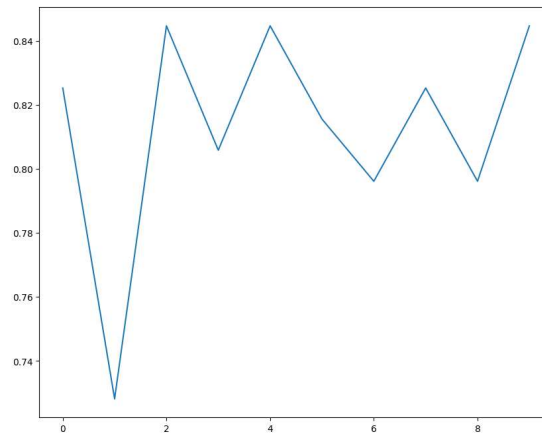


在正则化项增强时，最终预测结果变得更好，接近理想值：0.825。同时，l2正则化的效果比l1更加明显，这是由于其梯度直接与w的值相关，而不是只与符号相关。

3. 最佳准确率

在以下参数下进行训练，并使用10次随机 8:2留出结果，得到

- `penalty = l2`
- `gamma = 0.5`
- `fit_intercept = False`
- `lr = 0.001`
- `tol = 1e-7`
- `max_iter = 5e3`



平均结果为：**81.3%**