

机器学习与数据挖掘

Homework 3: Linear Regression and Logistic Regression

中山大学计算机学院 计算机科学与技术

19335174 施天予

一、Linear Regression

(a) 线性回归模型有 3 个参数

$$h(x) = w_0 + w_1x_1 + w_2x_2$$

梯度下降单轮迭代

$$w_0 = w_0 - \rho \frac{1}{n} \sum_{i=1}^n (h(x^{(i)}) - y^{(i)})$$

$$w_1 = w_1 - \rho \frac{1}{n} \sum_{i=1}^n ((h(x^{(i)}) - y^{(i)}) \cdot x_1^{(i)})$$

$$w_2 = w_2 - \rho \frac{1}{n} \sum_{i=1}^n ((h(x^{(i)}) - y^{(i)}) \cdot x_2^{(i)})$$

计算误差

$$E = \frac{1}{n} \sum_{i=1}^n (h(x^{(i)}) - y^{(i)})^2$$

具体设置如下

迭代次数	学习率	参数初始化	图示迭代间隔
1500000	0.00015	0.0	100000

结果如图所示

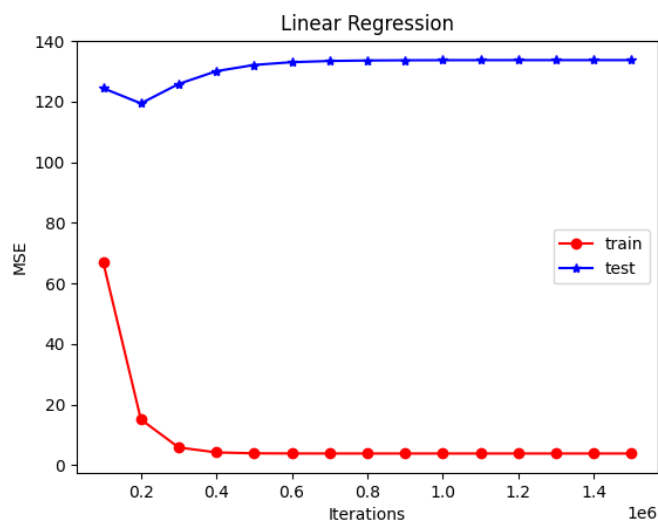


图 1: 训练样本和测试样本的误差

训练集误差下降至 0，测试集误差不能下降到 0，说明模型未能很好收敛，适当增大学习率可以改善

(b) 由于学习率增大，numpy 中的 float64 出现越界现象，所以降低了迭代次数。同时更改了图示的迭代间隔，便于观察。最后更改设置如下

迭代次数	学习率	参数初始化	图示迭代间隔
2000	0.0002	0.0	200

结果如图所示

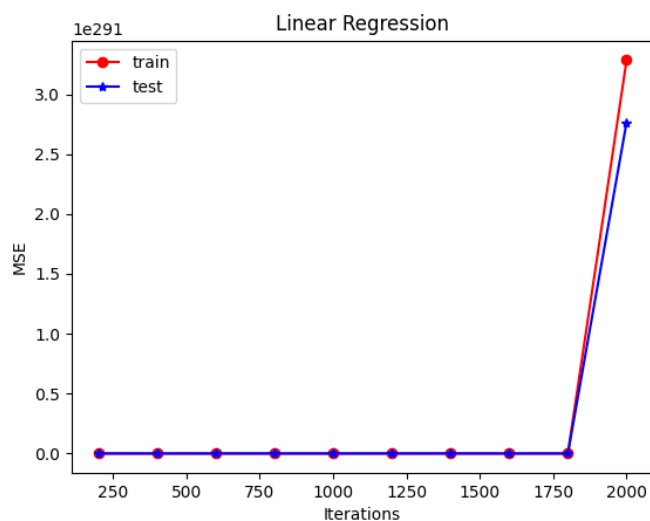


图 2: 训练样本和测试样本的误差

发现在最后一轮迭代时误差突然增大，导致结果难以正确收敛，因此需要适当降低学习率，从而加快模型收敛。

(c) 随机梯度下降设置与 (a) 相同

迭代次数	学习率	参数初始化	图示迭代间隔
1500000	0.00015	0.0	100000

每次迭代随机选 1 个点

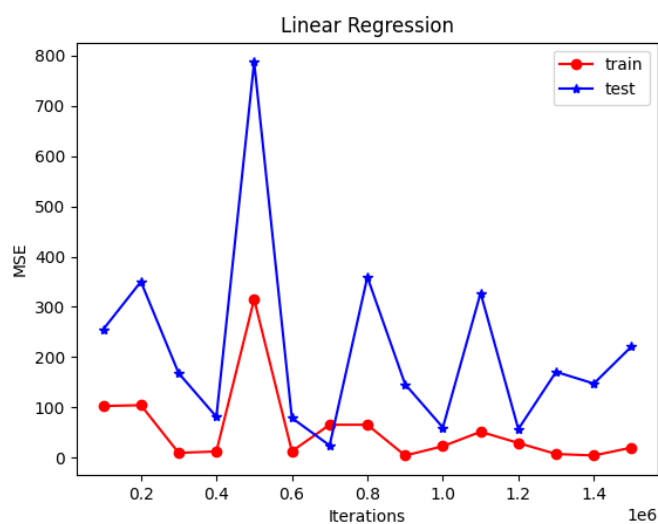


图 3: 训练样本和测试样本的误差

每次迭代随机选 10 个点

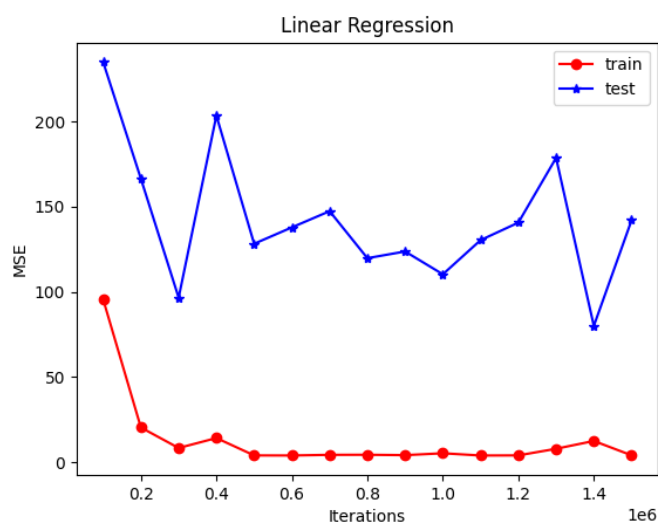


图 4: 训练样本和测试样本的误差

随机梯度下降训练时间大大减小，误差虽会随着迭代次数增加而减小，但是波动较大。

二、Exercise Two: Logistic Regression

(a) Sigmoid 函数：

$$P(x) = \frac{1}{1 + e^{-wx}}$$

对数似然函数：

$$f = \sum_{i=1}^n y^{(i)}(w_0 + \sum_{j=1}^p w_j x_j^{(i)}) - \ln(1 + \exp(w_0 + \sum_{j=1}^p w_j x_j^{(i)}))$$

其实 n 是样例个数，p 是特征个数，(i) 表示采样索引，j 表示特征索引。

(b) 对上式对数似然函数求偏导：

$$\frac{\partial f}{\partial w_0} = \sum_{i=1}^n (y^{(i)} - \frac{1}{1 + e^{-(w_0 + \sum_{k=1}^p w_k x_k^{(i)})}})$$

$$\frac{\partial f}{\partial w_j} = \sum_{i=1}^n (y^{(i)} - \frac{1}{1 + e^{-(w_0 + \sum_{k=1}^p w_k x_k^{(i)})}}) x_j^{(i)}$$

(c) 使用全局梯度下降，优化参数为 $w_0, w_1, w_2, w_3, w_4, w_5, w_6$ 。参数更新公式如下，其中 η 为学习率

$$w_j = w_j + \eta \frac{\partial f}{\partial w_j}$$

迭代次数	学习率	参数初始化	图示迭代间隔
200	0.00015	0.0	20

训练集和测试集误差都能降低为 0，模型收敛较快

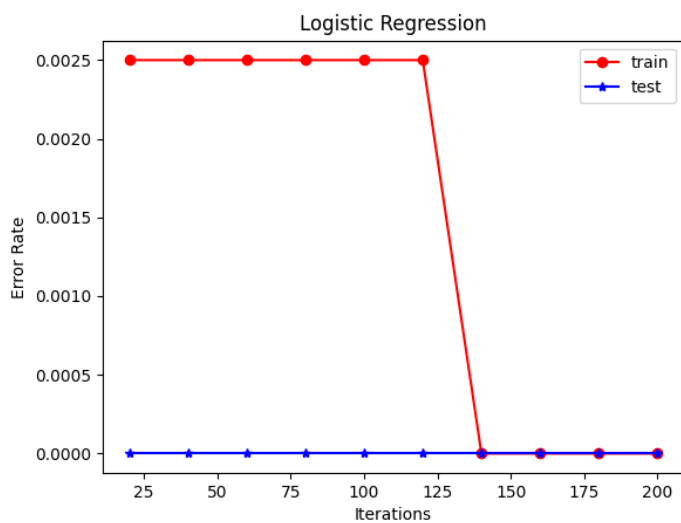


图 5: 训练样本和测试样本的误差

学习率设为 0.0002，模型收敛更快

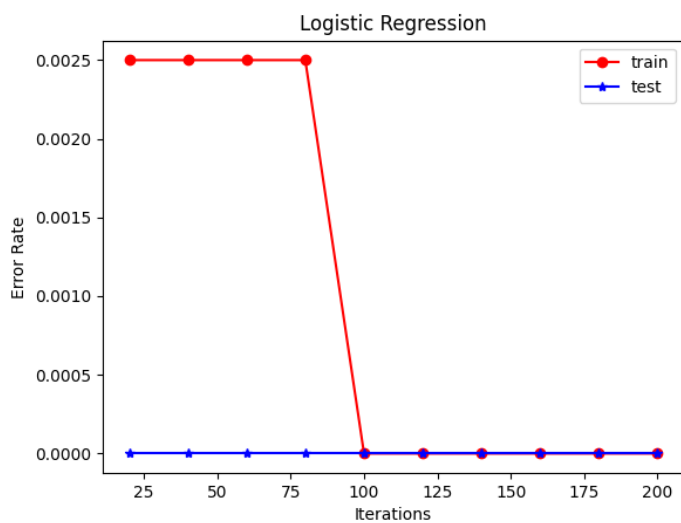


图 6: 训练样本和测试样本的误差

使用随机梯度下降

迭代次数	学习率	参数初始化	图示迭代间隔
150000	0.00015	0.0	10000

每次迭代随机选 1 个点

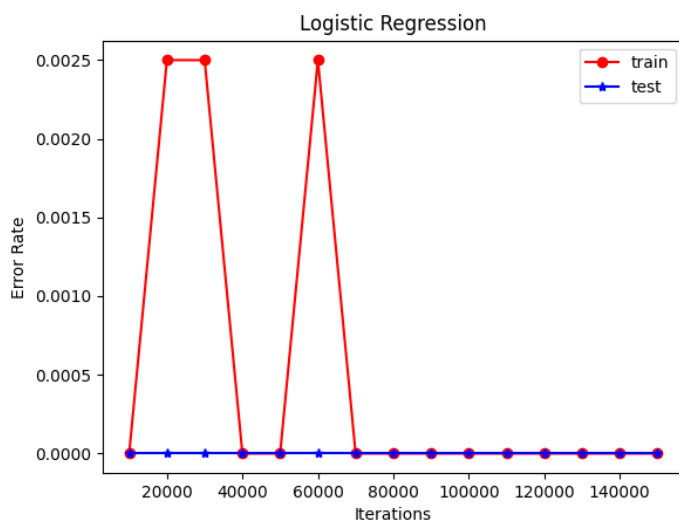


图 7: 训练样本和测试样本的误差

每次迭代随机选 10 个点

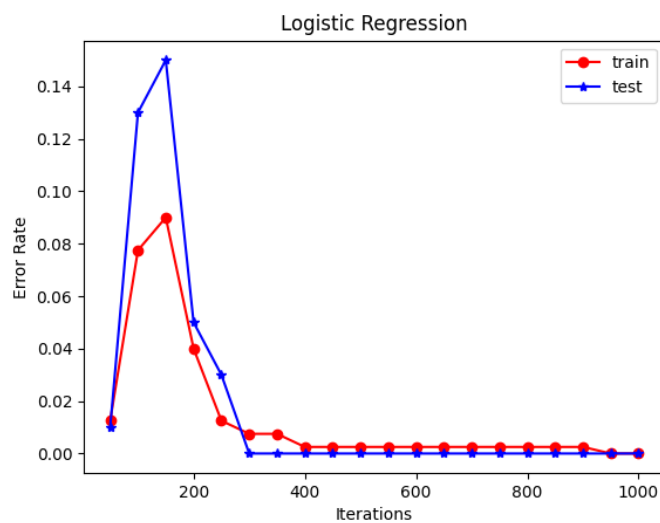


图 8: 训练样本和测试样本的误差

使用随机梯度下降，误差波动较大，但迭代多次错误率都能为 0

(d) 迭代次数足够多时，错误率就能为 0

(e) 随机梯度下降，每次迭代随机选 1 个点

迭代次数	学习率	参数初始化	图示迭代间隔
150000	0.00015	0.0	10000

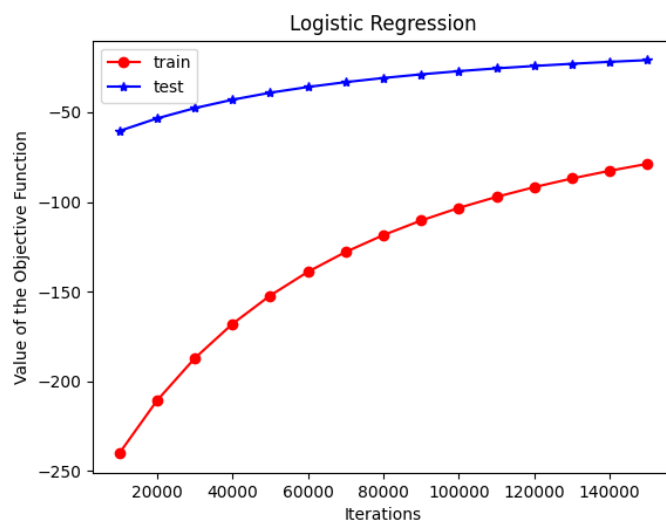


图 9: 训练样本和测试样本的误差

发现 150000 次迭代后还未收敛。增大学习率至 0.002，每次迭代随机选 10 个点

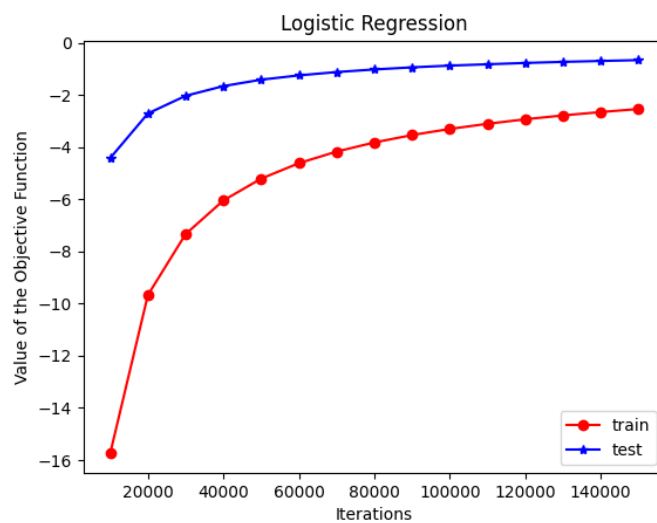


图 10: 训练样本和测试样本的误差

约 90000 次迭代后模型收敛

(f) 统一迭代次数和学习率

迭代次数	学习率	参数初始化
100	0.00015	0.0

全局梯度下降

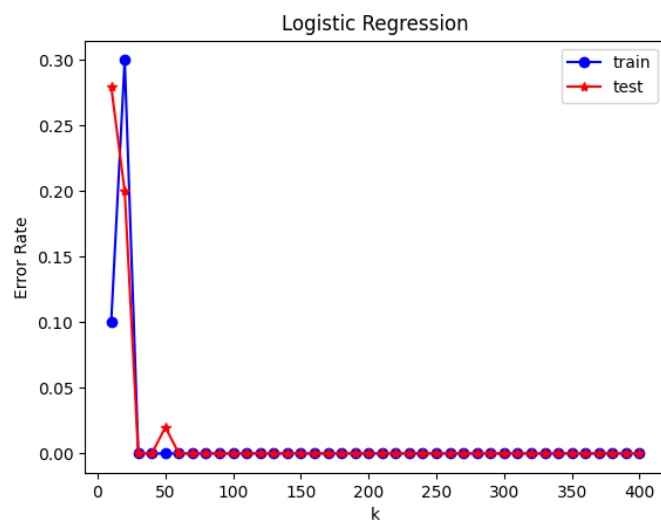


图 11: 训练样本和测试样本的误差

随机梯度下降，每次迭代随机选 1 个点

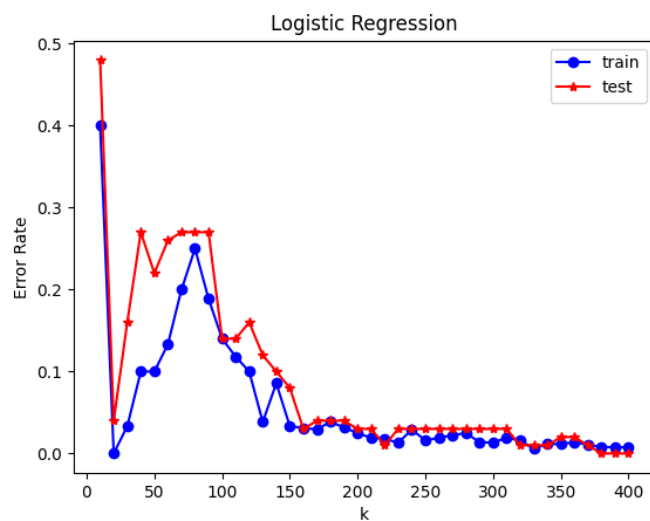


图 12: 训练样本和测试样本的误差

全局梯度下降在训练样本较少时错误率也能降为 0。随机梯度下降随着训练样本的增加，错误率整体逐渐降低，但由于随机性会导致错误率的波动。