

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**   **陈宇坤**

**学 号 201530611326**

**邮 箱 972640770@qq.com**

**指导教师**  **吴庆耀**

**提交日期** **2017年 12 月 2日**

## 1. 实验题目: 线性回归、线性分类与梯度下降

## 2. 实验时间：2017年 12 月 2日

## 3. 报告人: 陈宇坤

## 4. 实验目的:

1. 进一步理解线性回归和梯度下降的原理。
2. 在小规模数据集上实践。
3. 体会优化和调参的过程

## 数据集以及数据分析：

线性回归使用的是[LIBSVM Data](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/" \t "_blank)中的[Housing](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/regression.html" \l "housing" \t "_blank)数据，包含506个样本，每个样本有13个属性。请自行下载scaled版本，并将其切分为训练集，验证集

## 实验步骤:

线性回归和梯度下降

1.读取实验数据，使用sklearn库的load\_svmlight\_file函数读取数据。

2.将数据集切分为训练集和验证集，本次实验不切分测试集。使用train\_test\_split函数切分数据集。

3.线性模型参数初始化，可以考虑全零初始化，随机初始化或者正态分布初始化。

4.选择Loss函数及对其求导，过程详见课件ppt。

5.求得所有样本对Loss函数的梯度。

6.取梯度的负方向记为*D*。

7.更新模型参数，。为学习率，是人为调整的超参数。

8.在训练集上测试并得到Loss函数值，在验证集上测试并得到Loss函数值。

9.重复步骤5-8若干次，画出和随迭代次数的变化图

线性分类和梯度下降

1.读取实验数据，使用sklearn库的load\_svmlight\_file函数读取数据。

2.将数据集切分为训练集和验证集，本次实验不切分测试集。使用train\_test\_split函数切分数据集。

3.支持向量机模型参数初始化，可以考虑全零初始化，随机初始化或者正态分布初始化。

4.选择Loss函数及对其求导，过程详见课件ppt。

5.求得所有样本对Loss函数的梯度。

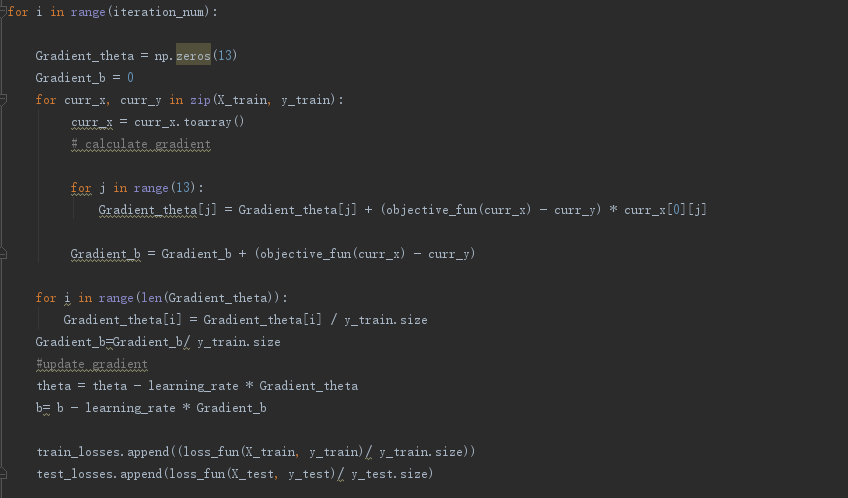
6.取梯度的负方向，记为。

7.更新模型参数，。为学习率，是人为调整的超参数。

8.选择合适的阈值，将计算结果大于阈值的标记为正类，反之为负类。在训9练集上测试并得到Loss函数值，在验证集上测试并得到Loss函数值。

9.重复步骤5-8若干次，画出和随迭代次数的变化图。

## 7. 代码内容:



## 8. 选择的评估方法:

评估方法为留出法。

## 9. 模型参数的初始化方法:

全零初始化参数。

## 10.选择的loss函数及其导数:

线性回归loss函数：

导数：

线性分类Loss函数：

导数：

## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：

线性回归：学习率为0.1 迭代次数：100

线性分类：学习率为0.0001 迭代次数：200

阈值为0.1 常数C=0.9

## 评估结果（根据选择的评估方法）：蓝色为train\_loss，红色为test\_loss，绿色为线性分类准确度

线性回归：

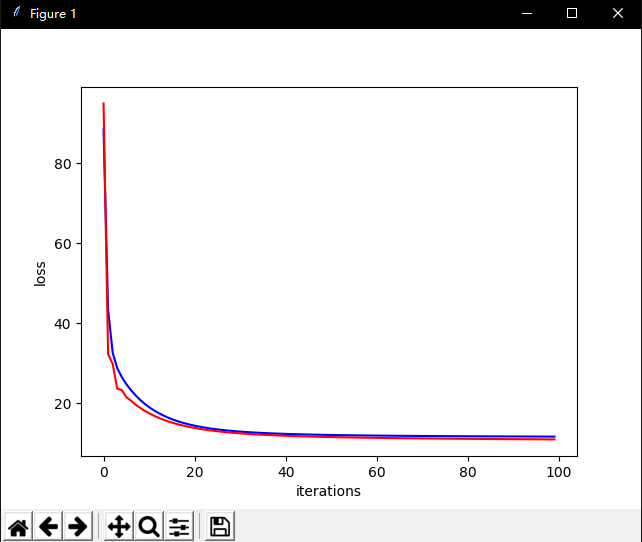


图1 学习率：0.3 迭代次数：100

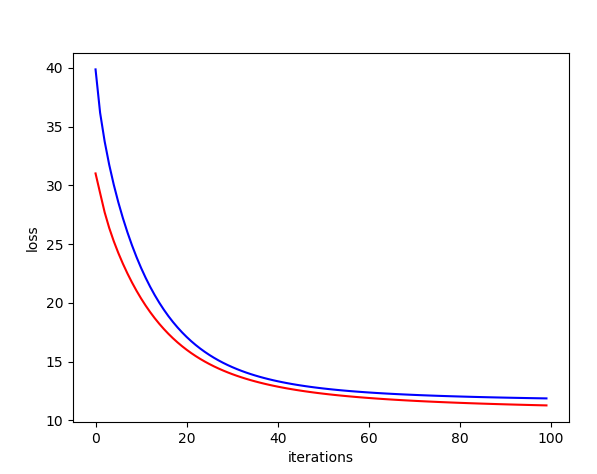


图2 学习率0.2 迭代次数100

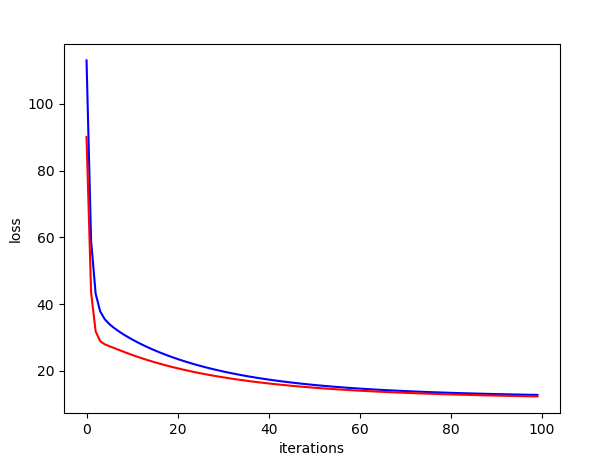


图3 学习率0.1 迭代次数100

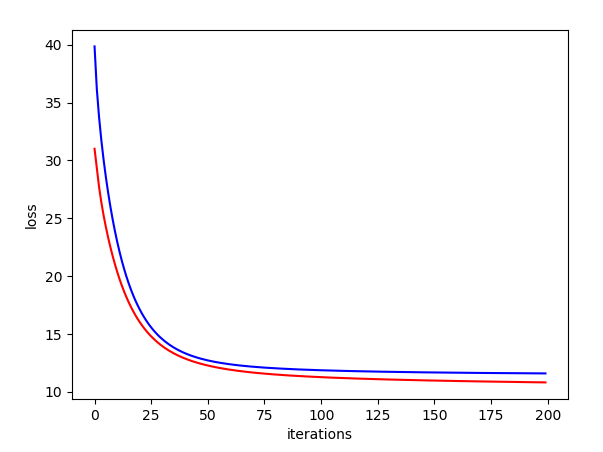


图4 学习率0.2 迭代次数200

线性分类：

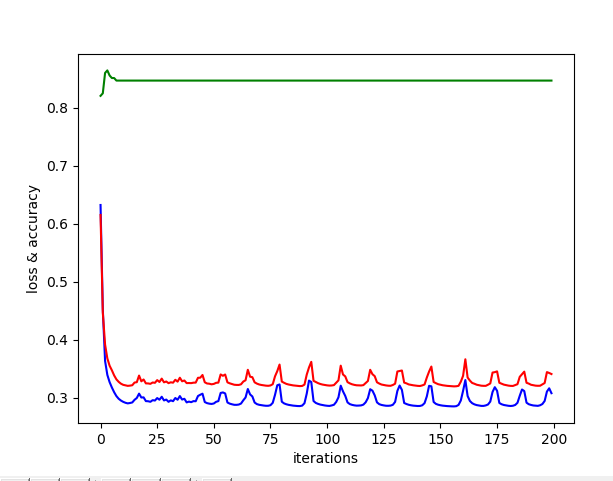


图5 学习率0.001 迭代次数200

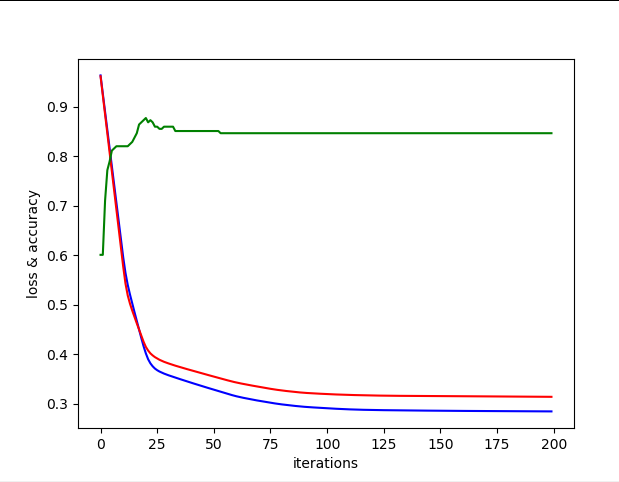


图6 学习率：0.0001 迭代次数200

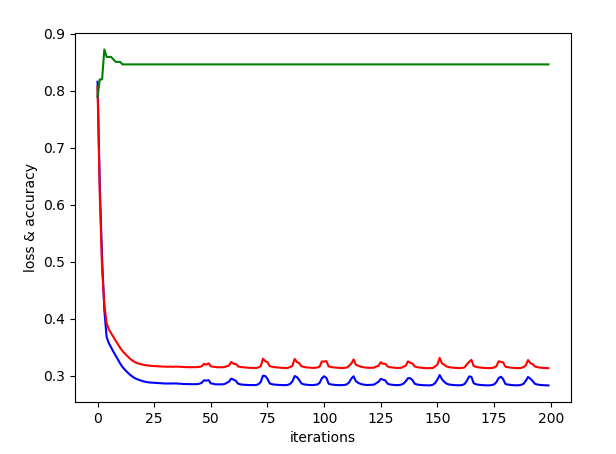


图7 学习率0.0005 迭代次数200

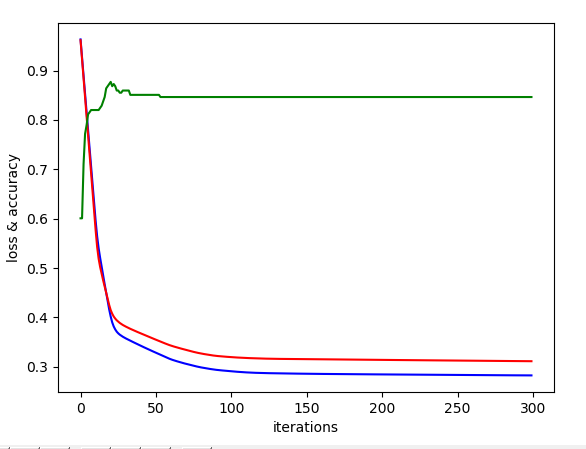


图8学习率0.0001 迭代次数300

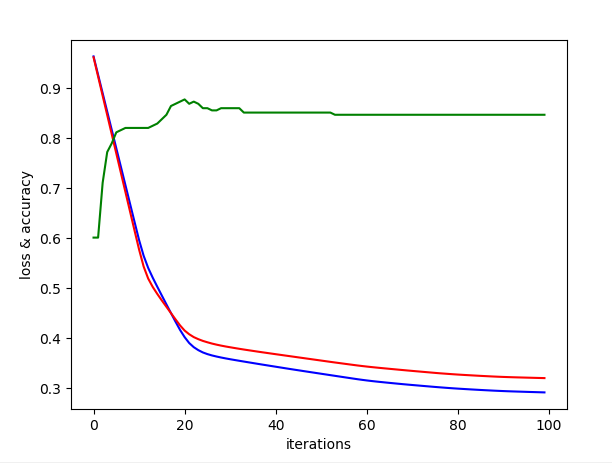


图9 学习率0.0001 迭代次数100

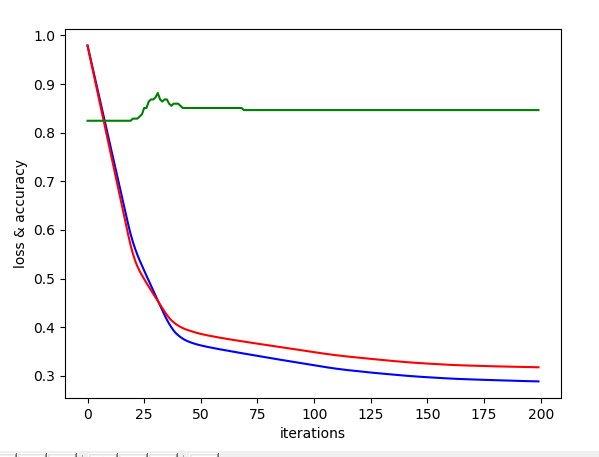


图10 学习率0.0001 迭代次数200 阈值0.5

## 预测结果（最佳结果）：

## loss曲线图：

线性回归：

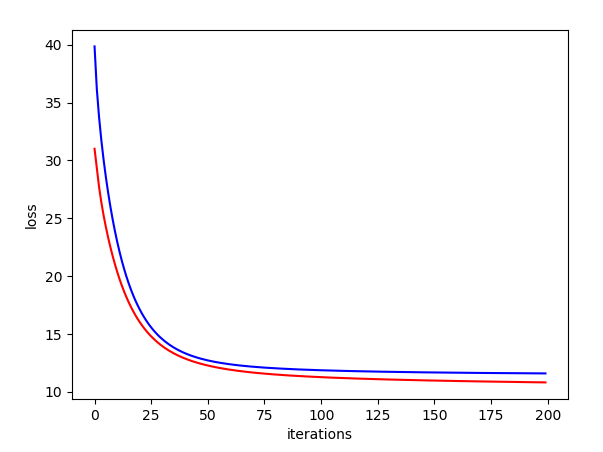


图11 学习率0.1 迭代次数200

线性分类：

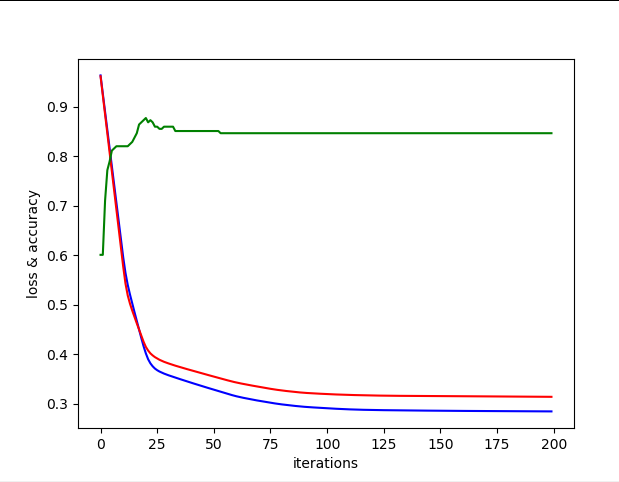


图12 学习率0.0001 迭代次数200

## 12.实验结果分析:

经过试验发现回归以及预测的效果与超参数的调整有着直接的联系，若训练次数太少，则Loss无法收敛到最小值。如果学习率设置过大，则会引起Loss波动，难以收敛。学习率设置过小，则要训练更多的次数才能收敛。

## 13.对比线性回归和线性分类的异同点：

线性回归于线性分类其实本质上是一致的，对一个给定的x，通过一个线性模型预测出一个y，只是对于回归来说一般一个x对应一个y，而分类问题则是把要预测的y规定为只有两个而已。

## 14.实验总结：

通过这次实验，我对线性回归和分类问题以及梯度下降的原理有了更深入的了解，线性回归和分类都是通过一个线性模型对数据进行预测，而梯度下降则是利用计算机强大的计算能力，一步步沿着模型参数的梯度的反方向前进，最终收敛到一个极小值。此外，第一次尝试了用Python进行矩阵运算。还有更好地意识机器学习中模型的预测效果受超参数的影响，参数调整是机器学习重要的一环。