

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**

**学 号 201530612033**

**邮 箱** [**478647411@qq.com**](mailto:478647411@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日上午9:00-12:00

## 3. 报告人: 李锐泳

## 4. 实验目的:

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

## 5. 数据集以及数据分析：

* 线性回归使用的是LIBSVM Data中的Housing数据，包含506个样本，每个样本有13个属性。
* 线性分类使用的是LIBSVM Data中的australian数据，包含690个样本，每个样本有14 个属性。

## 6. 实验步骤:

6.1 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。
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. 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。

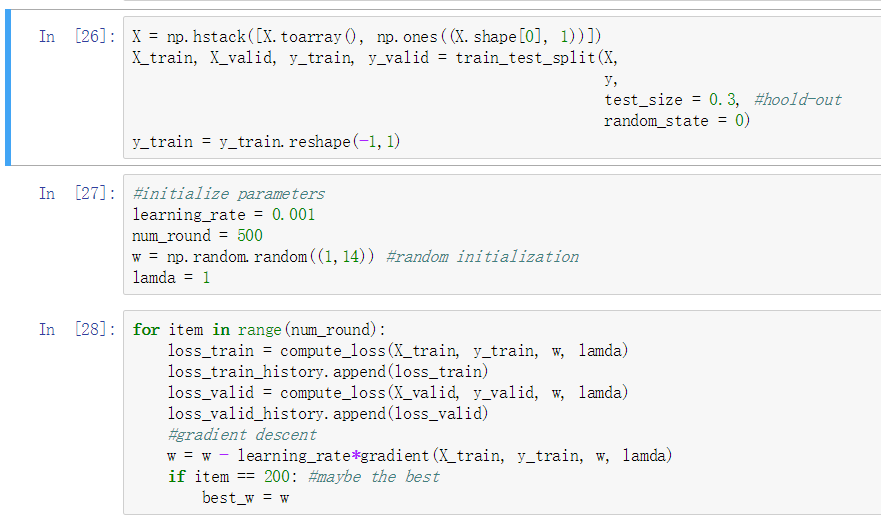
6.2 线性分类和梯度下降

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。
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:

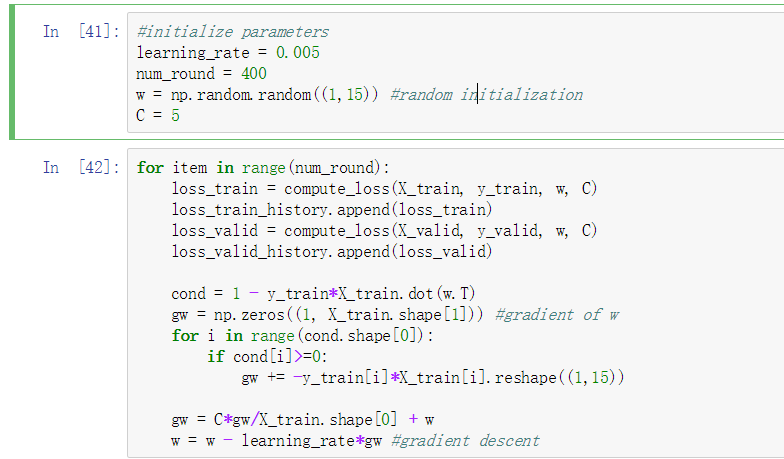
线性回归：





线性分类：





## 8. 选择的评估方法（留出法，交叉验证，k折交叉验证等）: 留出法

## 9. 模型参数的初始化方法: 随机初始化

## 10.选择的loss函数及其导数:

线性回归：

线性分类：

## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：

线性回归：

learning\_rate = 0.001

num\_round = 500

lamda = 1

线性分类：

learning\_rate = 0.005

num\_round = 400

C = 5

## 评估结果（根据选择的评估方法）：

## 预测结果（最佳结果）：

线性回归：

w: [[-3.18922665 -2.19946668 -0.90552238

-2.31820422 -1.33181976 1.32352814

1.43697608 -1.63212561 -1.51424543

-0.77851404 0.43708053 3.87844836

-1.87579209 4.58181091]]

learning\_rate = 0.001

lamda = 1

线性分类：

w: [[0.00594207 0.01776026 0.0394187 0.0795921 0.12730458 0.07669979 0.07092018 0.87742154 0.12281802 0.0465154 0.01508042 0.14583387

-0.01587872 0.00930932 0.12388915]]

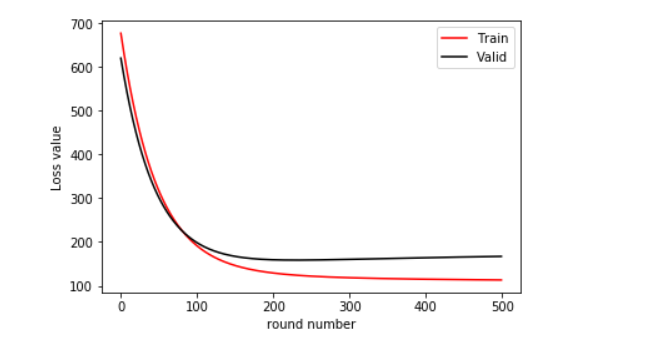
learning\_rate = 0.005

C = 5

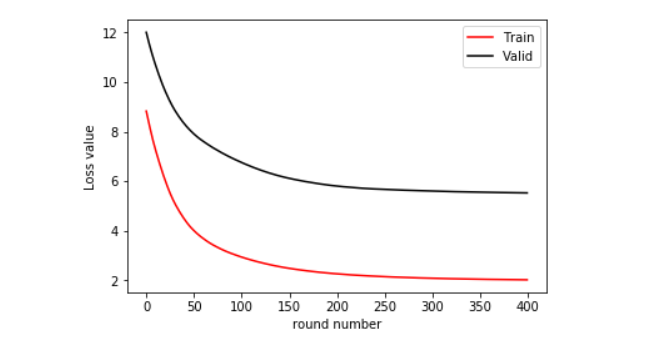
Accuracy = 0.8647342995169082（validation set）

## loss曲线图：

线性回归：



线性分类：



## 12.实验结果分析:

从loss曲线图可以直观看出，随着训练次数增加，loss不断减少。在线性回归中，轮次超过200后，有过拟合现象，在线性分类中，则在轮次达到400左右拟合效果仍然很好。

## 13.对比线性回归和线性分类的异同点：

相同：对于线性回归和线性分类，两者都是线性的模型，都有一个loss函数来评估拟合效果，都可以采用梯度下降方法来不断逼近最小的loss。

不同：回归问题主要是找到一堆对应的w参数值来拟合训练集，而分类问题主要是找到对应类别的label。

## 14.实验总结：

通过这次实验，我更加深入地了解了线性回归和线性分类以及梯度下降的原理，并自己动手实现，也学会了迭代地优化参数。