

**《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**

**邮 箱**

**指导教师**

**提交日期** **2017年 月 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:2015级软件工程5班甄淑怡

## 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

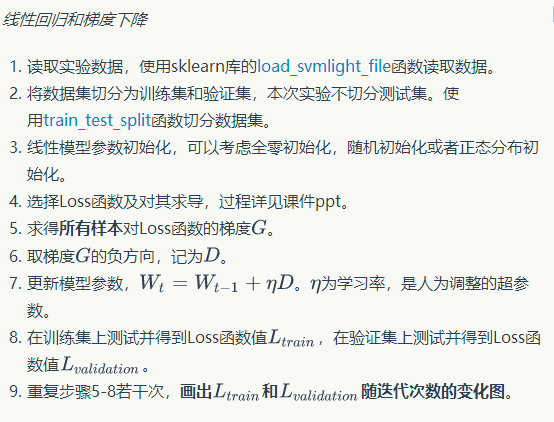
3.体会优化和调参的过程

## 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。

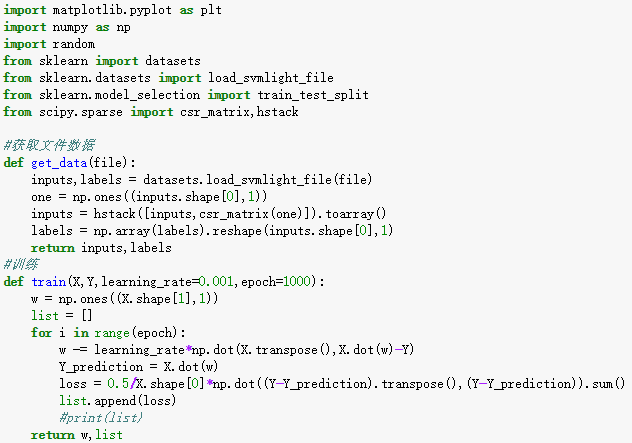
## 6. 实验步骤:



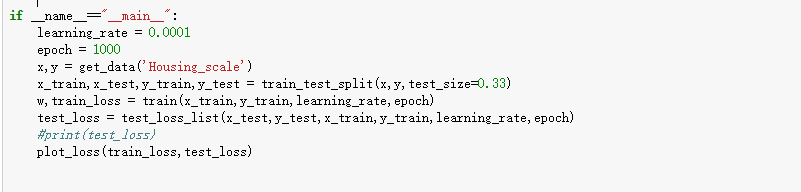


## 7. 代码内容:

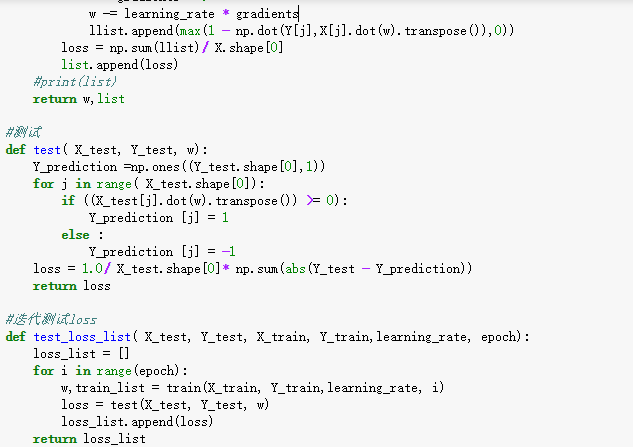
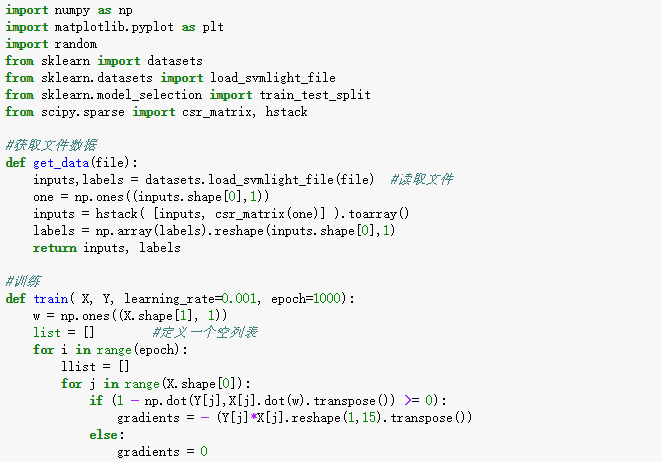
线性回归

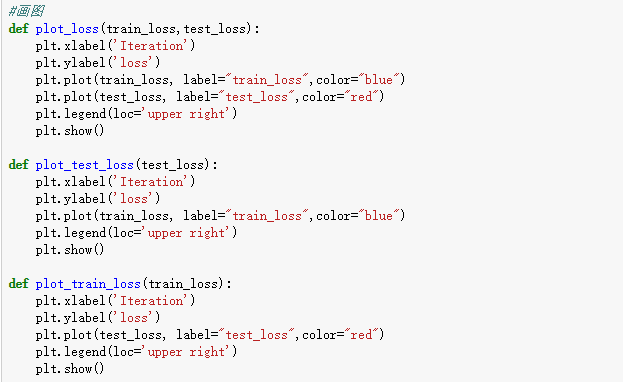


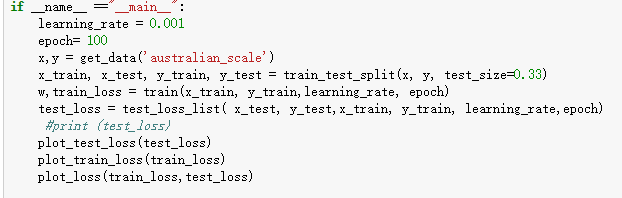




线性分类







## 8. 选择的评估方法（留出法，交叉验证，k折交叉验证等）:

线性回归：留出验证法

线性分类：留出验证法

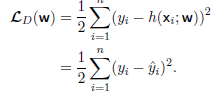
## 9. 模型参数的初始化方法:

线性回归：全零初始化

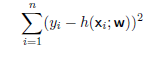
线性分类：全零初始化

## 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=0.0001

epoch = 1000

线性分类：learning\_rate = 0.001

epoch = 100

## 评估结果（根据选择的评估方法）：

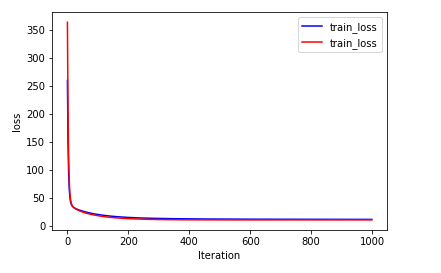
线性回归：训练数据集与测试数据集结果十分契合

线性分类：训练数据集与测试数据集所得结果存在不足，测试数据集的loss函数值不稳定

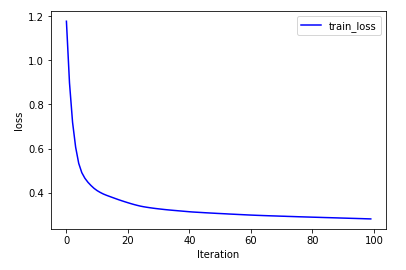
## 预测结果（最佳结果）：

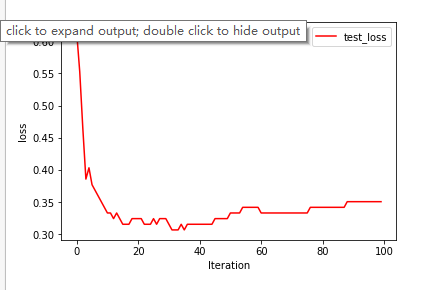
## loss曲线图：

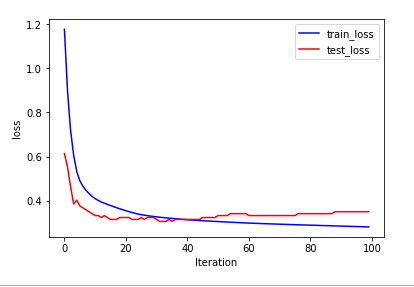
线性回归：



线性分类：







## 12.实验结果分析:

线性回归：线性回归，从loss函数的曲线可知，模型拟合效果很好，loss函数值从迭代第60次开始便处于一个较低值

线性分类：拟合结果不如线性回归，测试数据集的loss函数曲线有很大波动，模型拟合效果一般，这源于参数选择不够，经验不足。

## 13.对比线性回归和线性分类的异同点：

线性回归用于分析两个变量x,y之间的关系，而线性分类则是解决分类问题，需要一个阈值来将数据二分。

总的来说两个问题本质上都是一致的，就是模型的拟合，但是分类问题的y值更加离散化一些，而且同一个y值可能对应着一大批x，这些x是具有一定范围的。

## 14.实验总结：

这次的实验是我第一次自己用代码实现线性回归与线性分类，通过编码与找错，加深了对这两个算法的理解，也对建模以及对模型拟合的过程有了一定程度的了解，课堂上的知识不再仅限于纸上谈兵。