

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**  袁修竹

**学 号 201530613535**

**邮 箱 1225427085@qq.com**

**指导教师**  **吴庆耀**

**提交日期** **2017年12月 7日**

## 1. 实验题目: 线性回归和分类

## 2. 实验时间：2017年 12月 2日

## 3. 报告人:袁修竹

## 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；线性分类使用的是LIBSVM Data中的australian数据

## 6. 实验步骤:

**（1）线性回归和梯度下降**

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

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。

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。

选择Loss函数及对其求导。

求得所有样本对Loss函数的梯度G。

取梯度的负方向，记为D。

更新模型参数，Wt =Wt-1 + thetaD。theta为学习率，是人为调整的 超参数。

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

重复步骤5-8若干次，画出Ltrain和LossValidation随迭代次数的变 化图。

**（2）线性分类和梯度下降**

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

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。

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。

选择Loss函数及对其求导。

求得所有样本对Loss函数的梯度G。

取梯度的负方向，记为D。

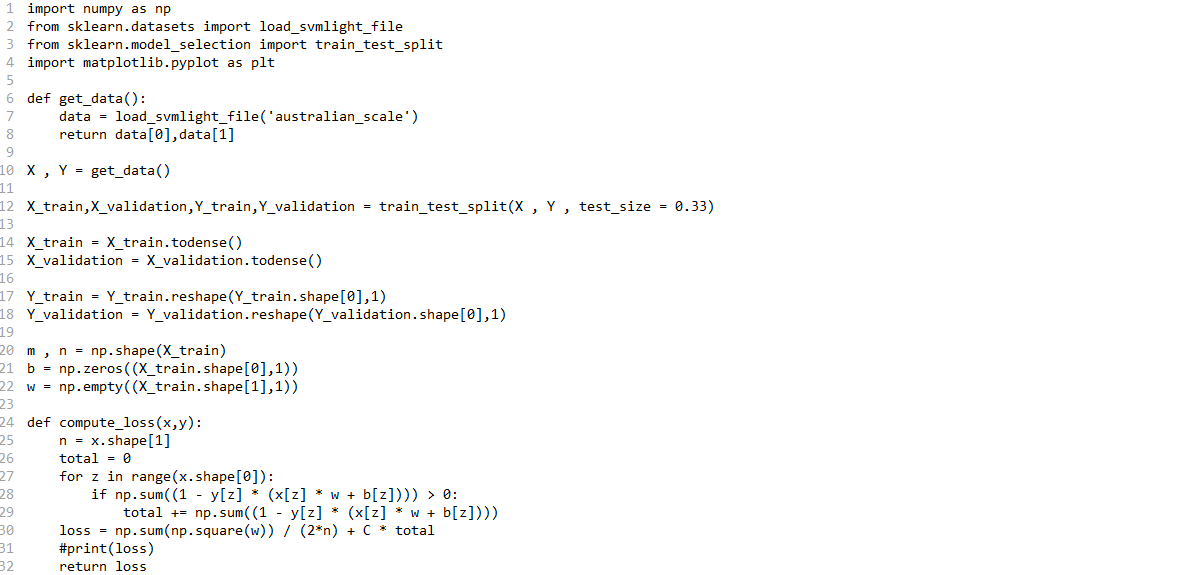
更新模型参数Wt =Wt-1 + thetaD。theta为学习率，是人为调整的超 参数。

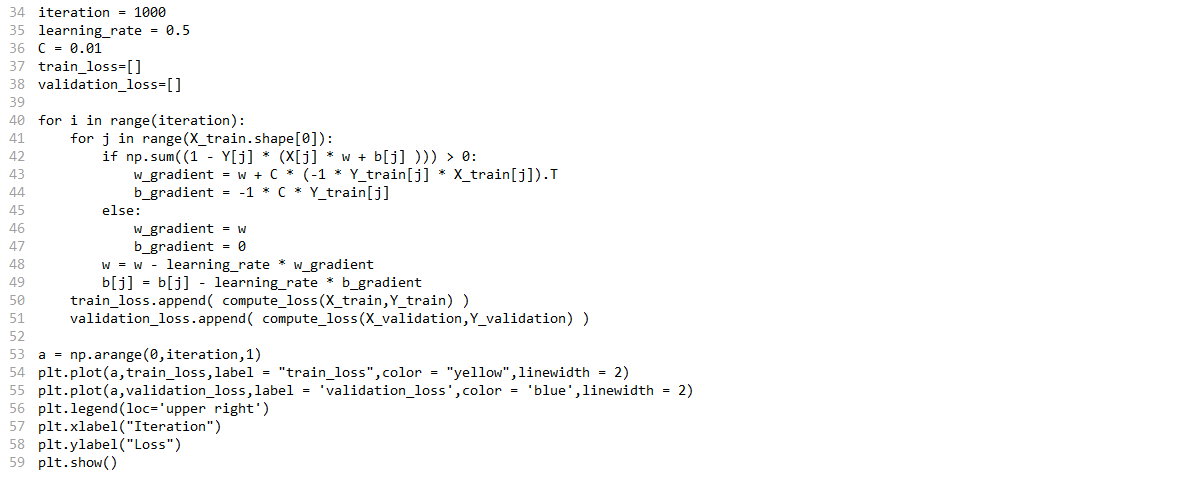
在训练集上测试并得到Loss函数值Ltrain，在验证集上测试并得到 LossValidation函数值。

重复步骤5-8若干次，画出Ltrain和LossValidation随迭代次数的变 化图。

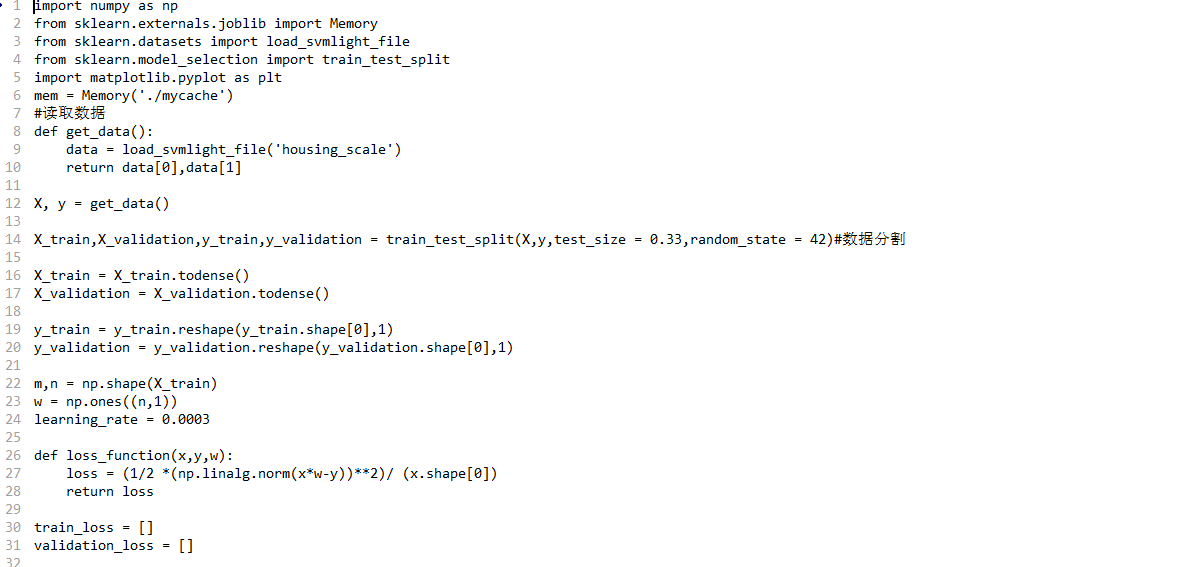
## 7. 代码内容:

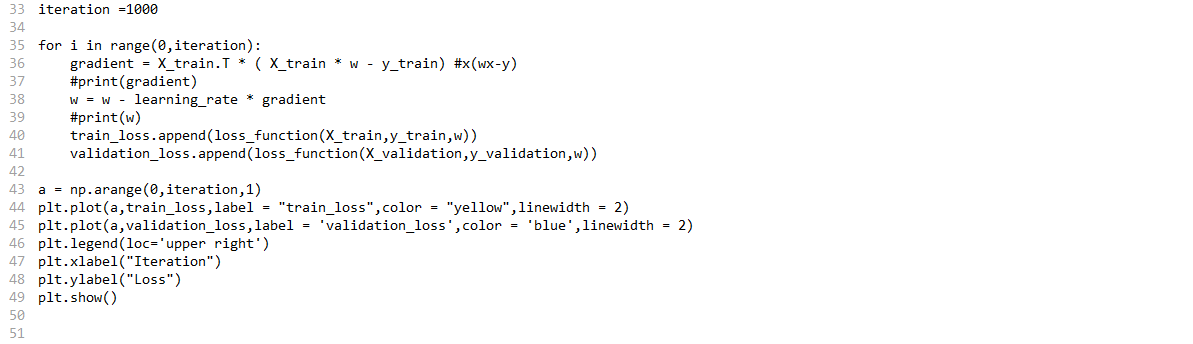
（1）线性回归





（2）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）:都是留出法

## 9. 模型参数的初始化方法:

## 

## 10.选择的loss函数及其导数:



## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：

（1）线性回归

theta=0.01

（2）线性分类

theta=0.05；C=0.01

## 评估结果（根据选择的评估方法）：

（1）线性回归

每次迭代更新theta，迭代次数固定

（2）线性分类

C不随迭代次数更新，theta进行更新

## 预测结果（最佳结果）：

（1）线性回归

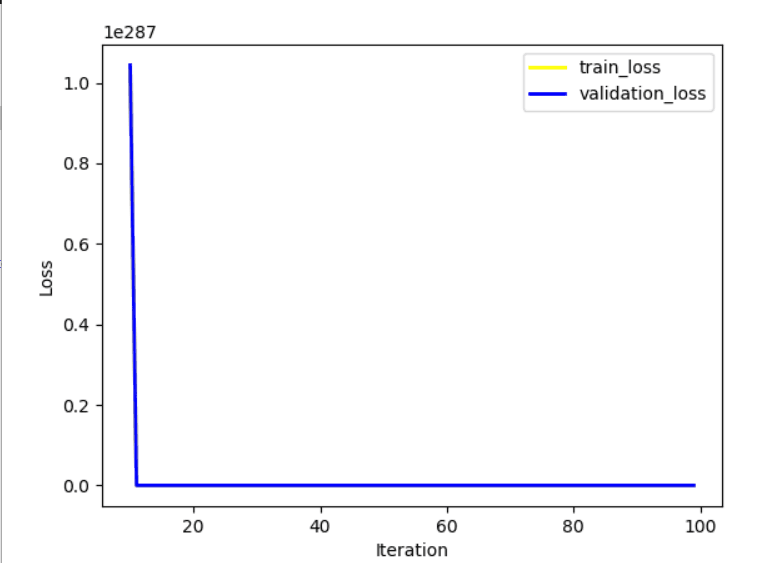
theta=0.0003

（2）线性分类

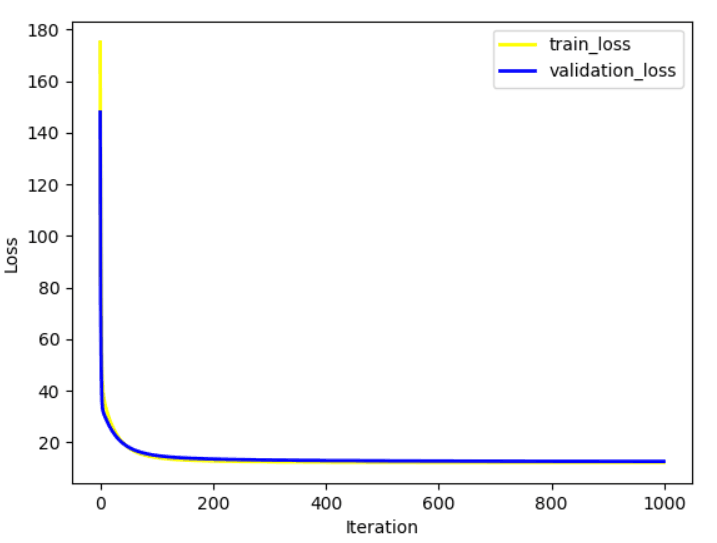
theta=0.05；C=0.01

## loss曲线图：

（1）线性分类



（2）线性回归



## 13.对比线性回归和线性分类的异同点：

## 线性回归，是利用数理统计中回归分析，来确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法，回归问题通常是用来预测一个值；类问题是用于将事物打上一个标签，通常结果为离散值。

## 14.实验总结：

通过此次实验掌握了线性回归和线性分类的基本方法，对过程有了进一步的了解与认识。