

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**   **柯杰智**

**学 号 201530611845**

**邮 箱 263439122@qq.com**

**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: 柯杰智

## 4. 实验目的:

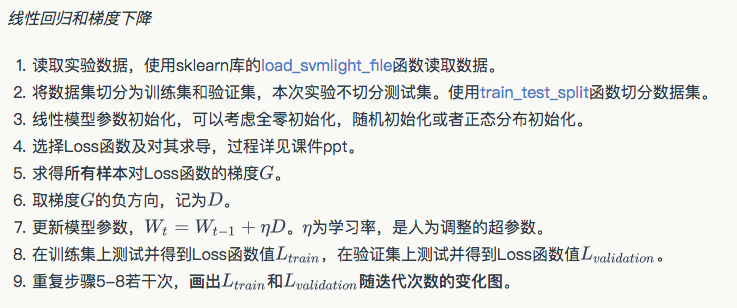
1. 进一步理解线性回归和梯度下降的原理。
2. 在小规模数据集上实践。
3. 体会优化和调参的过程。

## 5. 数据集以及数据分析：

线性回归使用的是[LIBSVM Data](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/)中的[Housing](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/regression.html#housing)数据，包含506个样本，每个样本有13个属性。其中label表示房价，最大值为50，最小值为5，平均值为22.53，中位数为21.2，极差为45，方差为84.42，标准差为9.19。

线性分类使用的是[LIBSVM Data](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/)中的[australian](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html#australian)数据，包含690个样本，每个样本有14 个属性。其中label表示该记录的类别，+1为正类，-1为负类。其中正类有307个样本，负类有383个样本。

## 6. 实验步骤:

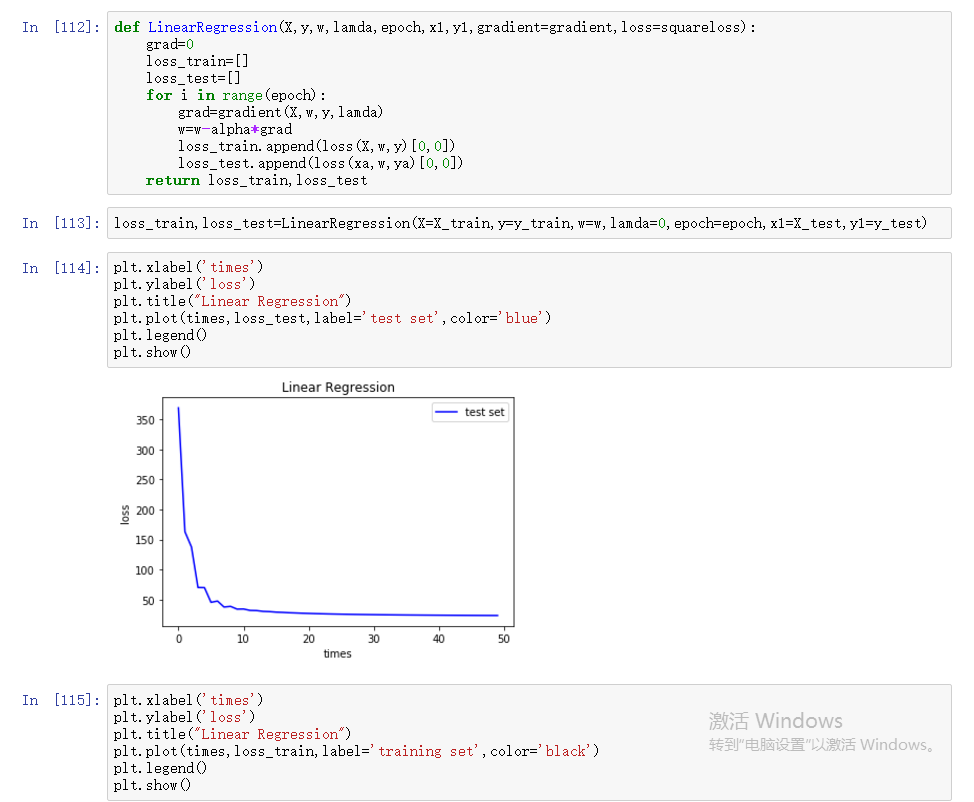


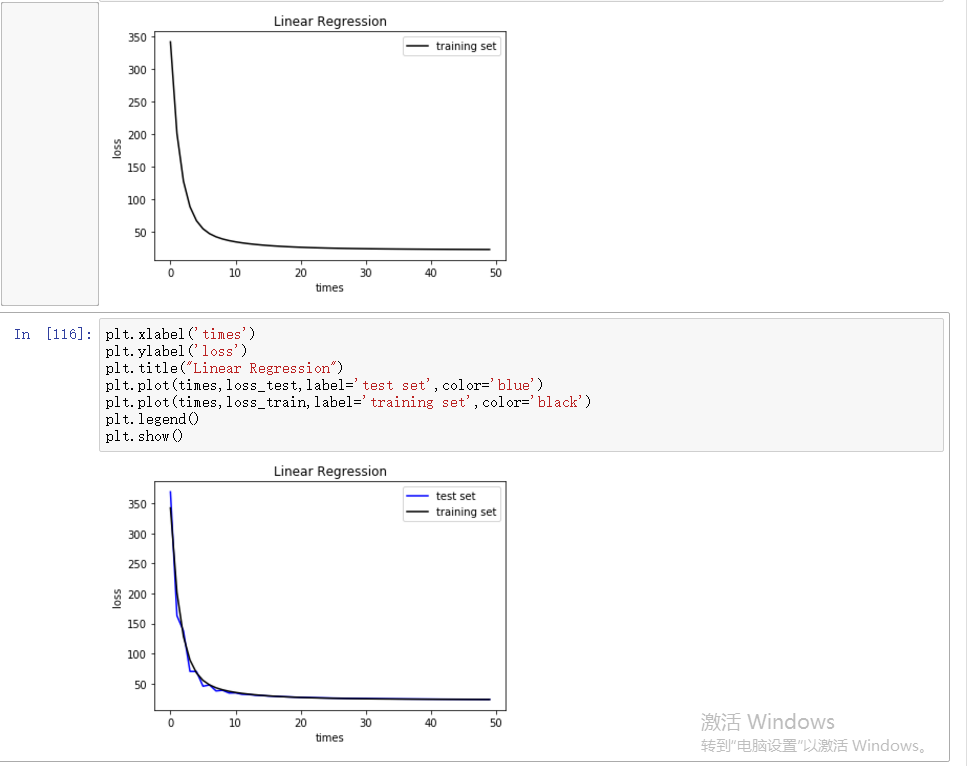


## 7. 代码内容:

线性回归：

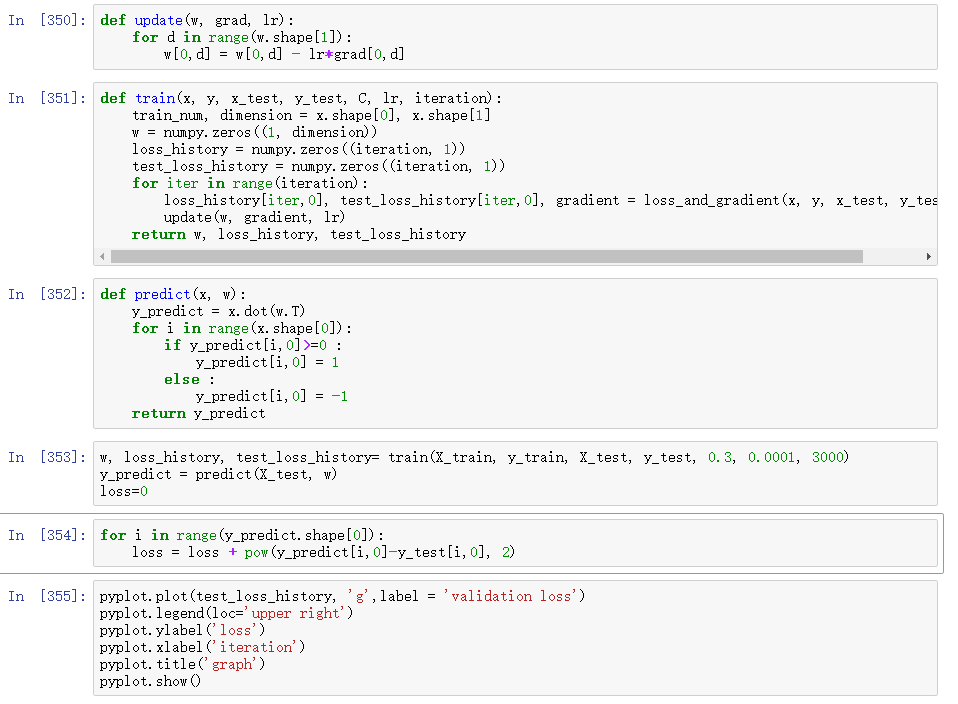


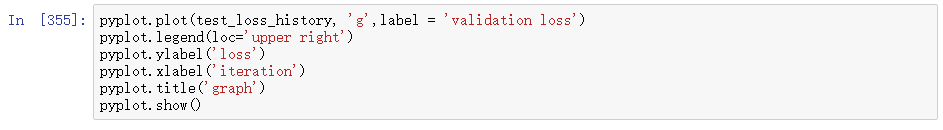


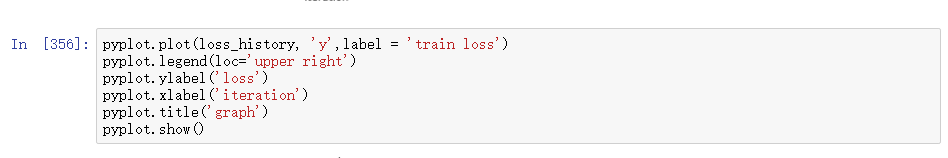


线性分类：









## 8. 选择的评估方法:

选择的评估方法是留出法。将训练集和测试集以7:3的比例随机划分，用训练集训练线性回归和线性分类模型，并在测试集上计算损失，记录训练集的损失和测试集的损失并做出图像。

## 9. 模型参数的初始化方法:

全零初始化

## 10.选择的loss函数及其导数:

loss function:

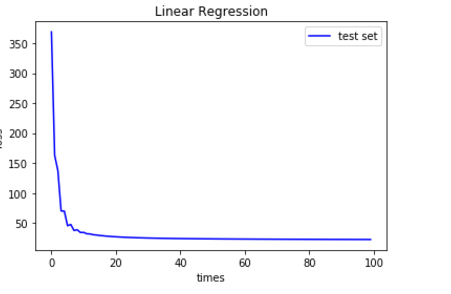
gradient:

## 11.实验结果和曲线图:

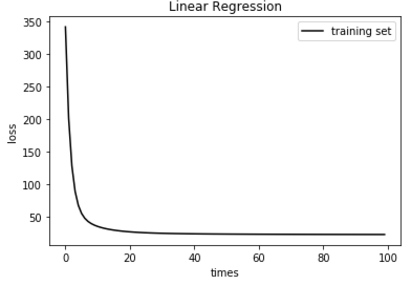
## 超参数选择（η,epoch等）：

η=0.001，epoch=100

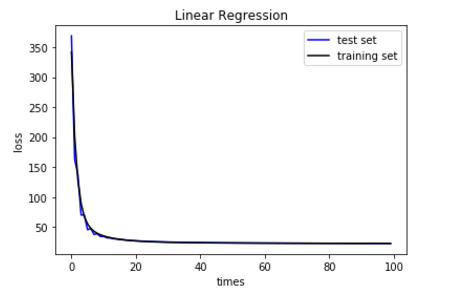
## 评估结果（根据选择的评估方法）：



## 预测结果（最佳结果）：



## loss曲线图：



12.实验结果分析: 本实验以梯度下降为方法对线性回归模型（ ）进行训练。在实验中，θ被初始化为零，在每一次迭代更新θ参数时，利用损失函数对θ的导数，乘以学习率η得到新的θ，重复此过程直至θ收敛。试验结果的loss图显示了每一次迭代对θ更新后在训练集和测试集上的损失，我们从图中可以看到随着迭代次数的增加，在训练集和测试集上的损失都是下降的，并稳定在一个较小的值上，说明模型训练完成，θ收敛到局部最优解，满足要求。

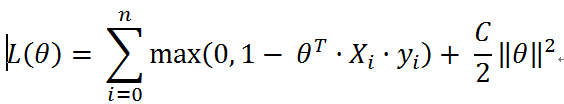
## 8. 选择的评估方法:

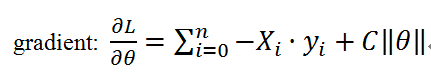
选择的评估方法是留出法。将训练集和测试集以7:3的比例随机划分，用训练集训练线性回归和线性分类模型，并在测试集上计算损失，记录训练集的损失和测试集的损失并做出图像。

## 9. 模型参数的初始化方法:

全零初始化

## 10.选择的loss函数及其导数:

loss function: 

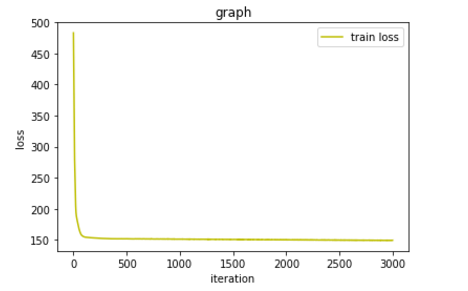
gradient: 

## 11.实验结果和曲线图:

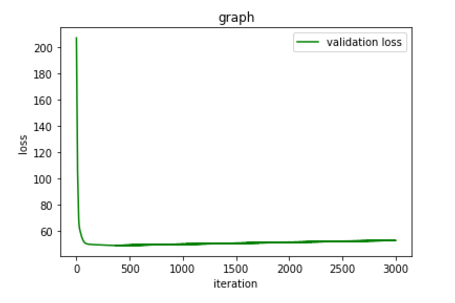
## 超参数选择（η,epoch等）：

η=0.0001，epoch=3000

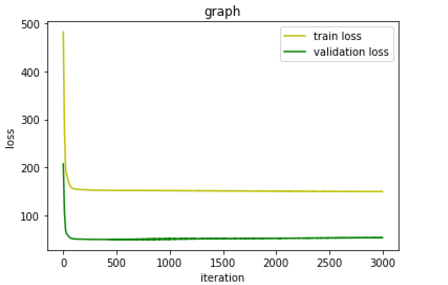
## 评估结果（根据选择的评估方法）：



## 预测结果（最佳结果）：



## loss曲线图：



12. **本实验以梯度下降为方法对线性回归模型（ ，y ）进行训练。在实验中，θ被初始化为零，在每一次迭代更新θ参数时，利用损失函数对θ的导数，乘以学习率η得到新的θ，重复此过程直至θ收敛。试验结果的loss图显示了每一次迭代对θ更新后在训练集和测试集上的损失，我们从图中可以看到随着迭代次数的增加，在训练集和测试集上的损失都是下降的，并稳定在一个较小的值上，说明模型训练完成，θ收敛到局部最优解，满足要求。**

13.对比线性回归和线性分类的异同点：同：

对于这两类问题都有以下几个步骤,

• 如何选取一个合理的模型。

• 选择一个合适的损失函数

• 采取梯度下降法求出最好的模型参数

异：

总的来说两个问题本质上都是一致的，就是模型的拟合。而且， 同一个y值可能对应着一大批的x, 这些x是具有一定范围的。 所以分类问题更多的是 (一定区域的一些x) 对应着 (一个y). 而回归问题的模型更倾向于 (很小区域内的x，或者一般是一个x) 对应着 (一个y).

## 14.实验总结：