

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员 薛志翔**

**学 号 201530613313**

**邮 箱 545589205@qq.com**

**指导教师**  **吴庆耀**

**提交日期** **2017年12月8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:薛志翔

## 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。使用的是scaled版本，将其切分为训练集，验证集。

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。使用的是scaled版本，将其切分为训练集，验证集。

## 6. 实验步骤:



## 7. 代码内容:

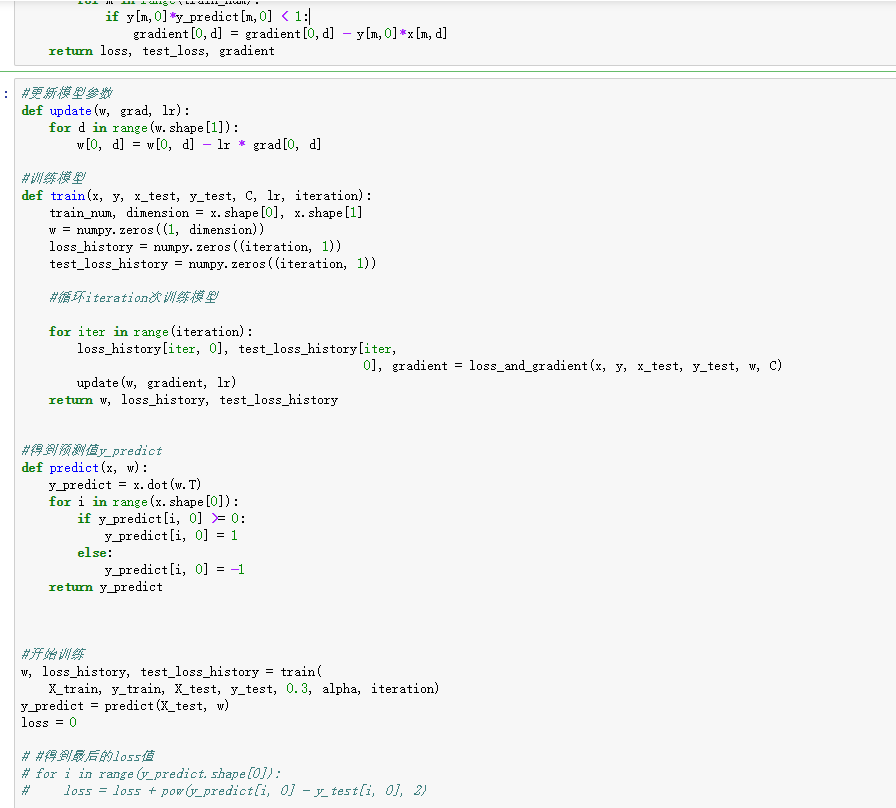
（针对线性回归和线性分类分别填写8-12内容）

线性回归：



线性分类：







线性回归：

## 8. 选择的评估方法（留出法，交叉验证，k折交叉验证等）:

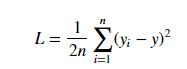
使用了留出法，将训练集和测试集以0.67：0.33的比例划分，用训练集训练模型，并在测试集和训练集上计算loss值。

## 9. 模型参数的初始化方法:

模型参数的初始化方法是全零初始化。

## 10.选择的loss函数及其导数:

Loss函数：



导数：

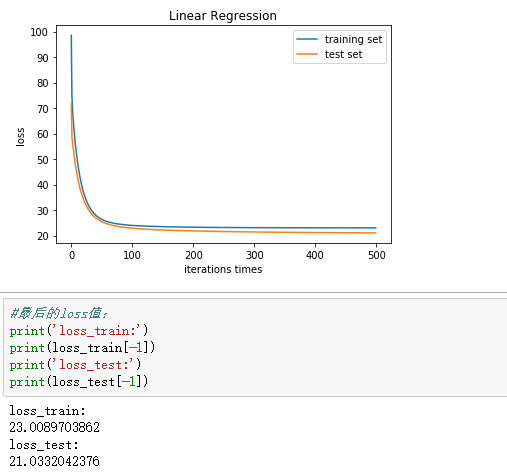


## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：

η=0.001，epoch=1000

## 评估结果（根据选择的评估方法）：



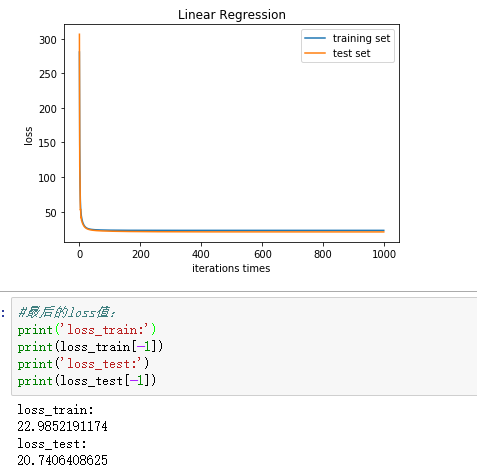
loss\_train:

23.0089703862

loss\_test:

21.0332042376

## 预测结果（最佳结果）：



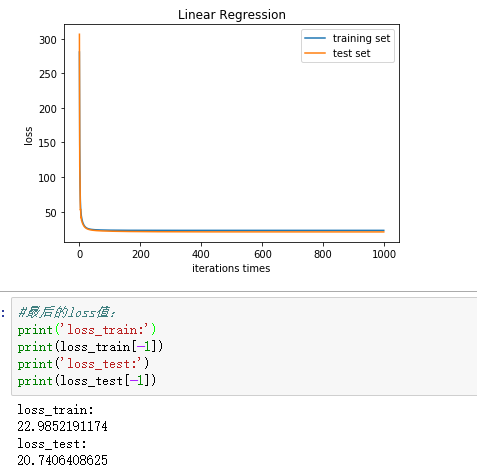
loss\_train:

22.9852191174

loss\_test:

20.7406408625

## loss曲线图：



## 12.实验结果分析:

使用梯度下降的方法进行线性回归。线性回归模型的参数采用全零初始化，在选取了合适的η学习率，进行多次迭代之后，loss最后收敛到一个较小的值，说明我们的模型训练完成，取得了不错的效果。

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）:

使用了留出法，将训练集和测试集以7：3的比例划分，用训练集训练模型，并在测试集和训练集上计算loss值。

## 9. 模型参数的初始化方法:

模型参数的初始化方法是全零初始化。

## 10.选择的loss函数及其导数:

Loss函数：屏幕剪辑

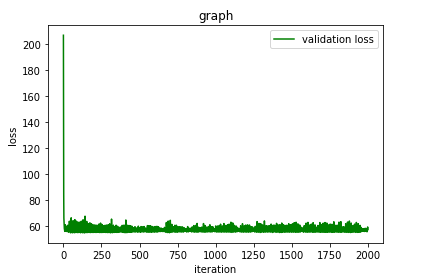
gradient:屏幕剪辑

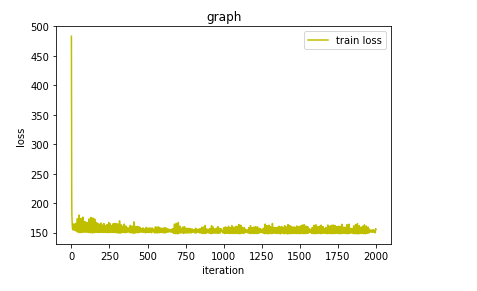
## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：

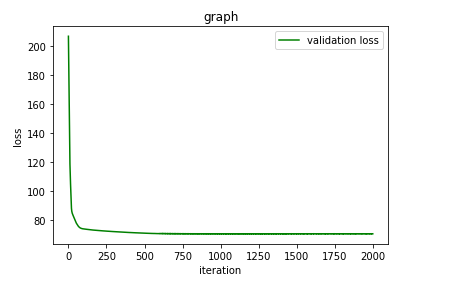
η= 0.0001 epoch = 2000

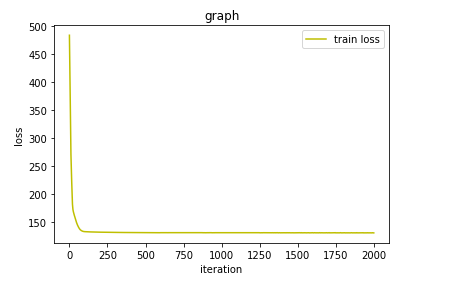
## 评估结果（根据选择的评估方法）：



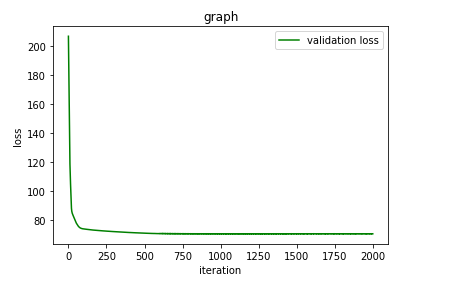


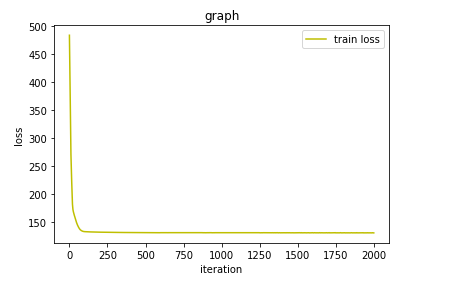
## 预测结果（最佳结果）：





## loss曲线图：





## 12.实验结果分析:

使用梯度下降的方法进行线性分类。模型的参数采用全零初始化，在选取了合适的η学习率，进行多次迭代之后，loss最后收敛到一个较小的值，说明我们的模型训练完成，取得了不错的效果。

## 13.对比线性回归和线性分类的异同点：

相同点：

解决这两类问题，本质上都是在构造一个模型，并选取一个合适的loss函数及其梯度函数，然后通过调参以及迭代，得到最符合数据的参数，也就是对应的模型。

不同点：

两种问题的loss函数显著不同。

## 14.实验总结：

在这一次实验中，我开始将课上学到的机器学习的理论知识初步运用起来。我发现了理论与实际操作之间还是有很大的不同，将理论知识合理利用起来是我们的目标。在这个过程中，我遇到了不少困难，模型的选择，loss函数的选择，参数的选择等等，都是我们需要解决的难题。在查阅资料，询问同学之后，我解决了问题，并取得了不错的实验成功。