

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**  **王博**

**学 号 201530381731**

**邮 箱 765246150@qq.com**

**指导教师**

**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

## 3. 报告人:王博

## 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。请自行下载scaled版本，并将其切分为训练集，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。请自行下载scaled版本，并将其切分为训练集，验证集。

## 6. 实验步骤:

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，过程详见课件ppt。

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。

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，过程详见课件ppt。

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。

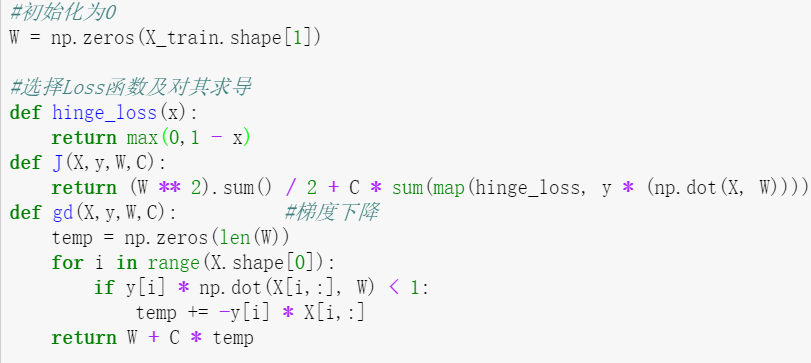
9.重复步骤5-8若干次，画出和随迭代次数的变化图。

## 7. 代码内容:

回归



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）:

## 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函数：J= (X\*theta-y).T\*(X\*theta-y)/(2\*m)

导数：G= X.T\*(y-X\*theta)/m

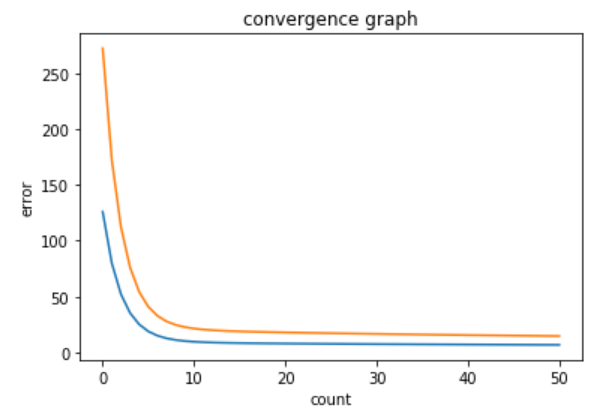
## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：η=0.005

## 评估结果（根据选择的评估方法）：

## 预测结果（最佳结果）：error=20

## loss曲线图：



## 12.实验结果分析:

迭代20次后收敛，learning\_rate=0.005时收敛效果比较明显

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）:

交叉验证

## 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:

hinge\_loss=max(0,1,-x)

loss函数：(W \*\* 2).sum() / 2 + C \* sum(map(hinge\_loss, y \* (np.dot(X, W))))

导数：

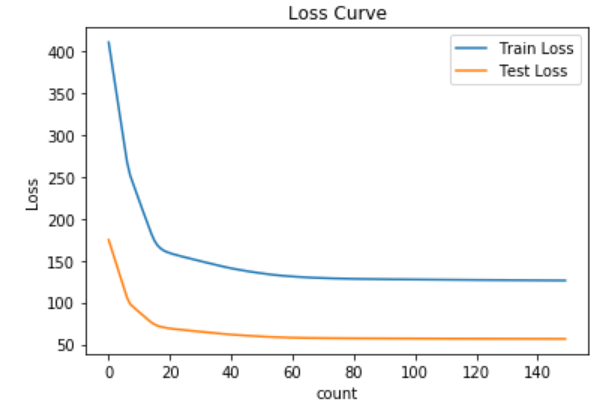
## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：η

## 评估结果（根据选择的评估方法）：交叉验证

## 预测结果（最佳结果）：Test\_loss=60

## loss曲线图：



## 12.实验结果分析:

Test集比training集的loss函数还要低，不知是什么原因。。。

## 13.对比线性回归和线性分类的异同点：

同：回归和分类本质上都是对一个新的输入预测其输出值，本质是一样的

异：回归是连续的，分类是离散的。输出的取值范围不同。

## 14.实验总结：

初次尝试python实现线性回归和线性分类，由于不熟悉，在numpy上花了不少时间和功夫。通过这次试验，我深刻感受到了理论和实践结合的重要性，调参也是需要技巧的，不能随便拿个数就来试。由于时间紧迫，做的很不仔细，下次一定谨记教训，认真书写报告。