

**《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 201530611951**

**邮 箱 786699266@qq.com**

**指导教师**  **吴庆耀**

**提交日期** **2017年 12 月 7日**

## 实验题目: 线性回归、线性分类与梯度下降

## 实验时间：2017年 12 月 2日

## 3. 报告人: 李嘉灏

## 4. 实验目的:

### 线性回归和梯度下降

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

### 线性分类和梯度下降

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：

Housing\_scale数据集，共506个样品，每个样品有13个特点。您需要下载缩放版。下载后，你应该把它分成训练集、验证集。

Australian\_scale数据集，包括690个样本，每个样本有14个特点。您需要下载缩放版。下载后，你应该把它分成训练集、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G。

6. 取梯度G的负方向，记为D。

7. 更新模型参数Wt=Wt-1+ηD，。η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G。

6.取梯度G的负方向，记为D。

7.更新模型参数，。为学习率，是人为调整的超参数。

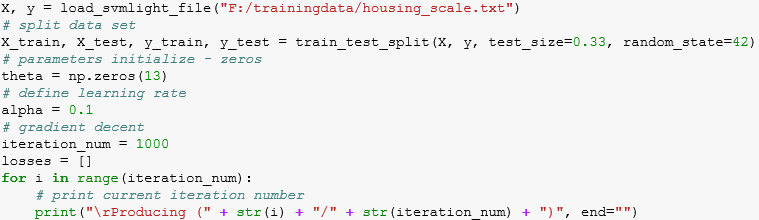
8.选择合适的阈值，将计算结果大于阈值的标记为正类，反之为负类。在训练集上测试并得到Loss函数值，在验证集上测试并得到Loss函数值。

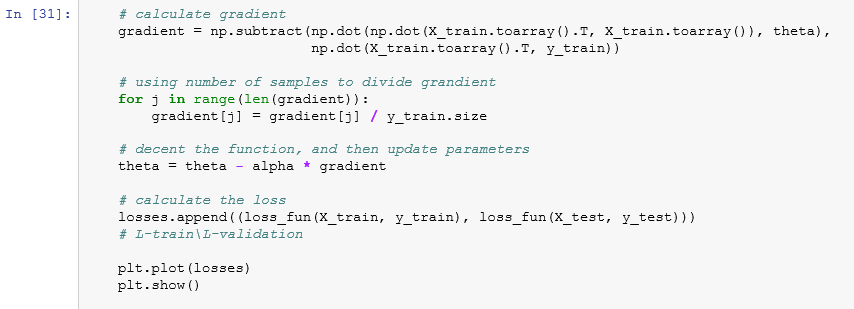
9.重复步骤5-8若干次，画出和随迭代次数的变化图。

## 7. 代码内容:

（针对线性回归和线性分类分别填写8-12内容）

**线性回归的梯度下降：**

****

****

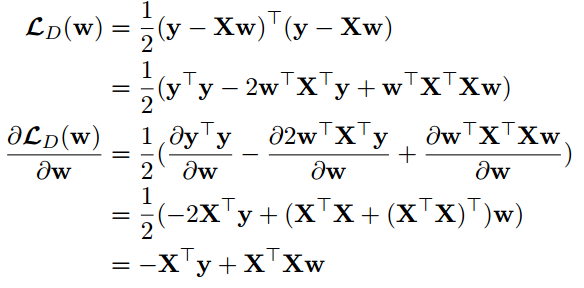
### 线性分类



## 8. 选择的评估方法（留出法，交叉验证，k折交叉验证等）:

## 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:



## 11.实验结果和曲线图:

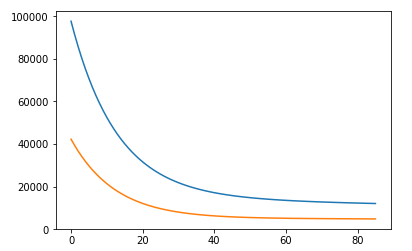
## 超参数选择（η,epoch等）：η（学习率）

## 评估结果（根据选择的评估方法）：

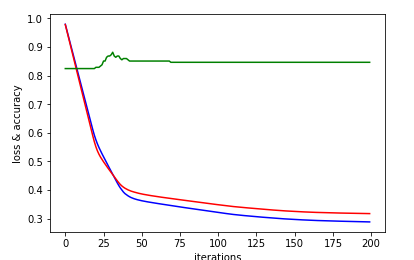
## 预测结果（最佳结果）：测试集Loss曲线最终值会略大于训练集Loss曲线

## loss曲线图：

**线性回归：**

****

**线性分类：**

****

## 12.实验结果分析:

线性分类实验结果基本正确，但是线性回归验证集的实验数据比训练集的实验数据要大，调整了学习率还有样本大小以及训练次数等参数也没有变化，原因不明

## 13.对比线性回归和线性分类的异同点：

总的来说两个问题本质上都是一致的，就是模型的拟合（匹配）。

分类问题的y值(也称为label), 更离散化一些. 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：

此次试验我们分别在较小的数据集上对线性分类和线性回归以及其所对应的梯度下降情况作了模拟和实验，通过此次试验，我更加明白了这两者之间的共同点和差异所在，也对我机器学习有很大的帮助