

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**  胡日扬

**学 号 201530611616**

**邮 箱 1126453431@qq.com**

**指导教师**  **谭明奎**

**提交日期** **2017年 12 月 7 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: 胡日扬

## 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。

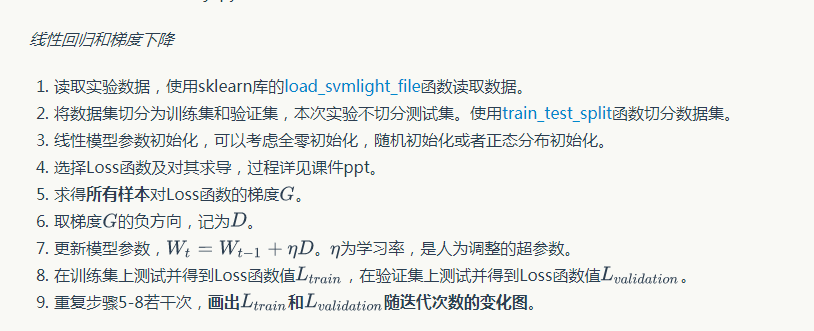
## 数据集以及数据分析：

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13个属性。

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14个属性。

## 实验步骤:

线性回归和梯度下降：



线性分类与梯度下降：



## 7. 代码内容:

## 线性回归：

**def** loss(W, X, y):

ans = 0

m = X.shape[0]

**for** i **in** range(0, m):

ans += (X[i].dot(W)-y[i])\*\*2

**return** ans/(2\*m)

**def** update(W, X, y):

ret = np.zeros((X.shape[1],1))

H = X.dot(W)-y

**for** i **in** range(0,X.shape[1]):

ret[i] = np.sum(H \* X[:, i:i+1]) / X.shape[1]

**return** ret

线性分类：

**def** loss(W, b, X, y, C):

ans = 0

m = X.shape[0]

**for** i **in** range(0, m):

ans += max(0, 1-y[i]\*(X[i].dot(W)+b))

**return** ans\*C + np.sum(W\*W)/2

**def** Gwb(x, y, W, b):

**if** 1-y\*(x.dot(W)+b) >= 0:

**return** -y\*x, -y

**else**:

**return** np.zeros(len(x)), 0

## 选择的评估方法（留出法，交叉验证，k折交叉验证等）:

留出法

## 模型参数的初始化方法:

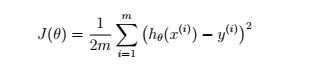
线性回归：随机初始化

线性分类：全零初始化

## 选择的loss函数及其导数:

线性回归：

Loss函数：

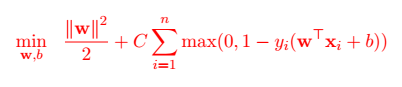


导数：

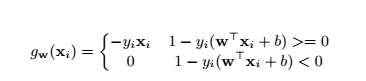


线性分类：

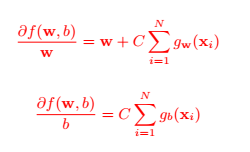
Loss函数：



导数：







## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：

线性回归：学习率=0.0001，迭代次数=1000

线性分类：学习率=0.0001， 迭代次数=250，C=0.8

## 评估结果（根据选择的评估方法）：

线性回归：

Loss值=29.6

线性分类：

Loss值=0.236

## 预测结果（最佳结果）：

线性回归：

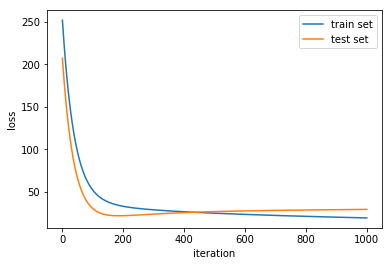
Loss值=20.1

线性分类：

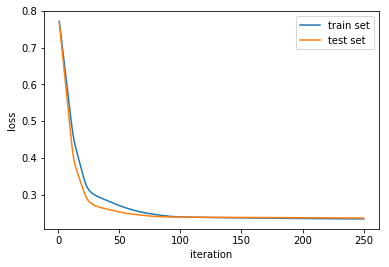
Loss值=0.216

## loss曲线图：

线性回归：



线性分类：



## 实验结果分析:

线性回归：调参之后loss值收敛到较小的值。

线性分类：调参之后loss值收敛到较小的值且正确率达85%。

## 13.对比线性回归和线性分类的异同点：

异：线性回归用来处理回归问题，线性分类用来处理分类问题， 且线性分类使用了核函数

同：都是使用线性的函数来拟合

## 14.实验总结：

本次实验中我学会了Jupyter notebook的使用和玄学调参技术，初步了解了调参的艺术。