

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**  黄浩填

**学 号 201530611708**

**邮 箱** [**1025686131@qq.com**](mailto:1025686131@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 8 日

## 3. 报告人:黄浩填

## 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。

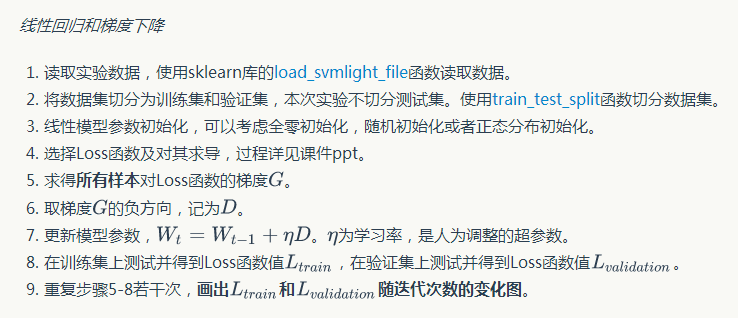
（3）体会和优化调参的过程。

## 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，并将其切分为训练集，验证集。

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，并将其切分为训练集，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：

def loss (X, W, y):

tmp = X.dot(W) - y

return (tmp \* tmp).sum() / X.shape[0] / 2

def grad (X, W, y):

tmp = X.dot(W) - y

return X.T.dot(tmp) / X.shape[0]

线性分类：

Def loss (X, W, y):

ret = 0

tmp = X.dot(W) \* y

for i in range(0,tmp.shape[0]):

if tmp[i][0] < 1 :

ret += 1 - tmp[i][0]

global C

ret \*= C

ret += ((W \* W).sum() - W[W.shape[0]-1][0] \* W[W.shape[0]-1][0]) / 2

return ret / X.shape[0]

def grad (X, W, y):

tmp = X.dot(W) \* y

X\_tmp = []

y\_tmp = []

for i in range(0,tmp.shape[0]):

if tmp[i][0] <= 1 :

X\_tmp.append(X[i])

y\_tmp.append(y[i])

X\_tmp = np.array(X\_tmp)

y\_tmp = np.array(y\_tmp)

global C

tmp = X\_tmp.T.dot(y\_tmp) \* C

ret = W - tmp

ret[ret.shape[0]-1][0] = -1.0 \* C \* y\_tmp.sum()

return ret

## 8. 选择的评估方法（留出法，交叉验证，k折交叉验证等）:

留出法

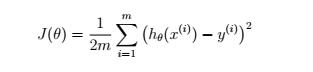
## 9. 模型参数的初始化方法:

随机初始化

## 10.选择的loss函数及其导数:

线性回归：

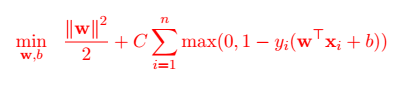
Loss 函数：



导数：

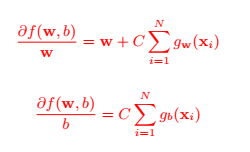
D:\我的文档\Tencent Files\1025686131\Image\C2C\~LZ99WG]~LGW(B6QD7N@MYQ.png

线性分类：



D:\我的文档\Tencent Files\1025686131\Image\C2C\~LZ99WG]~LGW(B6QD7N@MYQ.png

D:\我的文档\Tencent Files\1025686131\Image\C2C\0486YNFJFP@PQ4Q3Z72[71N.png



## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：

线性回归：epoch=10000

η=0.02

线性分类： epoch=5000

C=0.1

η=0.002

## 评估结果（根据选择的评估方法）：

线性回归： loss = 10.3978

线性分类： loss = 0.0334

## 预测结果（最佳结果）：

线性回归： loss = 8.234

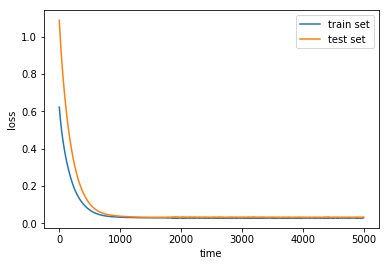
线性分类： loss = 0.0332

## loss曲线图：

线性回归：



线性分类：



## 12.实验结果分析:

Loss函数收敛达到预期结果，预测值也较为符合

## 13.对比线性回归和线性分类的异同点：

都是求多维空间中的一个超平面，而线性回归是使点极可能分布在超平面上，而线性分类则是将相同类型的点划分在超平面的同一边

## 14.实验总结：

学会体会到了调参对实验结果的影响，学会了Python在机器学习方面的应用。