

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**   **李嘉明**

**学 号 201530611975**

**邮 箱 707148473@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，线性分类和梯度下降

## 2. 实验时间：2017年 12 月 2 日

## 3. 报告人: 李嘉明

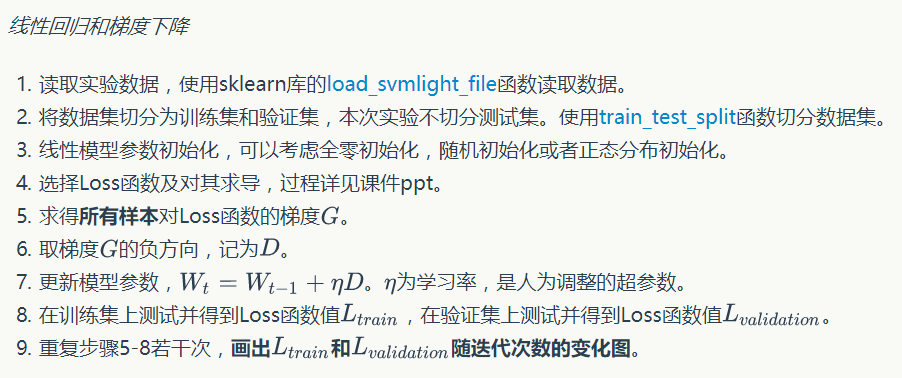
## 4. 实验目的:

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](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/" \t "https://www.zybuluo.com/chenyaofo/note/_blank)中的[Housing](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/regression.html" \l "housing" \t "https://www.zybuluo.com/chenyaofo/note/_blank)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](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/" \t "https://www.zybuluo.com/chenyaofo/note/_blank)中的[australian](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html" \l "australian" \t "https://www.zybuluo.com/chenyaofo/note/_blank)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。**

## 实验步骤:





## 代码内容:

线性回归：

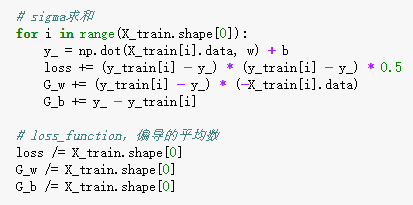
加载数据



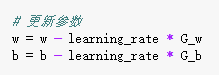
将数据分割为训练集和验证集



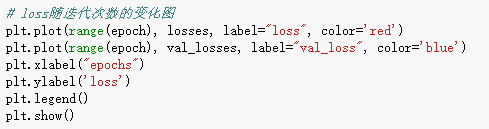
计算loss\_function及w，b的偏导数



通过learning\_rate更新参数



画图



线性分类：

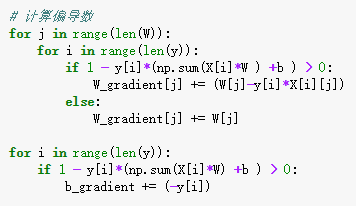
加载数据



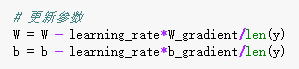
将数据分割为训练集和验证集



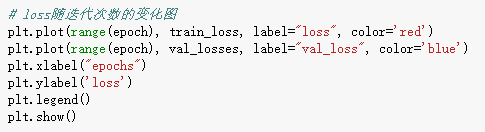
计算loss\_function及w，b的偏导数



通过learning\_rate更新参数



画图



（针对线性回归和线性分类分别填写8-12内容）

## 选择的评估方法（留出法，交叉验证，k折交叉验证等）:

将数据集按照8：2拆分为训练集和验证集，用训练集数据训练调参，在验证集上进行测试

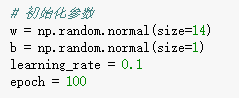
## 模型参数的初始化方法:

正态分布初始化

线性回归：

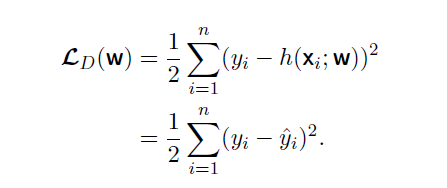


线性分类

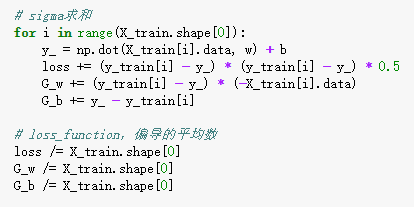


## 选择的loss函数及其导数:

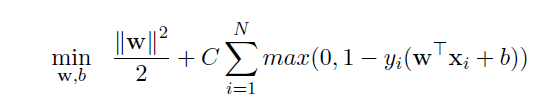
线性回归：

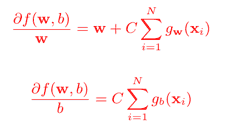
 

代码如下：

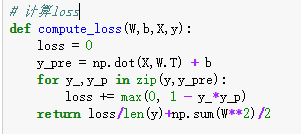


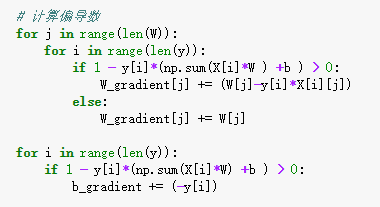
线性分类：





代码如下：





## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：

线性回归：

Learning\_rate = 0.001/0.01

Epoch = 1000

参数w,b 选择正态分布初始化

线性分类：

Learning\_rate = 0.001/0,01

Epoch = 100

参数w,b 选择正态分布初始化

## 评估结果（根据选择的评估方法）：

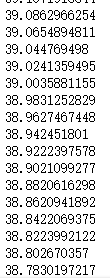
线性回归：

Learning\_rate = 0.001

Epoch = 1000

参数w,b 选择正态分布初始化

Train\_loss随epoch变化如下

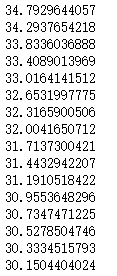


Learning\_rate = 0.001

Epoch = 1000

参数w,b 选择正态分布初始化

Train\_loss随epoch变化如下



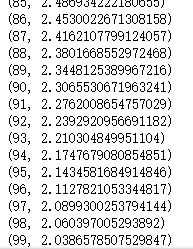
线性分类：

Learning\_rate = 0.01

Epoch = 1000

参数w,b 选择正态分布初始化

train\_loss 随epoch变化如下表

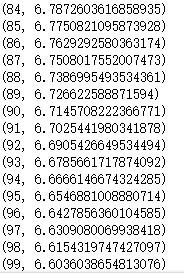


Learning\_rate = 0.01

Epoch = 1000

参数w,b 选择正态分布初始化

train\_loss 随epoch变化如下表



## 预测结果（最佳结果）：

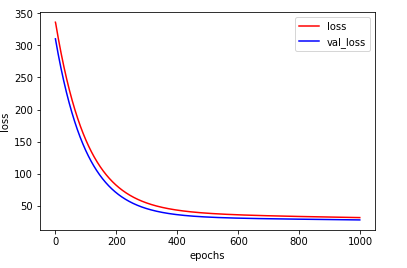
线性回归：

Learning\_rate = 0.01

Epoch = 1000

参数w,b 选择正态分布初始化

## loss曲线图：



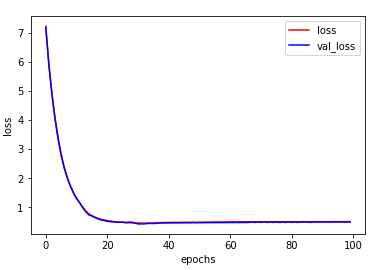
线性分类：

Learning\_rate = 0.1

Epoch = 1000

参数w,b 选择正态分布初始化

## loss曲线图：



## 12.实验结果分析:

线性分类/回归问题结果总是收敛，但是参数设置不同导致收敛结果不一样，如果学习率设置过小，可能收敛到极小值但非整体最小值，如果学习率设置过大，可能无法收敛到最小值而在周围摆动。在训练数据时要不断的调参，选择合适的loss\_function，以及不同验证方法

## 对比线性回归和线性分类的异同点：

线性回归问题通过线性方程，不断带入训练集数据，通过梯度下降的方式更新参数w，b，使其预测结果y\_pred逼近真实结果y。

线性分类问题通过训练训练集数据，然后利用梯度下降方式更新参数w，b，但是其预测结果y需要带入sigmoid函数（或者设立阈值）求得分类结果，本次实验为二分类，设立的阈值为0，将结果分为-1/1，从而求loss

## 14.实验总结：

通过本次实验，深入理解了线性分类，线性回归，梯度下降，以及支持向量机的原理，自己能够通过简单的算法实现其功能。理解了学习率，epoch对实验结果的影响，而w，b对实验结果影响并非很大，其能通过梯度下降不断更新进而逼近真实值。掌握了回归与分类问题的异同，为其他算法的学习提供了参考。