

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**  **黄启琛**

**学 号 201530611777**

**邮 箱** [**1049385379@qq.com**](mailto:1049385379@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1日

## 3. 报告人:黄启琛

## 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 中的Housing 数据，包含506个样本，每个样本有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，每个样本有14个属性

## 6. 实验步骤:

### 线性回归和梯度下降

读取实验数据，使用sklearn库的load\_svmlight\_file函数读取数据。

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。

线性模型参数初始化，使用全零初始化。

选择OLS损失函数及对其求导。

求得所有样本对Loss函数的梯度的负方向G。

更新模型参数w = w + learning\_rate \* G，learning\_rate是人为调整的超参数

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。

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

读取实验数据，使用sklearn库的load\_svmlight\_file函数读取数据。

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。

支持向量机模型参数初始化，使用全零初始化。

选择Hinge Loss函数及对其求导。

求得所有样本对Loss函数的梯度。

取梯度的负方向

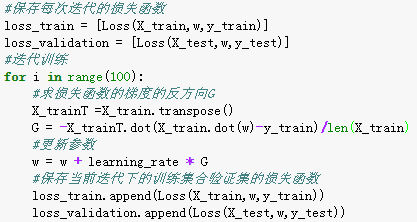
更新模型参数w -= learning\_rate \* gradient。learning\_rate为学习率，是人为调整的超参数。

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。

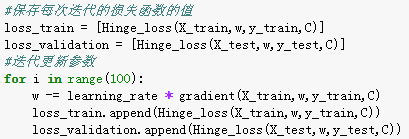
重复步骤5-8若干次，画出和随迭代次数的变化图。

## 7. 代码内容:

线性回归和梯度下降



线性分类和梯度下降



## 8. 选择的评估方法（留出法，交叉验证，k折交叉验证等）:

线性回归和梯度下降

留出法

线性分类和梯度下降

留出法

## 9. 模型参数的初始化方法:

线性回归和梯度下降

全零初始化

线性分类和梯度下降

全零初始化

## 10.选择的loss函数及其导数:

线性回归和梯度下降

loss =

gradient =

\*其中n为样本数

线性分类和梯度下降

loss =

gradient =

\*其中n为样本数，表示第i个样本的label，表示第i个样本的特征

## 11.实验结果和曲线图:

线性回归和梯度下降

超参数选择：

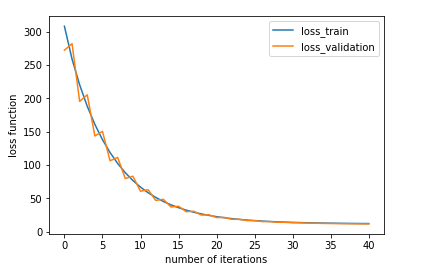
学习率learning\_rate为0.4

评估结果：

|  |  |
| --- | --- |
| learning\_rate | loss |
| 0.01 | 12.3474 |
| 0.02 | 11.2816 |
| 0.05 | 10.7088 |
| 0.1 | 10.4771 |
| 0.2 | 10.3914 |
| 0.25 | 10.3800 |
| 0.3 | 10.3735 |
| 0.4 | 10.3670 |
| 0.5 | 无法收敛 |

## 预测结果（最佳结果）： loss = 10.3670

## loss曲线图：



线性分类和梯度下降

超参数选择：

learning\_rate = 0.01

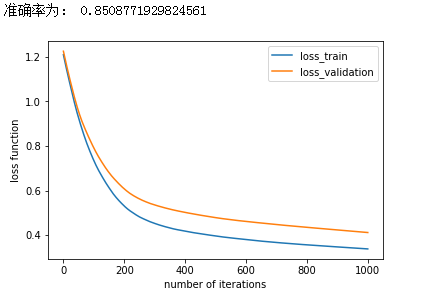
C = 0.9

评估结果：

|  |  |  |
| --- | --- | --- |
| Learning\_rate | C | 准确率 |
| 0.01 | 0.9 | 0.8509 |
| 0.05 | 0.9 | 0.8333 |
| 0.1 | 0.9 | 0.8333 |
| 0.01 | 0.8 | 0.8465 |
| 0.01 | 0.7 | 0.8509 |
| 0.01 | 0.6 | 0.8421 |
| 0.01 | 0.5 | 0.8289 |

预测结果（最佳结果）：准确率 = 0.8509

## loss曲线图：



## 12.实验结果分析:

因为这里线性回归和线性分类都是使用所有样本求得的梯度，对训练集数据而言，loss函数值一直下降；线性回归的验证集的loss函数出现轻微的震荡，但总体是下降的趋势，线性分类的验证集的loss函数一直下降，loss值大于训练集的loss值，属于正常现象，最终分类准确率达到85%

## 13.对比线性回归和线性分类的异同点：

线性回归和线性分类都是对样本特征进行简单的线性计算，从而达到预测的目的，不同点在于，线性回归的预测结果是一个实数值，线性分类的预测结果是分类结果；线性回归超参数只有learning\_rate，调参时直接使用loss函数值判断最佳的learning\_rate，线性分类的超参数除了learning\_rate以外，还有hinge\_loss的权重C，调参时，因为loss函数是C的线性函数，不能直接用loss函数对C调参，需要使用分类的准确率。

## 14.实验总结：

本次实验让我进一步理解了线性回归、线性分类和梯度下降的原理，熟悉了损失函数的计算和求导，清楚了参数的迭代更新过程，掌握了超参数的调整过程；与此同时，基于Python的各种工具包的使用有了更深的了解和认识，能够在得到新的数据的时候利用工具包进行相关的处理，