

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**  **李灿光**

**学 号 201530611937**

**邮 箱 1065561759@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: 李灿光

## 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:

1）、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取梯度的负方向G，记为D。

⑦更新模型参数，。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。

2）、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取梯度的负方向G，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:

1）、线性回归和梯度下降

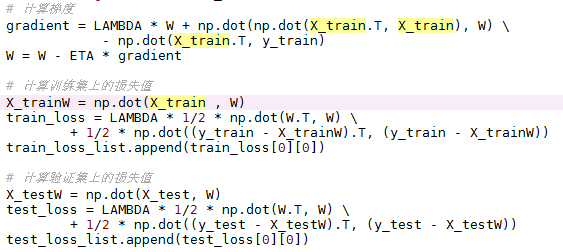


图1 线性回归梯度下降和loss计算的代码

2）、线性分类和梯度下降

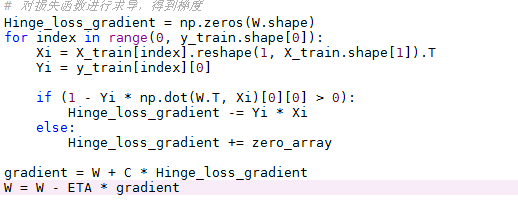


图2 线性分类梯度下降的代码

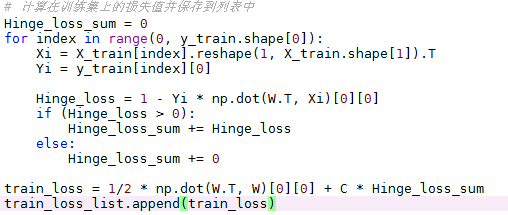


图3 线性分类loss计算的代码

## 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，从数据集中随机切分四分之一的样本作为验证集，剩下部分作为训练集。

## 9. 模型参数的初始化方法:

1）、线性回归和梯度下降

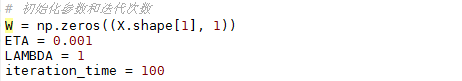


图4 线性回归初始化参数W、、和迭代次数

2）、线性分类和梯度下降

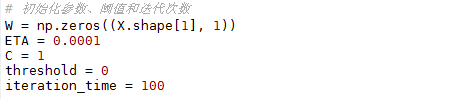


图5 线性分类初始化参数W、、C、阈值和迭代次数

## 10.选择的loss函数及其导数:

1）、线性回归和梯度下降

loss函数：

导数：

2）、线性分类和梯度下降

loss函数：

导数：记

则loss函数的导数为

## 11.实验结果和曲线图:

## 超参数选择（η,epoch等）：

1）、线性回归和梯度下降

iteration\_time = 100

2）、线性分类和梯度下降

C = 1

iteration\_time = 100

threshold = 0

## 评估结果（根据选择的评估方法）：

1）、线性回归和梯度下降

训练集上loss最小为4012.0963673122073

验证集上loss最小为2058.3742636076668

2）、线性分类和梯度下降

训练集上loss最小为160.0298752798023

验证集上loss最小为45.982449217649219

## 预测结果（最佳结果）：

1）、线性回归和梯度下降

验证集上loss最小为2058.3742636076668

2）、线性分类和梯度下降

验证集上正确率最高为0.9017341040462428

## loss曲线图：

1）、线性回归和梯度下降

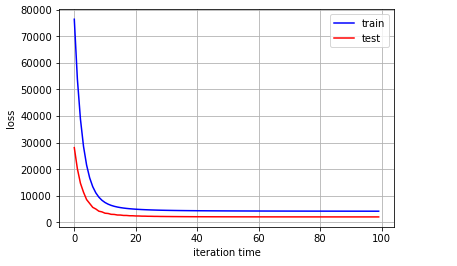


图6 线性回归loss值随迭代次数增加的图像

2）、线性分类和梯度下降

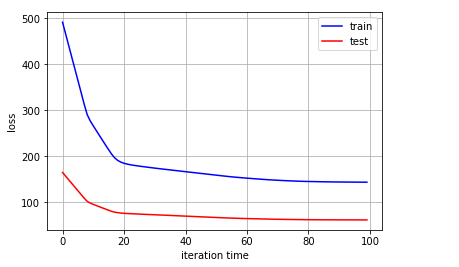


图7 线性分类loss值随迭代次数增加的图像

## 12.实验结果分析:

1）、线性回归和梯度下降

我实验时学习率分别取了0.1、0.01、0.001、0.0001，发现取0.1时偏大，loss值不会收敛，loss值图像会呈上升趋势，取0.01时也是同样的情况，取0.001时loss值图像开始呈下降趋势，会收敛，取0.0001时收敛速度减小。具体情况如下图：

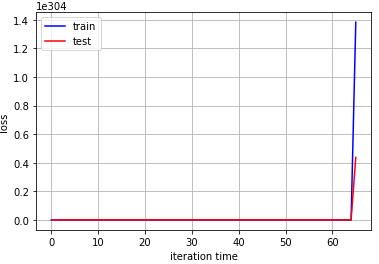


图8 线性回归

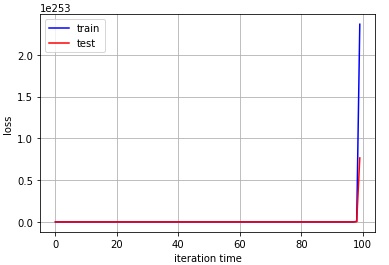


图9 线性回归

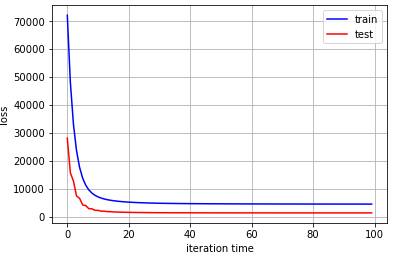


图10 线性回归

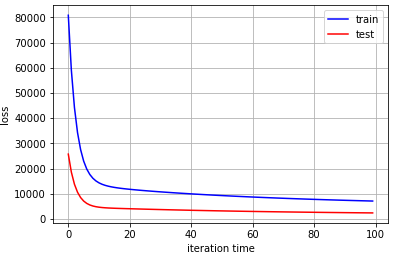


图11 线性回归

2）、线性分类和梯度下降

实验时学习率分别取了0.1、0.01、0.001、0.0001、0.00001，发现取0.1时loss值图像振荡幅度很大，取0.01时loss值图像振荡稍微减小，取0.001时振荡再减小，取0.0001时发现效果较好，取0.00001时由于学习率较小，所以收敛较慢，具体情况如下图：

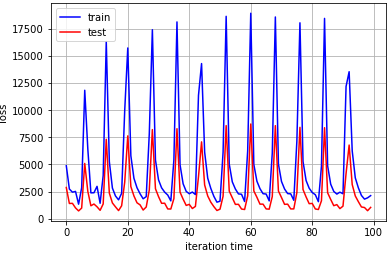


图12 线性回归

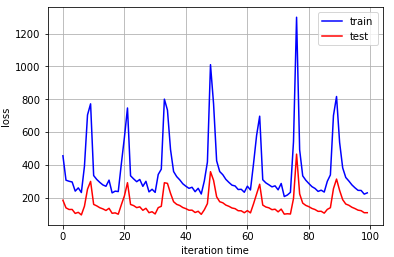


图13

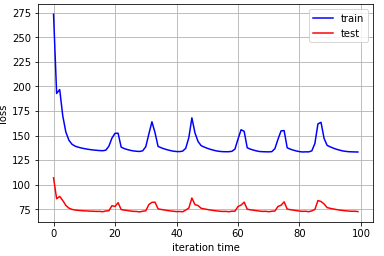


图14 线性回归

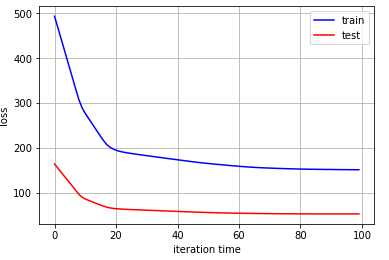


图15 线性回归

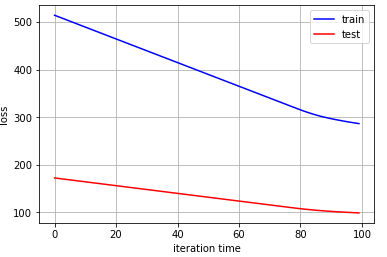


图16 线性回归

## 13.对比线性回归和线性分类的异同点：

相同点：线性回归和线性分类所用的线性函数是相同的。

不同点：预测的方式不一样——线性回归是将样本代入线性函数进行计算，所得结果作为预测值；线性分类需要设置一个阈值，将样本代入线性函数计算后，把结果与阈值进行比较，根据比较结果将样本归为正类或负类。

## 14.实验总结：

通过自己实现了线性回归与线性分类算法，我更加理解了矩阵求导的方法和梯度下降的原理。同时通过实践过程中的调参，我也获得了一些调参的经验。除此之外，我也理解了线性分类和线性回归的异同点。