

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**  **林艾琳**

**学 号 201530612200**

**邮 箱 743852586@qq.com**

**指导教师**  **吴庆耀**

**提交日期** **2017年12月8日**

## 1. 实验题目: 逻辑回归、线性分类与随机梯度下降

## 2. 实验时间：2017年12月2日

## 3. 报告人：林艾琳

## 4. 实验目的：

①对比理解梯度下降和随机梯度下降的区别与联系。

②对比理解逻辑回归和线性分类的区别与联系。

③进一步理解SVM的原理并在较大数据上实践。

## 数据集以及数据分析：

## **数据集**

## 实验使用的是[LIBSVM Data](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/" \t "https://www.zybuluo.com/chenyaofo/note/_blank)的中的[a9a](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html" \l "a9a" \t "https://www.zybuluo.com/chenyaofo/note/_blank)数据，包含32561 / 16281(testing)个样本，每个样本有123/123 (testing)个属性。

## **数据分析**

## 使用了python包：sklearn，numpy，jupyter，matplotlib等工具，以及逻辑回归、线性分类和随机梯度下降等方法。

## 实验步骤：

## **逻辑回归与随机梯度下降**

## 读取实验训练集和验证集。

## 逻辑回归模型参数初始化，可以考虑全零初始化，随机初始化或者正态分布初始化。

## 选择Loss函数及对其求导，过程详见课件ppt。

## 求得部分样本对Loss函数的梯度。

## 使用不同的优化方法更新模型参数（NAG，RMSProp，AdaDelta和Adam）。

## 选择合适的阈值，将验证集中计算结果大于阈值的标记为正类，反之为负类。在验证集上测试并得到不同优化方法的Loss函数值，，和。

## 重复步骤4-6若干次，画出，，和随迭代次数的变化图。

## **线性分类与随机梯度下降**

## 读取实验训练集和验证集。

## 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。

## 选择Loss函数及对其求导，过程详见课件ppt。

## 求得部分样本对Loss函数的梯度。

## 使用不同的优化方法更新模型参数（NAG，RMSProp，AdaDelta和Adam）。

## 选择合适的阈值，将验证集中计算结果大于阈值的标记为正类，反之为负类。在验证集上测试并得到不同优化方法的Loss函数值，，和。

## 重复步骤4-6若干次，画出，，和随迭代次数的变化图。

## 7. 代码内容：

（针对逻辑回归和线性分类分别填写8-11内容）

## 8. 模型参数的初始化方法:

## 9.选择的loss函数及其导数:

## 10.实验结果和曲线图:（各种梯度下降方式分别填写此项）

## 超参数选择：

## 预测结果（最佳结果）：

## loss曲线图：

## 11.实验结果分析:

## 12.对比逻辑回归和线性分类的异同点：

## 13.实验总结：