

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**  **吴国斌**

**学 号 201530613030**

**邮 箱 854370642@qq.com**

**指导教师**  **吴庆耀**

**提交日期** **2017年12月2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:吴国斌

## 4. 实验目的:

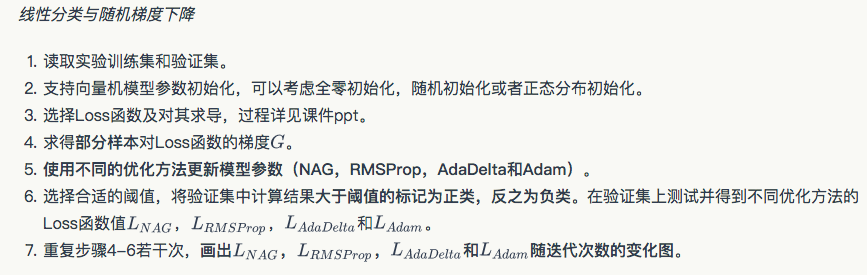
1. 对比理解梯度下降和随机梯度下降的区别与联系。
2. 对比理解逻辑回归和线性分类的区别与联系。
3. 进一步理解SVM的原理并在较大数据上实践。

## 5. 数据集以及数据分析：

实验使用的是[LIBSVM Data](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/" \t "_blank)的中的[a9a](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html" \l "a9a" \t "_blank)数据，包含32561 / 16281(testing)个样本，每个样本有123/123 (testing)个属性。请自行下载训练集和验证集。其中label表示该记录的类别，+1为正类，-1为负类。其中正类有11687个样本，负类有37155个样本。

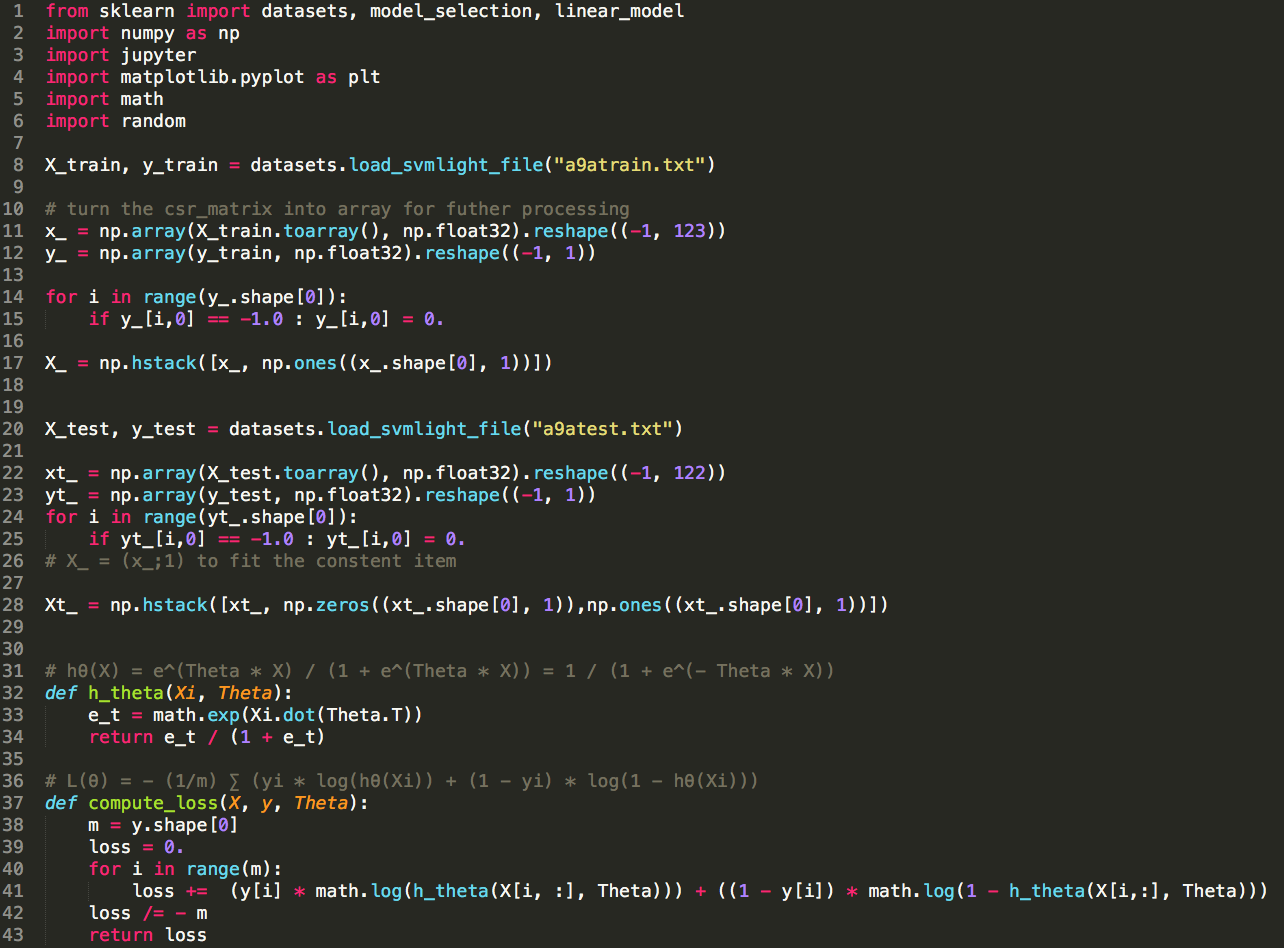
## 6. 实验步骤:





## 7. 代码内容:

逻辑回归：

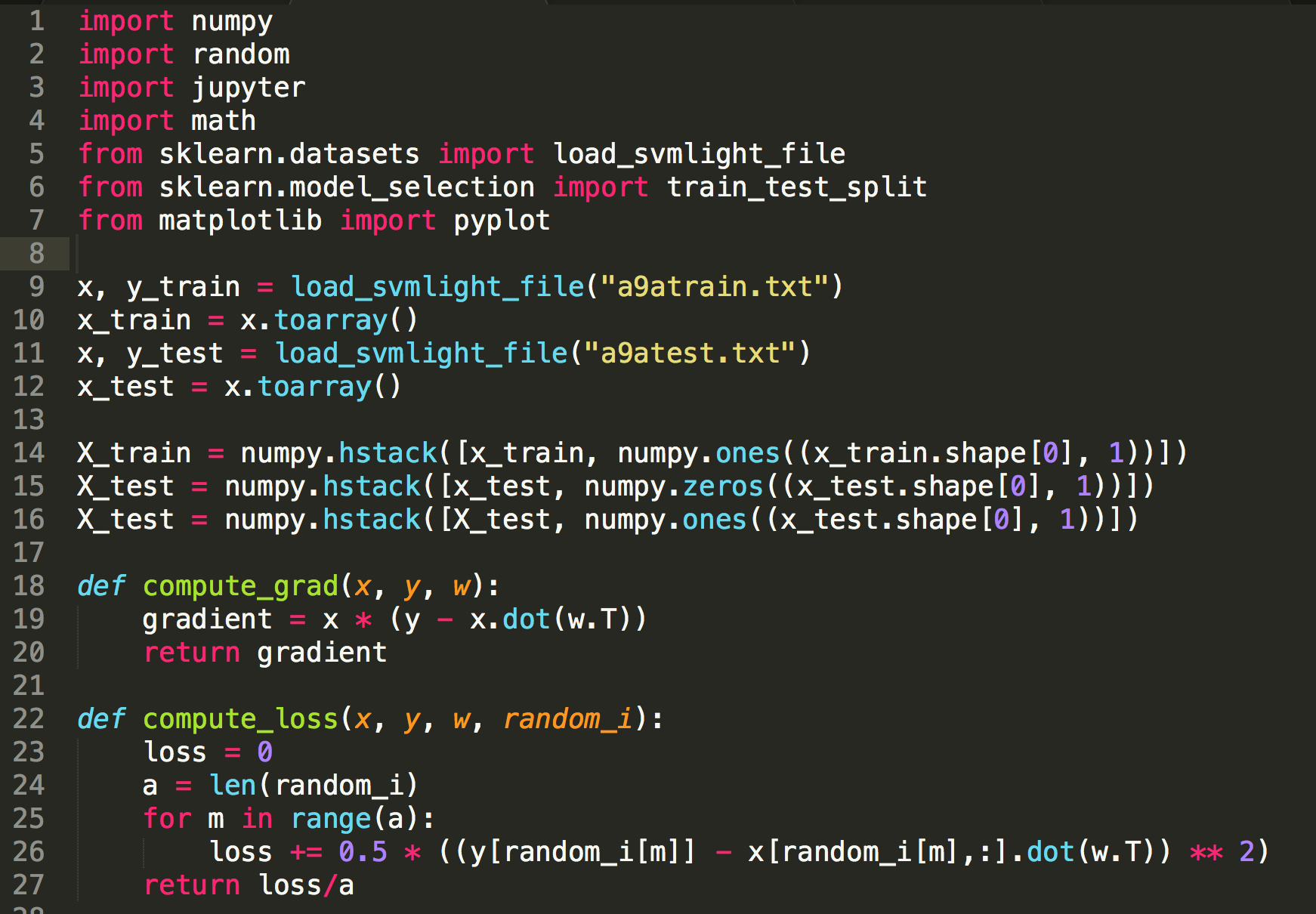


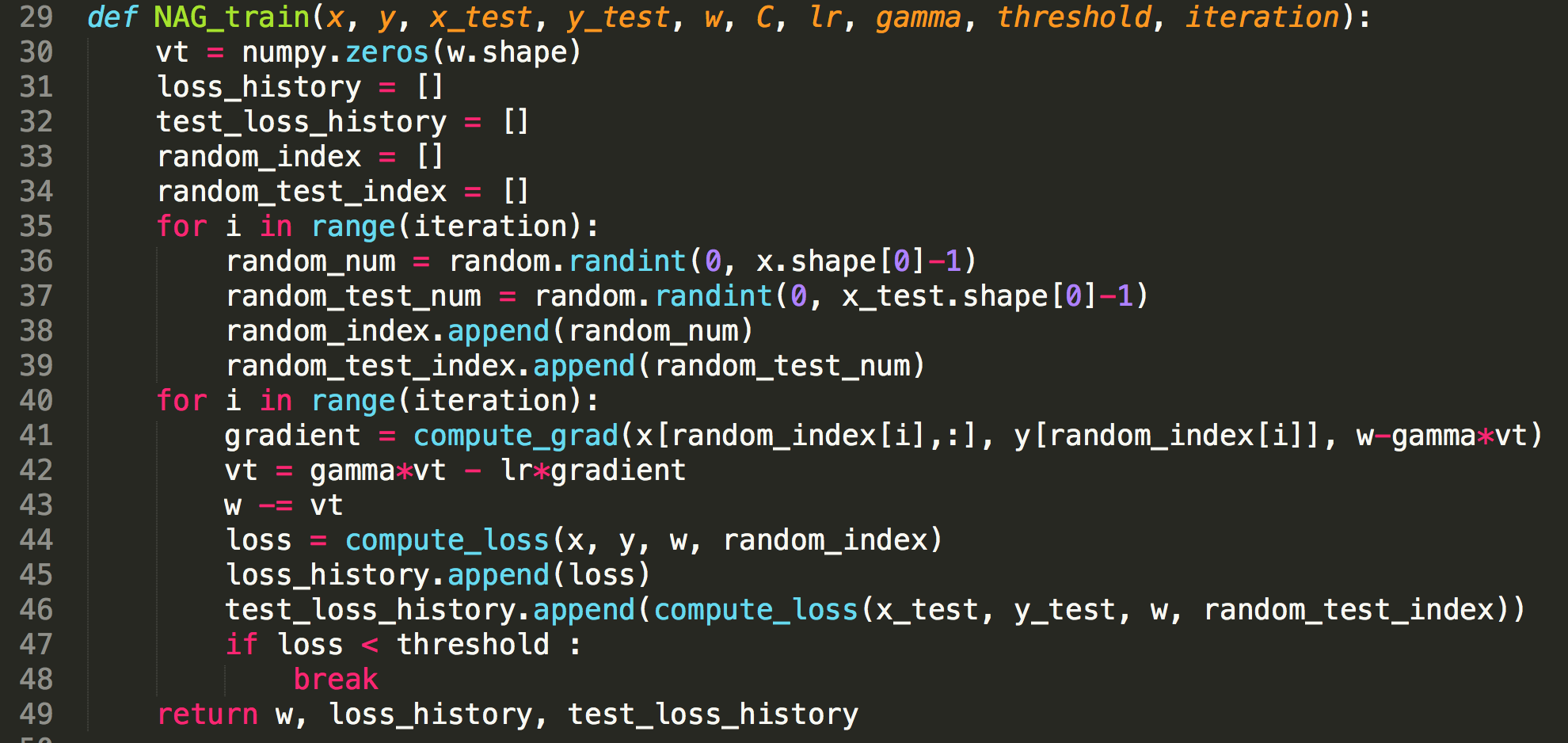


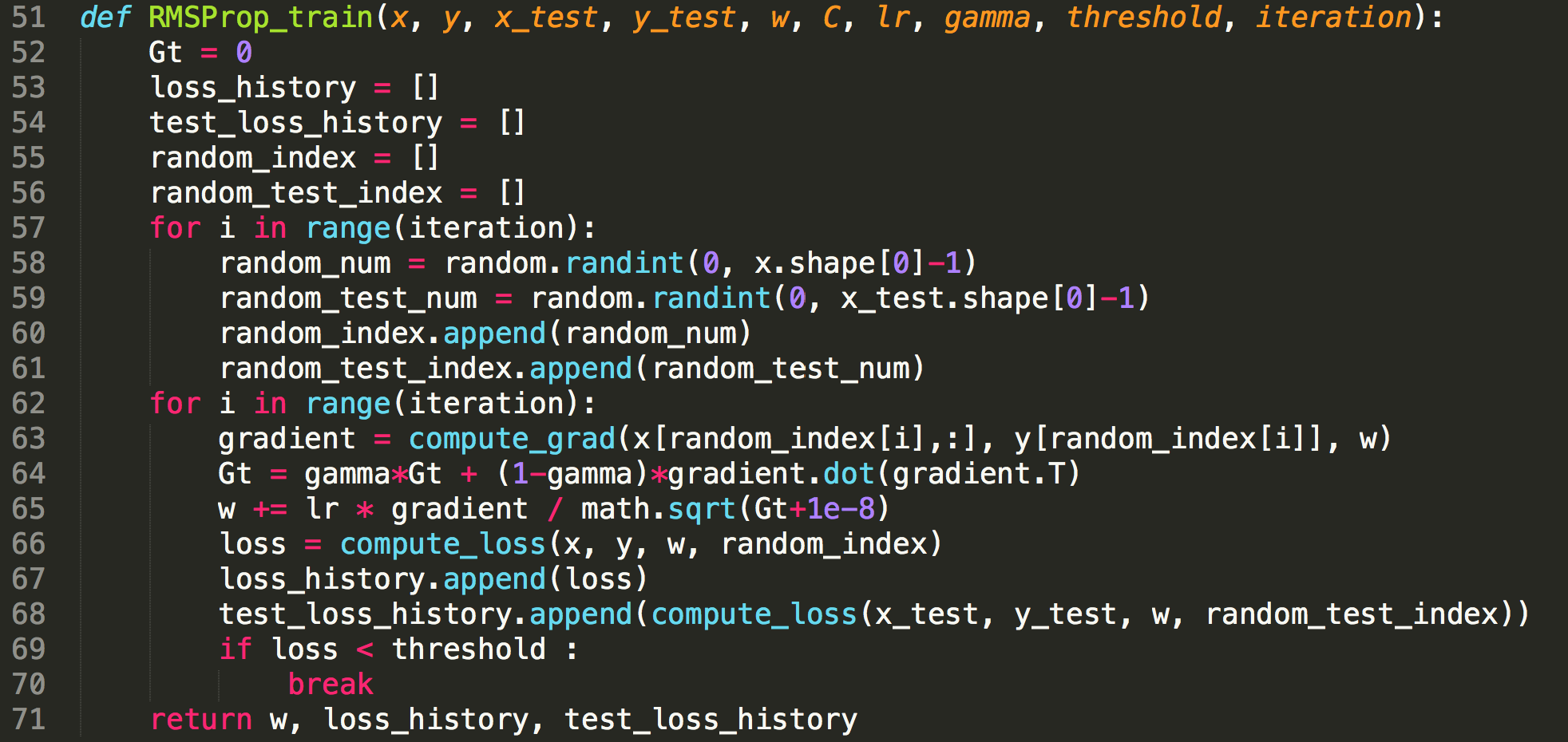


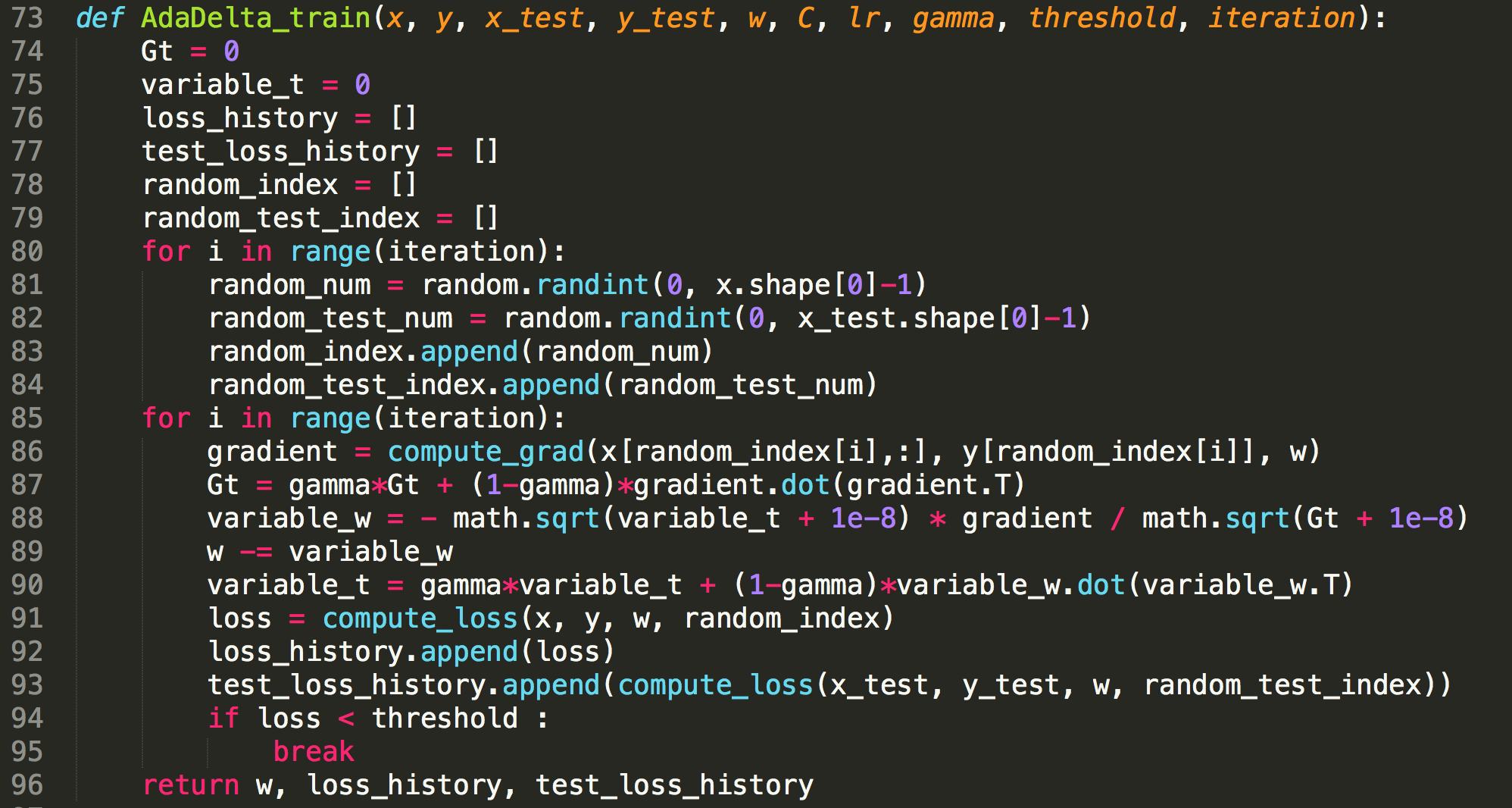


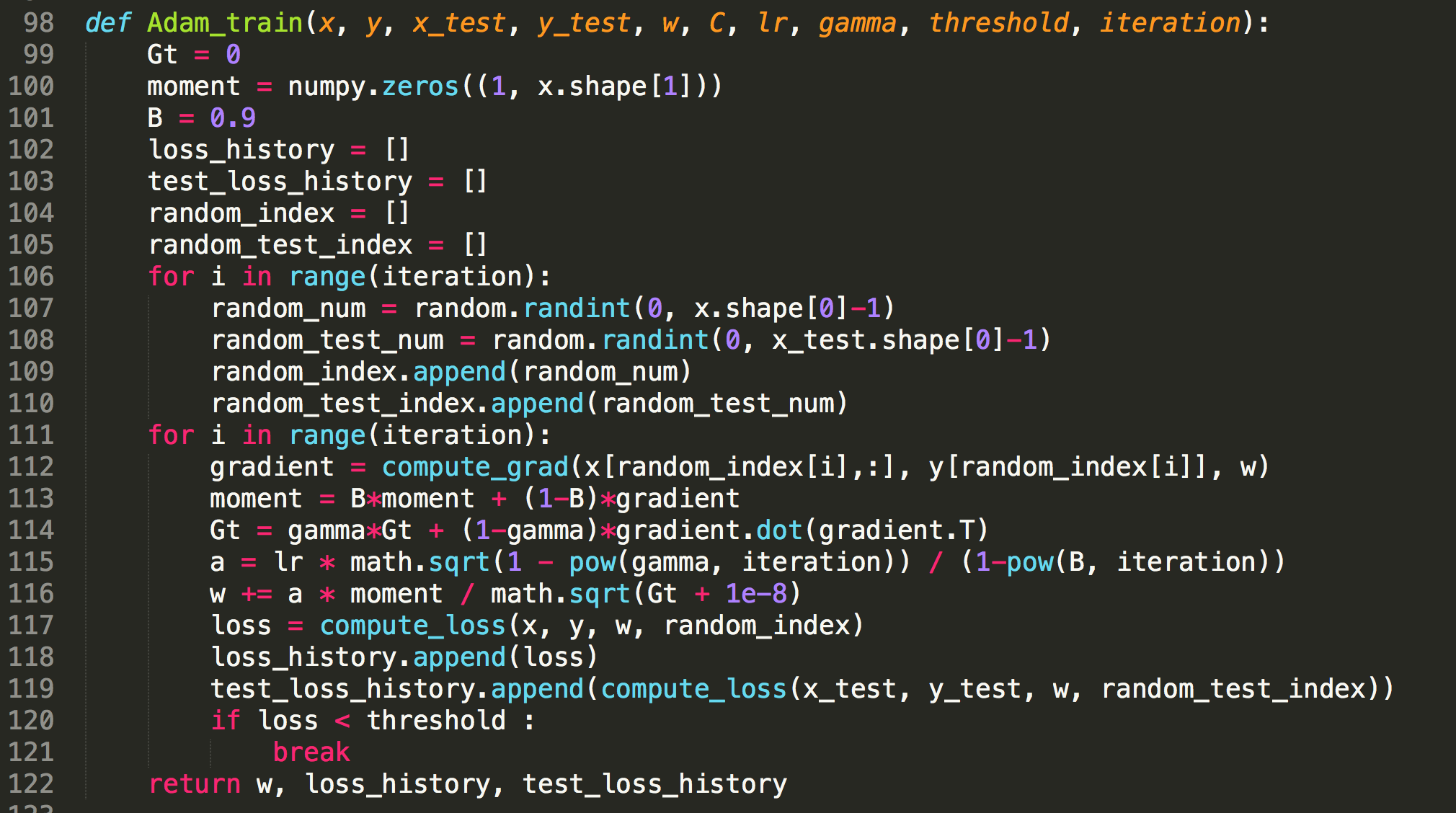
线性分类：

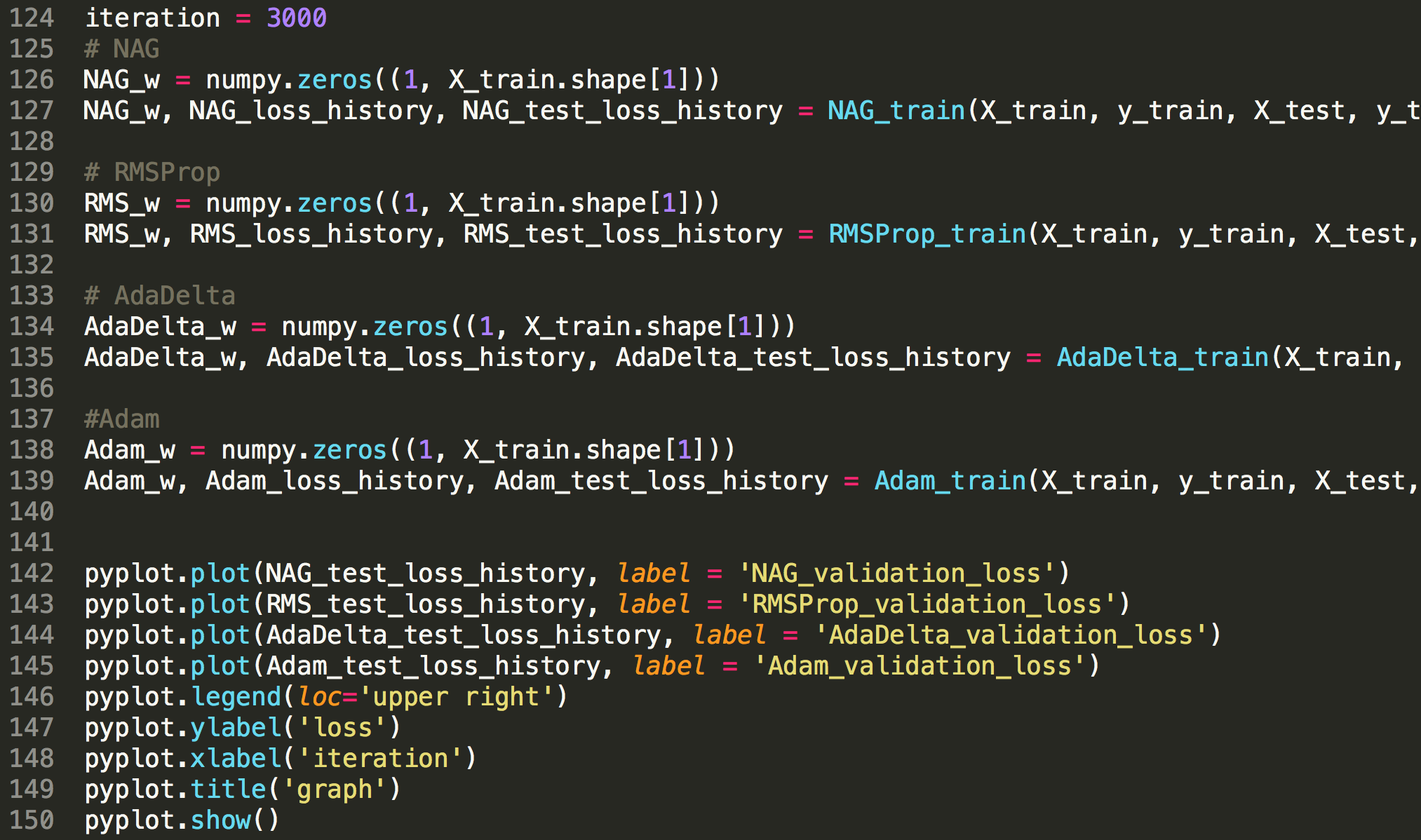












**逻辑回归：**

## 8. 模型参数的初始化方法:

模型参数的初始化方法采用的是全零初始化。

## 9.选择的loss函数及其导数:

loss function:

其中，

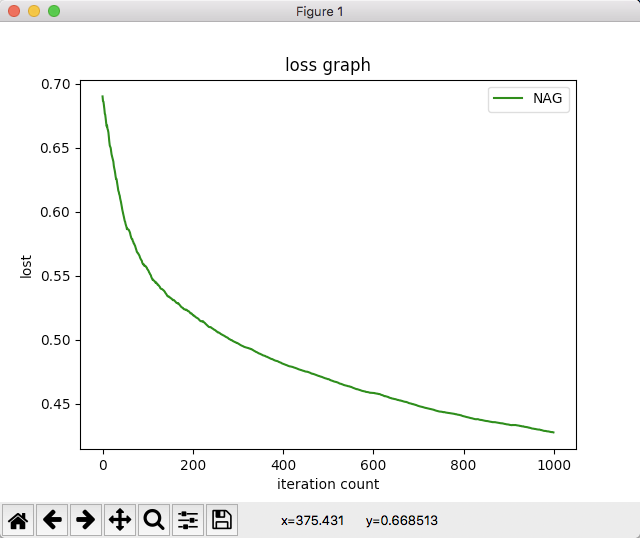
gradient:

## 10.实验结果和曲线图:

## NAG:

## 超参数选择：η=0.005 =0.9 epoch = 1000

## 预测结果（最佳结果）：



RMSProp:

## 超参数选择：η=0.005 = 0.9 = 1e-8 epoch = 1000

## 预测结果（最佳结果）：



AdaDelta:

## 超参数选择： = 0.9 = 1e-8 epoch=1000

## 预测结果（最佳结果）：

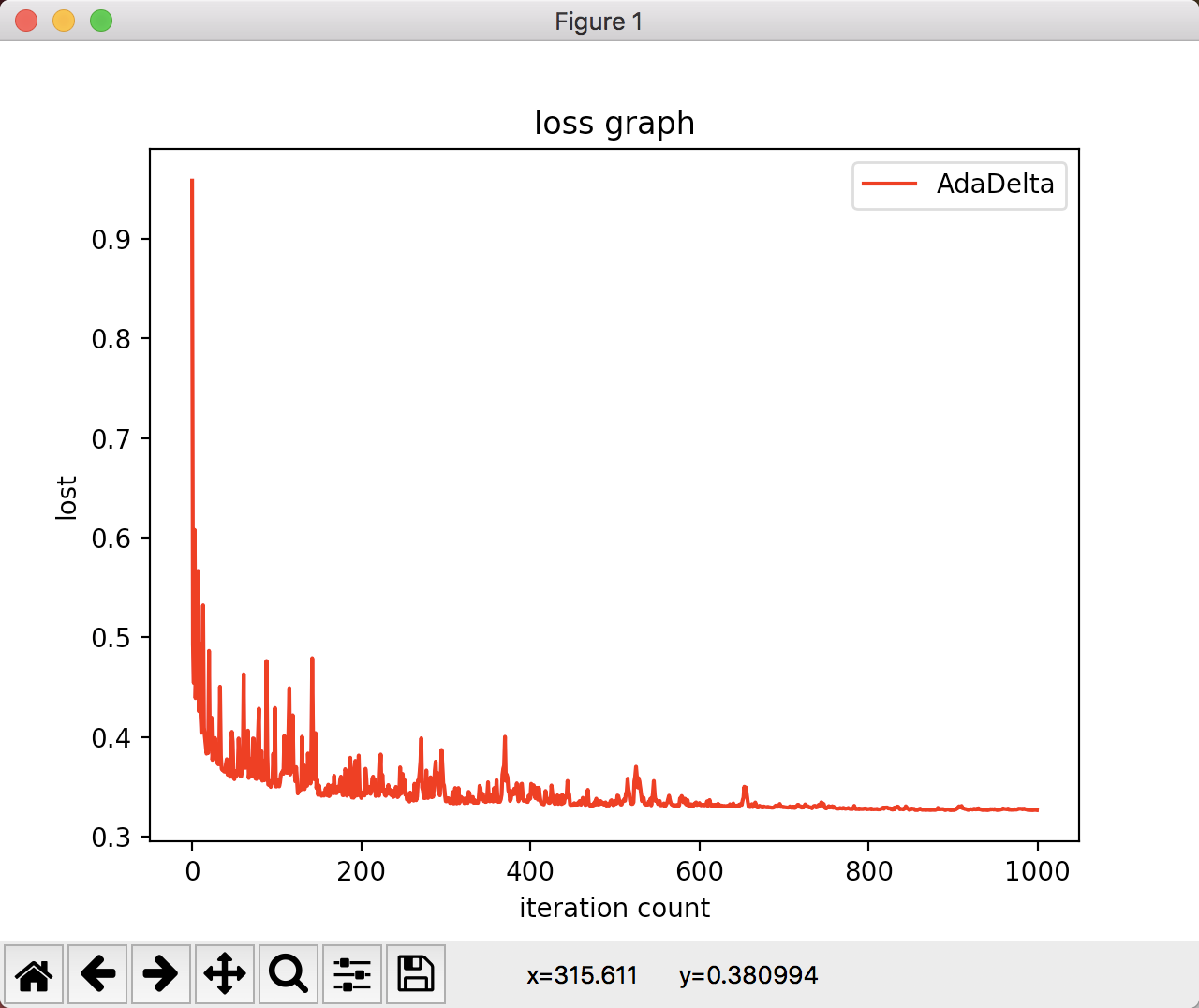
## loss曲线图：

Adam:

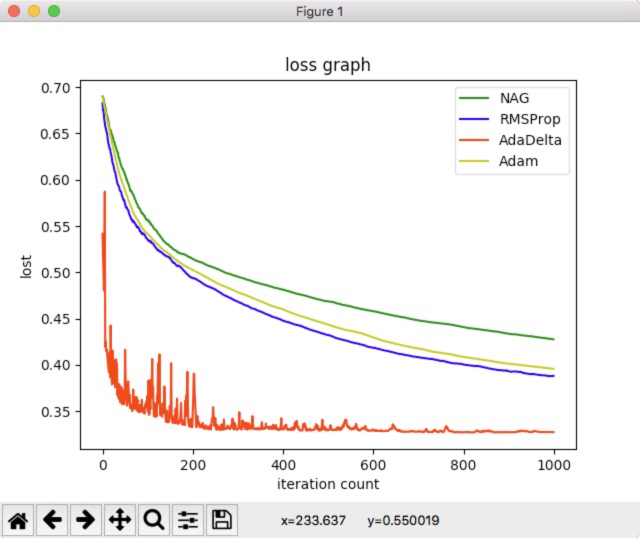
## 超参数选择：η=0.005 = 0.9 = 0.999

## = 1e-8 epoch=1000

## 预测结果（最佳结果）：



## loss曲线图：



## 11.实验结果分析:

AdaDelta较快的达到了局部最优解，但是震荡明显；NAG在收敛速度和收敛幅度上都较小，但是曲线较为平滑；RMSProp和Adam基本重合，收敛速度比NAG快，但是比AdaDelta慢。

总体而言，四种方法都可以较为迅速的收敛到局部最优解，比基本的梯度下降法更加优秀。

**线性分类：**

## 8. 模型参数的初始化方法:

模型参数的初始化方法采用的是全零初始化。

## 9.选择的loss函数及其导数:

loss function:

其中，

gradient:

## 10.实验结果和曲线图:

## NAG:

## 超参数选择：η=0.001 =0.9 epoch = 3000

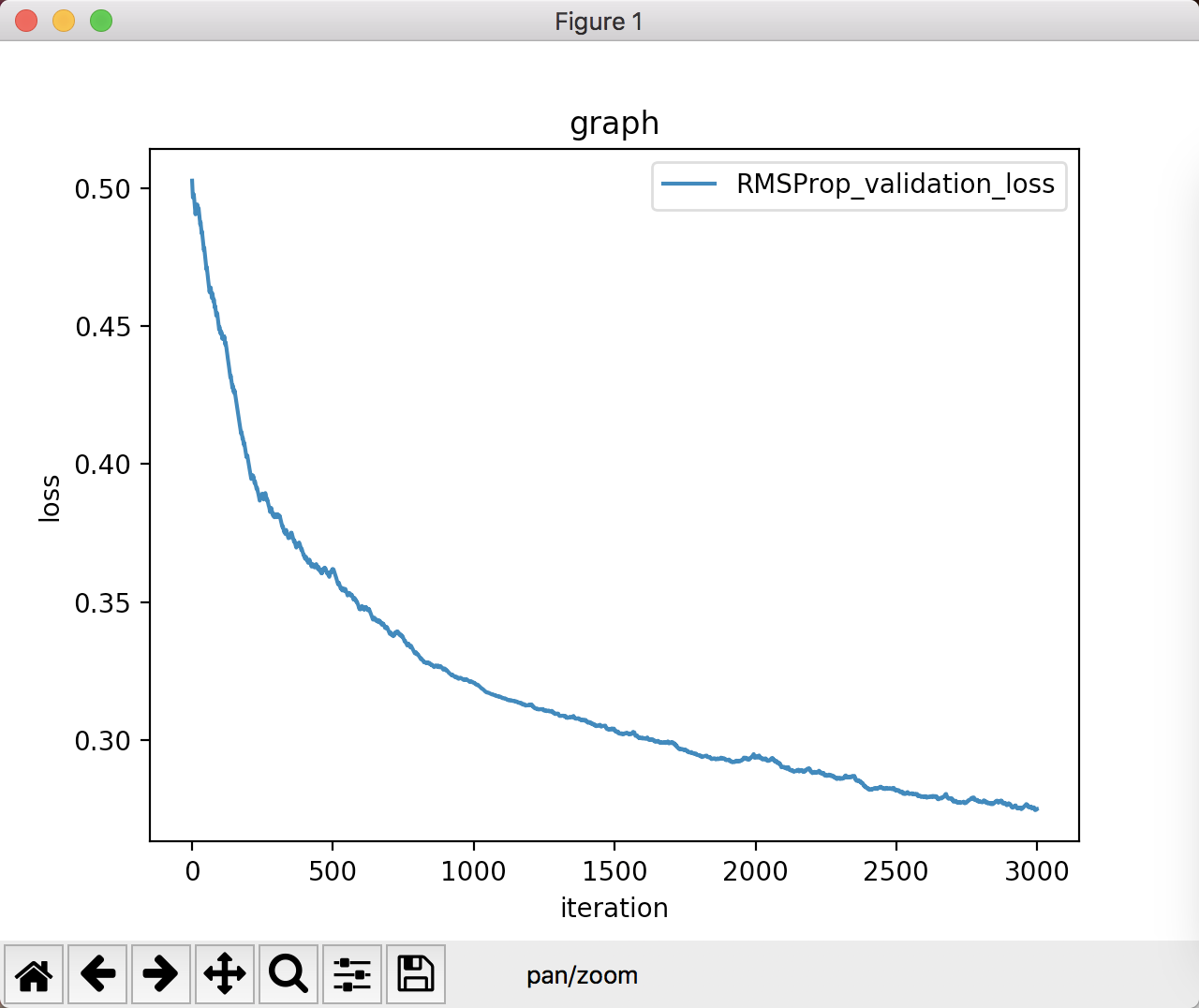
## 预测结果（最佳结果）：



RMSProp:

## 超参数选择：η=0.001 = 0.9 = 1e-8 epoch = 3000

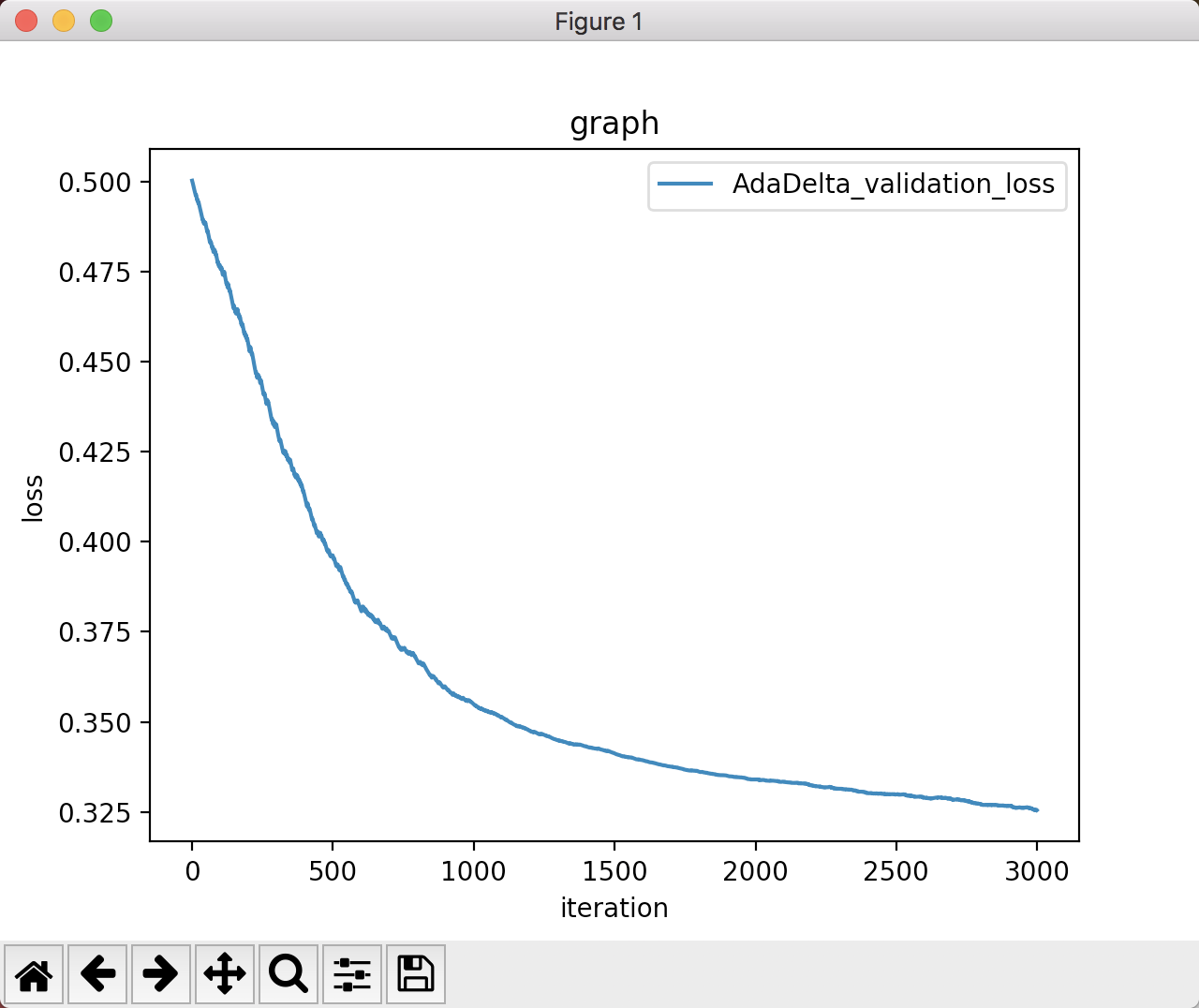
## 预测结果（最佳结果）：



AdaDelta:

## 超参数选择： = 0.9 = 1e-8 epoch=3000

## 预测结果（最佳结果）：

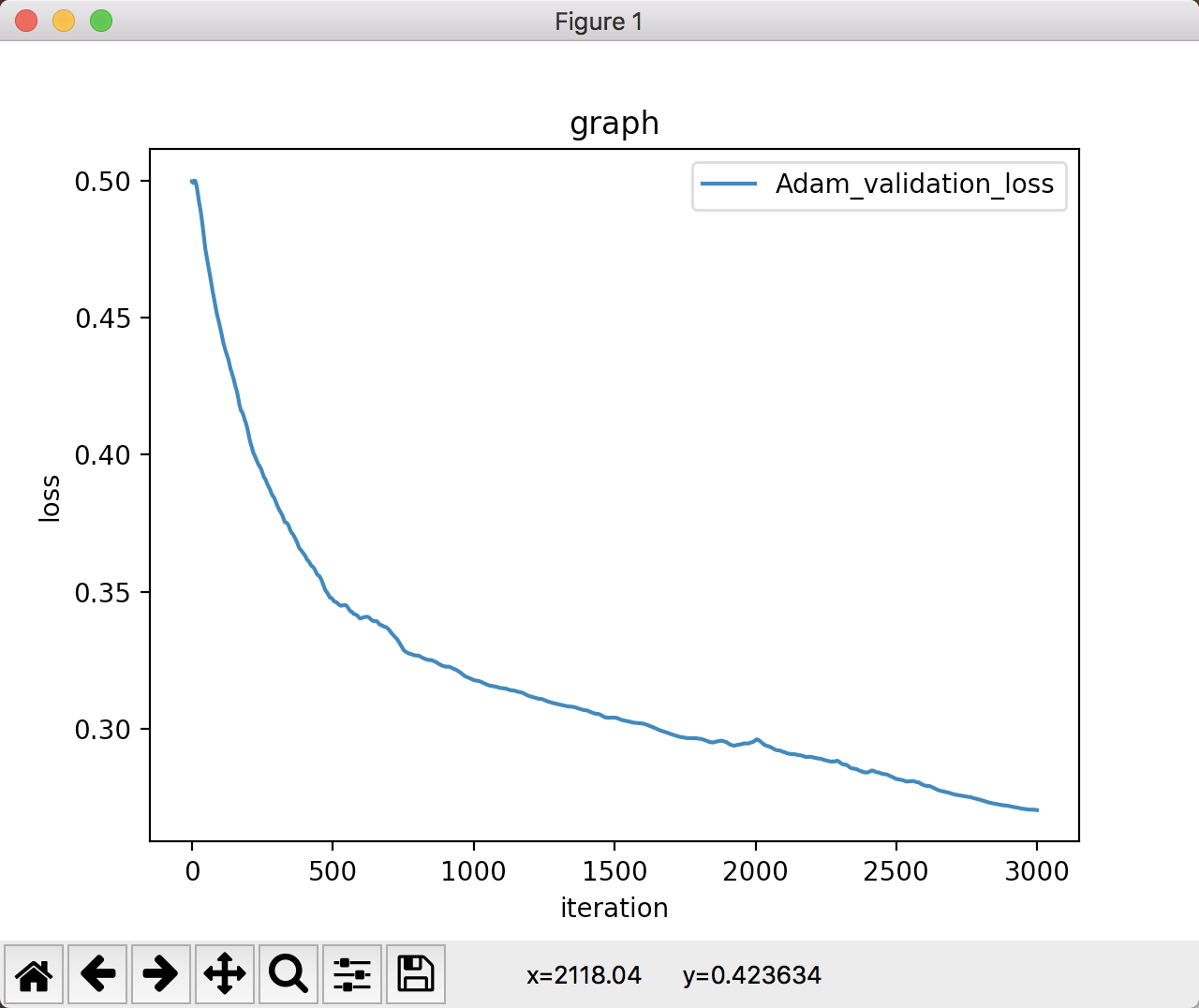


Adam:

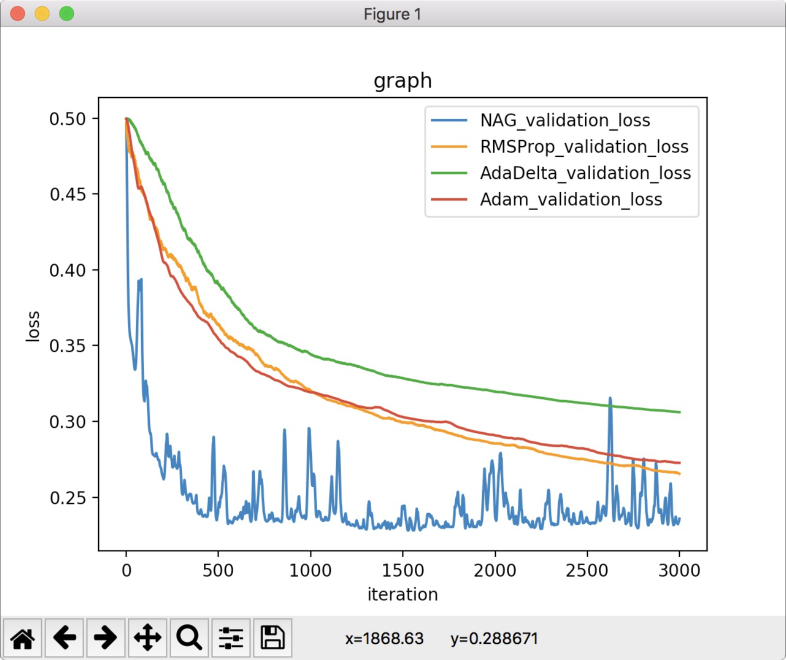
## 超参数选择：η=0.005 = 0.9 = 0.999

## = 1e-8 epoch=3000

## 预测结果（最佳结果）：



loss曲线：



## 11.实验结果分析:

NAG收敛速度最快，但是震荡明显；；RMSProp 与 Adam的收敛速度几乎持平，震荡不明显；AdaDelta收敛较慢但是曲线比较平滑。相比基本的随机梯度下降算法，这四种算法在不同程度上做了改进，更加优秀。

## 12.对比逻辑回归和线性分类的异同点：

**同：**

都属于分类问题，都用于预测。

**异：**

找最优超平面的方法不同，，形象点说，logistic模型找的那个超平面，是尽量让所有点都远离它，而SVM线性分类寻找的那个超平面，是只让最靠近中间分割线的那些点尽量远离，即只用到那些“支持向量”的样本。

逻辑回归只可以处理线性可分情况；SVM则二者皆可。

## 13.实验总结：

本次实验当中，我学习到了很多SGD在实践运用的经验，将课程中学习到的知识运用在实际问题上。但是由于自身对于知识把握懂得程度不高，在实现的过程中遇到了诸如无法正确实现SGD优化算法，调参不够灵活的问题，在总结反思之后解决了问题并顺利完成了实验。、

在对模型的训练过程中，我体会到了灵活调参的重要性。在一开始，因为超参数C, learning rate等设置得不合理，导致loss图像与预期相差甚远，模型参数无法收敛或者收敛过慢，跑数据集时间过长等问题，导致无法拟合数据；或者是因为迭代次数过少，参数还未收敛便停止了训练。在进行数次的不同的调参后，模型往预期方向改变，我也从中学得一些经验。