

**《机器学习》课程实验报告**

**学 院 软件学院**

**专 业 软件工程**

**组 员**  胡永浩

**学 号 201530611630**

**邮 箱 1458732467@qq.com**

**指导教师**  **吴庆耀**

**提交日期** **2017年12 月15 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15 日

## 3. 报告人:胡永浩

## 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：

1. 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。请自行下载训练集和验证集

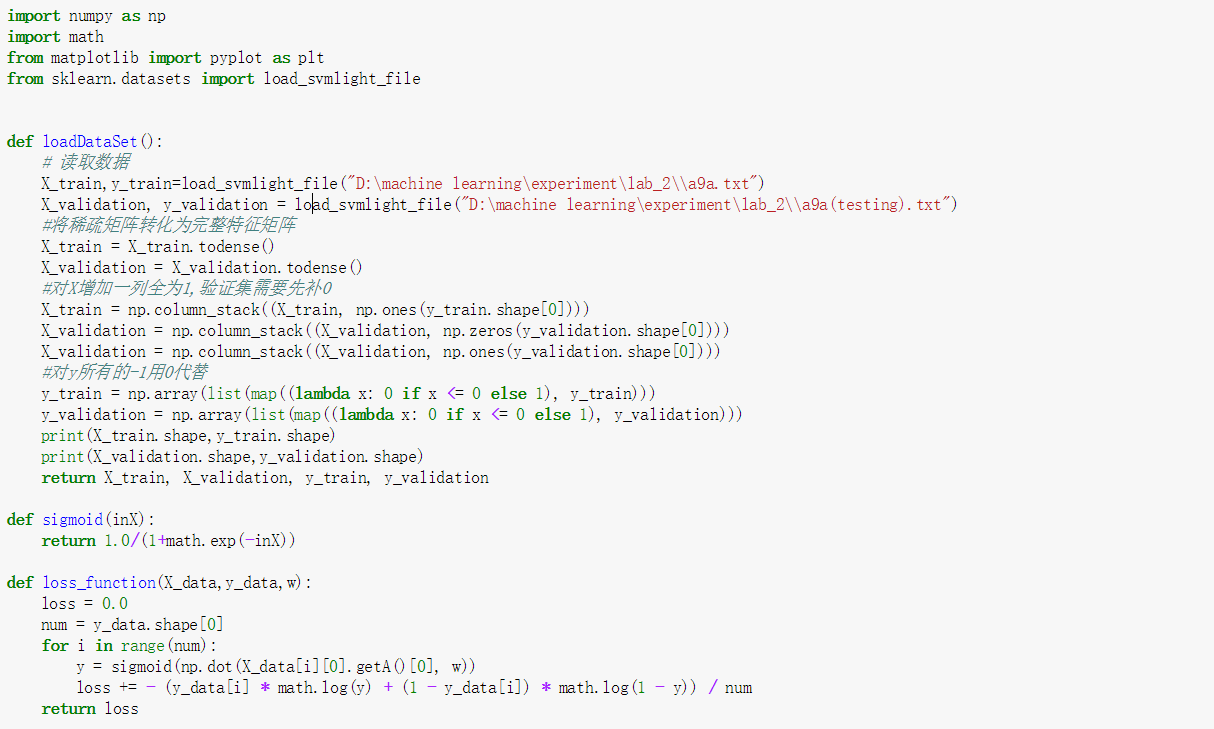
读取数据后要将稀疏矩阵转化为特征矩阵，并对X增加一列全为1，验证集需要先补零，对y所有的-1用0代替

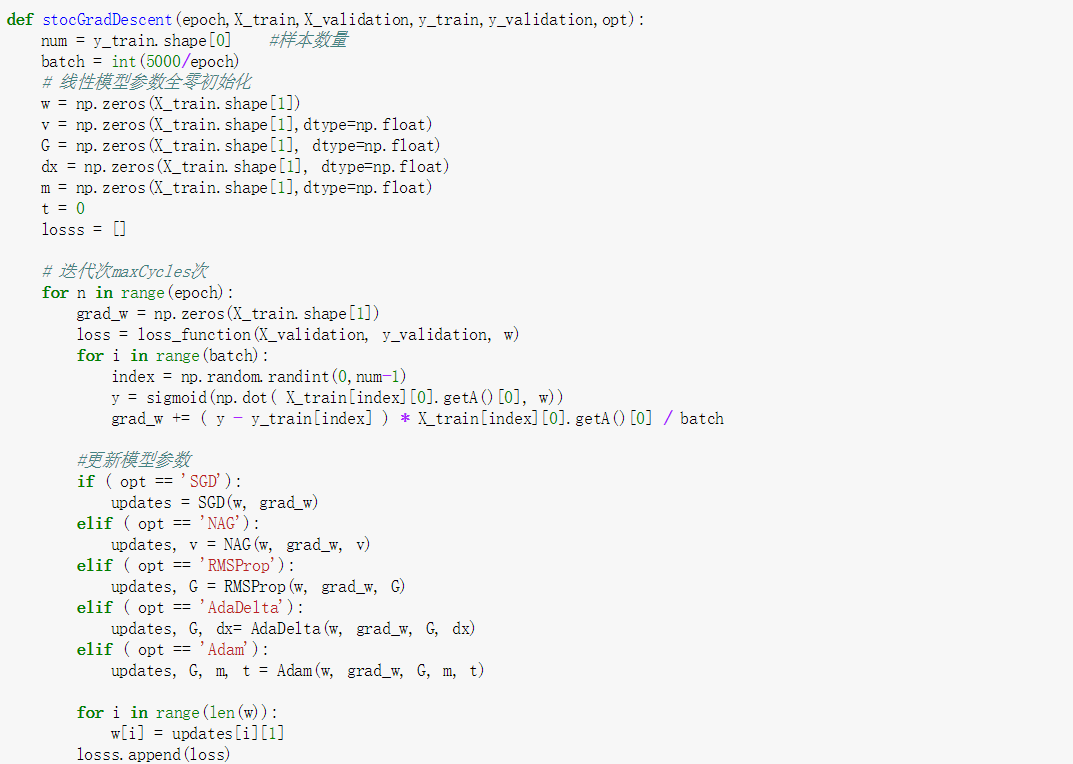
1）*逻辑回归与随机梯度下降*

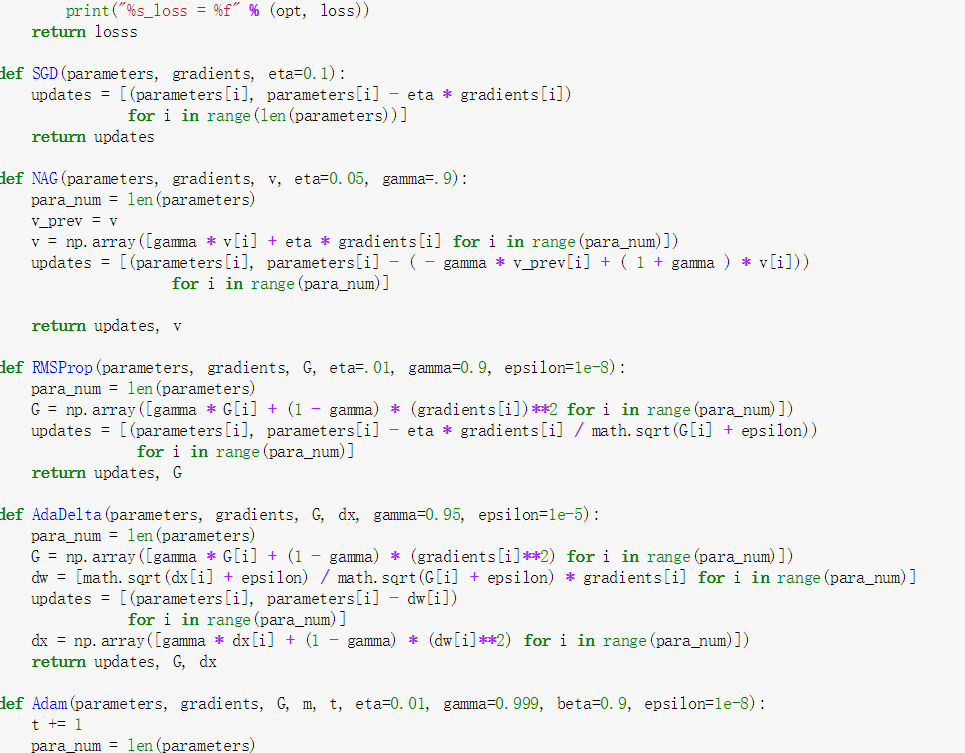
## 实验步骤:

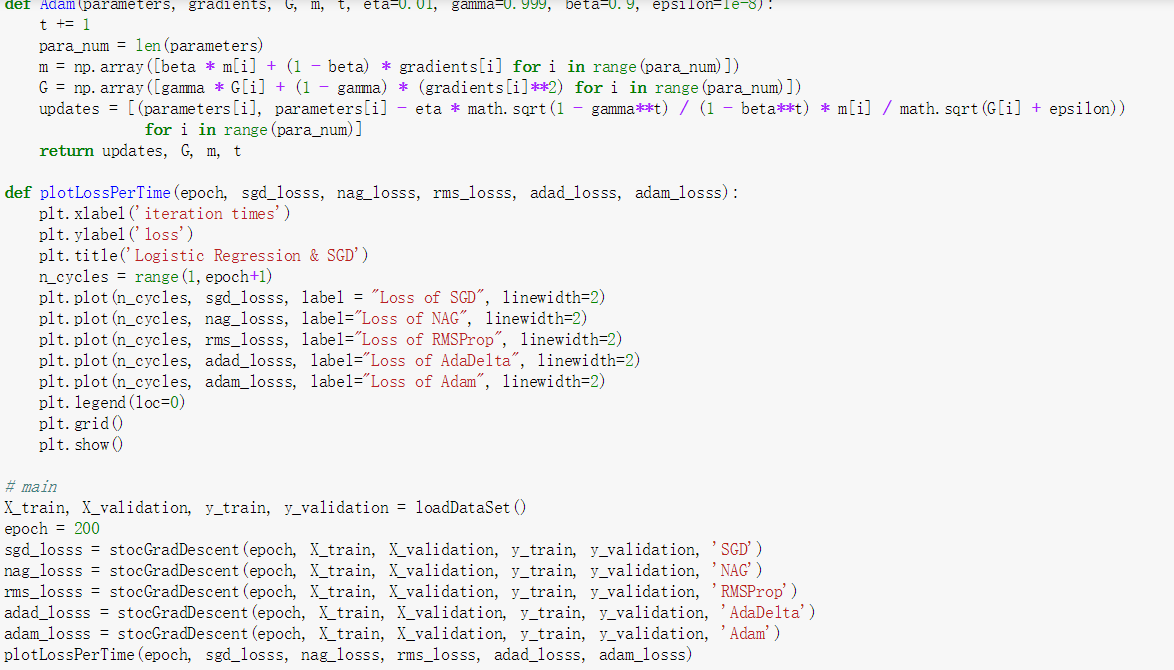
1. 读取实验训练集和验证集。
2. 逻辑回归模型参数初始化，可以考虑全零初始化，随机初始化或者正态分布初始化。
3. 选择Loss函数及对其求导，过程详见课件ppt。
4. 求得****部分样本****对Loss函数的梯度。
5. ****使用不同的优化方法更新模型参数（NAG，RMSProp，AdaDelta和Adam）****。
6. 选择合适的阈值，将验证集中计算结果****大于阈值的标记为正类，反之为负类****。在验证集上测试并得到不同优化方法的Loss函数值，，和。
7. 重复步骤4-6若干次，****画出，，和随迭代次数的变化图****。

## 代码内容:









## 8. 模型参数的初始化方法:全零初始化

## 9.选择的loss函数及其导数:

## 10.实验结果和曲线图:

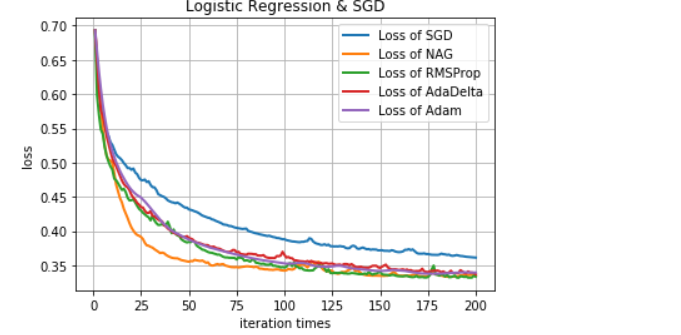
## 超参数选择：

## NAG(parameters, gradients, v, eta=0.05, gamma=.9

RMSProp(parameters, gradients, G, eta=.01, gamma=0.9, epsilon=1e-8

AdaDelta(parameters, gradients, G, dx, gamma=0.95, epsilon=1e-5

Adam(parameters, gradients, G, m, t, eta=0.01, gamma=0.999, beta=0.9, epsilon=1e-8



*2）线性分类与随机梯度下降*

## 6.实验步骤:

· 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。

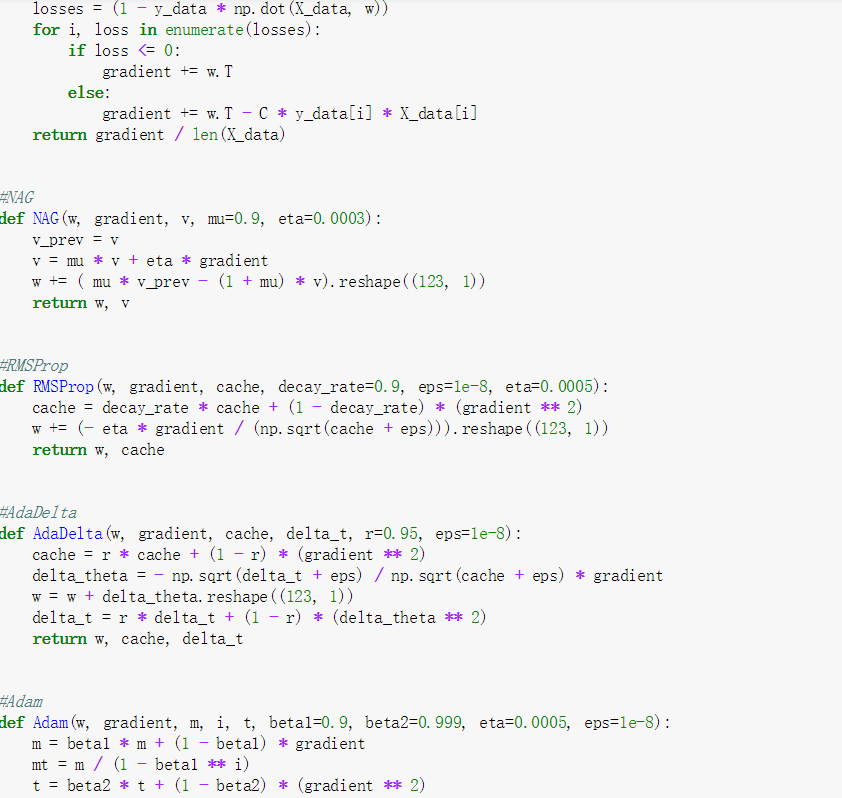
· ****使用不同的优化方法更新模型参数（NAG，RMSProp，AdaDelta和Adam）****。

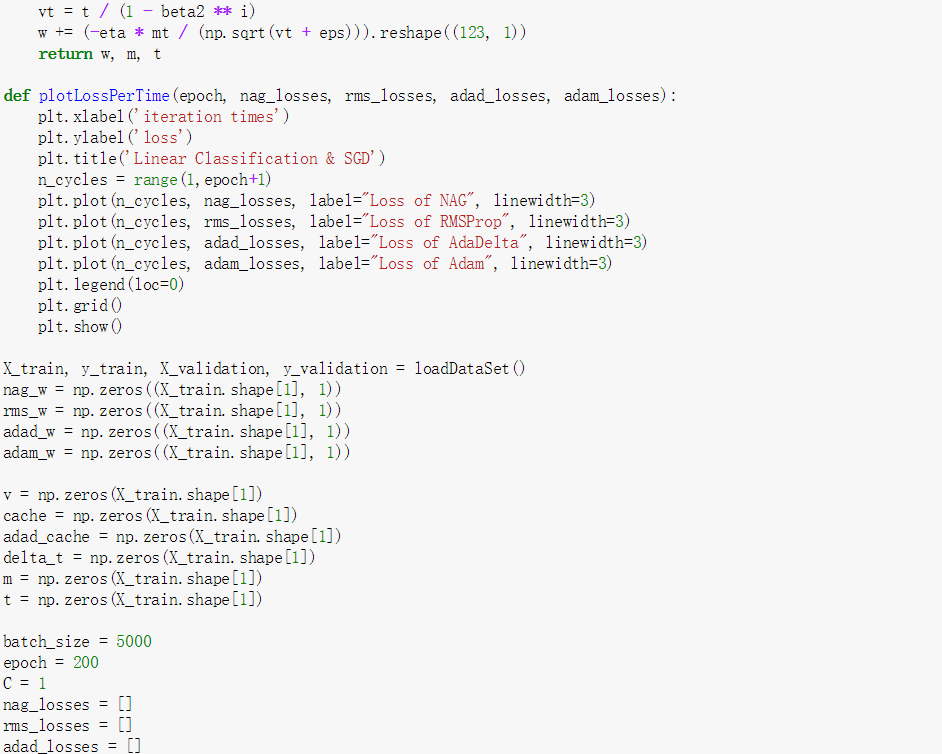
· 选择合适的阈值，将验证集中计算结果****大于阈值的标记为正类，反之为负类****。在验证集上测试并得到不同优化方法的Loss函数值，，和。

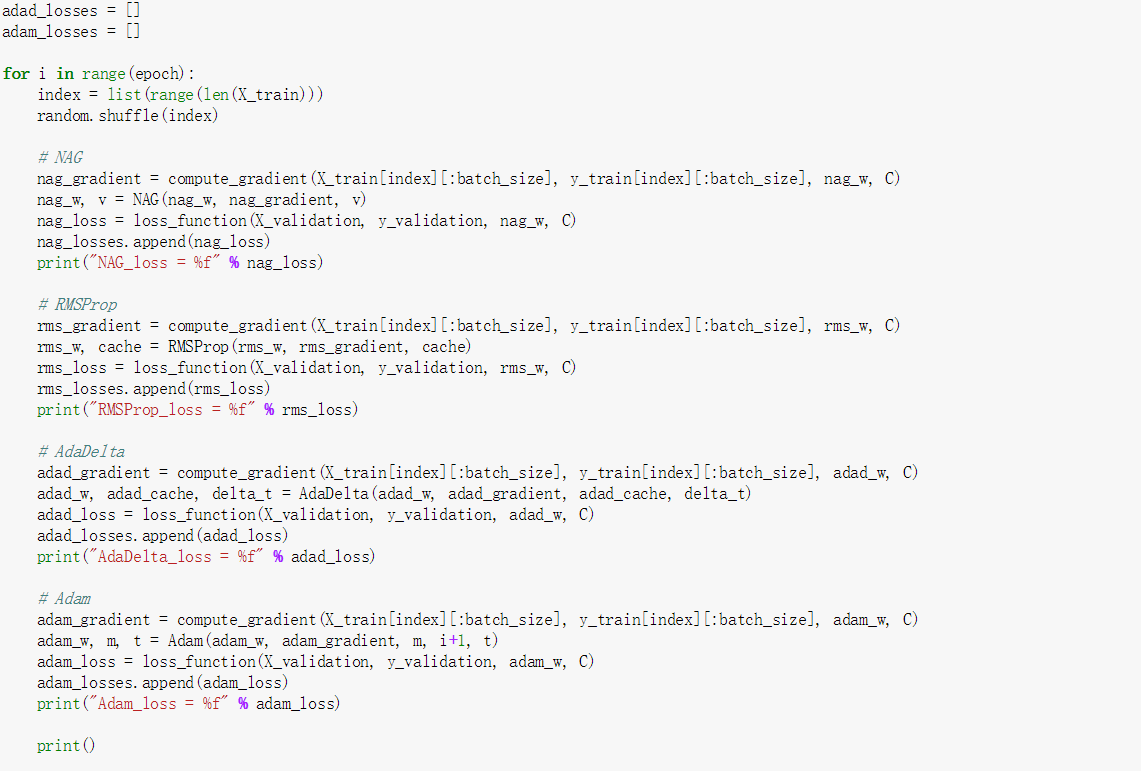
· 重复步骤4-6若干次，****画出，，和随迭代次数的变化图****。

## 7. 代码内容:









## 8. 模型参数的初始化方法:正态分布初始化

## 9.选择的loss函数及其导数:

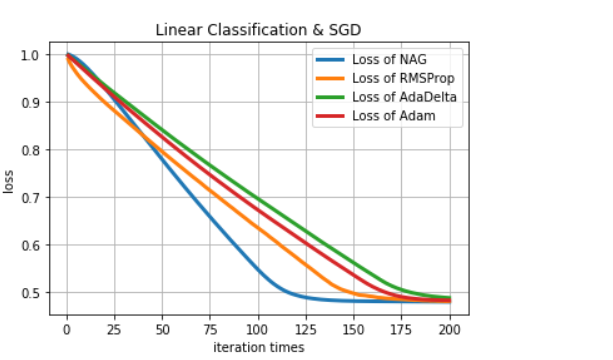
## 10.实验结果和曲线图:

## 超参数选择：NAG(w, gradient, v, mu=0.9, eta=0.0003

RMSProp(parameters, gradients, G, eta=.01, gamma=0.9, epsilon=1e-8

AdaDelta(w, gradient, cache, delta\_t, r=0.95, eps=1e-8

Adam(w, gradient, m, i, t, beta1=0.9, beta2=0.999, eta=0.0005, eps=1e-8



## 实验结果分析:

以NAG以外的其他方法更新参数收敛没有NAG快，但都成功收敛了

1. 对比逻辑回归和线性分类的异同点：

两种方法都是常见的分类算法,从[目标函数](https://www.baidu.com/s?wd=%E7%9B%AE%E6%A0%87%E5%87%BD%E6%95%B0&tn=44039180_cpr&fenlei=mv6quAkxTZn0IZRqIHckPjm4nH00T1YLrHTdnj-brj99nyFbrHKh0ZwV5Hcvrjm3rH6sPfKWUMw85HfYnjn4nH6sgvPsT6KdThsqpZwYTjCEQLGCpyw9Uz4Bmy-bIi4WUvYETgN-TLwGUv3EnHTsn1msnjTYPH6kn1cYrjmzn0" \t "https://zhidao.baidu.com/question/_blank)来看,区别在于逻辑回归采用的是logistical loss,svm采用的是hinge loss.这两个[损失函数](https://www.baidu.com/s?wd=%E6%8D%9F%E5%A4%B1%E5%87%BD%E6%95%B0&tn=44039180_cpr&fenlei=mv6quAkxTZn0IZRqIHckPjm4nH00T1YLrHTdnj-brj99nyFbrHKh0ZwV5Hcvrjm3rH6sPfKWUMw85HfYnjn4nH6sgvPsT6KdThsqpZwYTjCEQLGCpyw9Uz4Bmy-bIi4WUvYETgN-TLwGUv3EnHTsn1msnjTYPH6kn1cYrjmzn0" \t "https://zhidao.baidu.com/question/_blank)的目的都是增加对分类影响较大的数据点的权重,减少与分类关系较小的数据点的权重.SVM的处理方法是只考虑support vectors,也就是和分类最相关的少数点,去学习分类器.而逻辑回归通过非线性映射,大大减小了离分类平面较远的点的权重,相对提升了与分类最相关的数据点的权重.两者的根本目的都是一样的.此外,根据需要,两个方法都可以增加不同的正则化项,如l1,l2等等.所以在很多实验中,两种算法的结果是很接近的.  
但是逻辑回归相对来说模型更简单,好理解,实现起来,特别是大规模线性分类时比较方便.而SVM的理解和优化相对来说复杂一些.但是SVM的理论基础更加牢固,有一套结构化风险最小化的理论基础,虽然一般使用的人不太会去关注.还有很重要的一点,SVM转化为对偶问题后,分类只需要计算与少数几个支持向量的距离,这个在进行复杂核函数计算时优势很明显,能够大大简化模型和计算  
svm 更多的属于非参数模型,而logistic regression 是参数模型,本质不同.其区别就可以参考参数模型和非参模型的区别就好了.  
logic 能做的 svm能做,但可能在准确率上有问题,svm能做的logic有的做不了

## 实验总结：

进一步熟悉调参过程并且了解到了逻辑回归与线性分类各自的优缺点