

机器学习实验报告

学院: 软件学院

专业: 软件工程

|  |  |
| --- | --- |
| *作者:*  侯斯扬 | *指导教师:*  谭明奎 |
|  |  |
| *学号:*  201930380488 | *班级:*  软件2班 |

2021-10-7

# 逻辑回归和支持向量机

***摘要—对比理解梯度下降和批量随机梯度下降；理解逻辑回归和线性分类的区别；理解逻辑回归和支持向量机的原理***

## **介绍**

本实验的主要目的如下：对比理解梯度下降和批量随机梯度下降的区别与联系。对比理解逻辑回归和线性分类的区别与联系。进一步理解SVM的原理并在较大数据上实践。

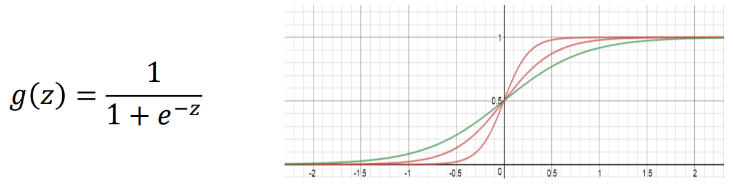
## **方法和理论**

1. 逻辑回归

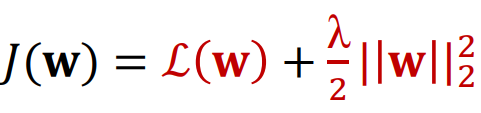
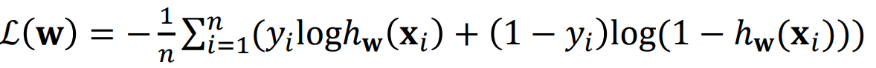
区别于线性回归：

线性回归：一般是指通过计算输入变量的加权和，并加上一个常数偏置项（截距项）来得到一个预测值。

逻辑回归：如何用连续的数值去预测离散的标签值？我们能否将线性回归输出的一个连续的数值变成一个标签呢？一个比较直观的想法是设定一个阈值，比如回归模型输出的y大于0时，属于正类，y小于0时属于负类，由此，我们有一个更好的方法，将y等于1和y等于100都归为正类的同时，也考虑它们各自属于正类的置信度，这便是逻辑回归

****逻辑回归的sigmoid函数

****逻辑回归通过估计样本属于正类或者负类的概率，再通过以上公式判断是属于哪个类别

逻辑回归的损失函数：（由对数的似然函数构造损失函数）

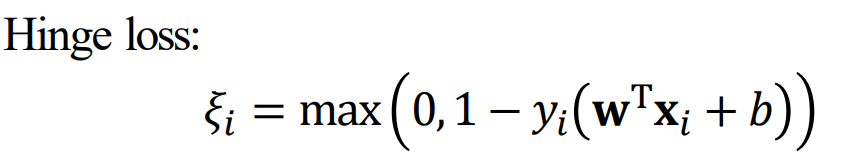
运用梯度下降（Gradient Descent）来求解损失函数的最小值

学习率：是一个大于0的数，能够控制沿着某个方向走多长一段距离（但不是步长）一般随着迭代次数的增加，学习率逐渐减小

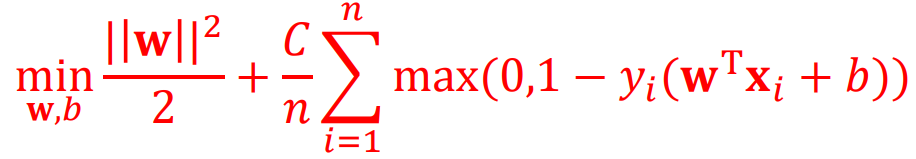
1. 支持向量机

在这个实验中我们讨论的是支持向量机的软间隔：实践中由于异常数据的存在，导致超平面不能完全将数据分为两部分，我们希望这样的样本越少越好

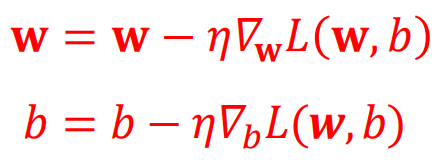
支持向量机的损失函数：



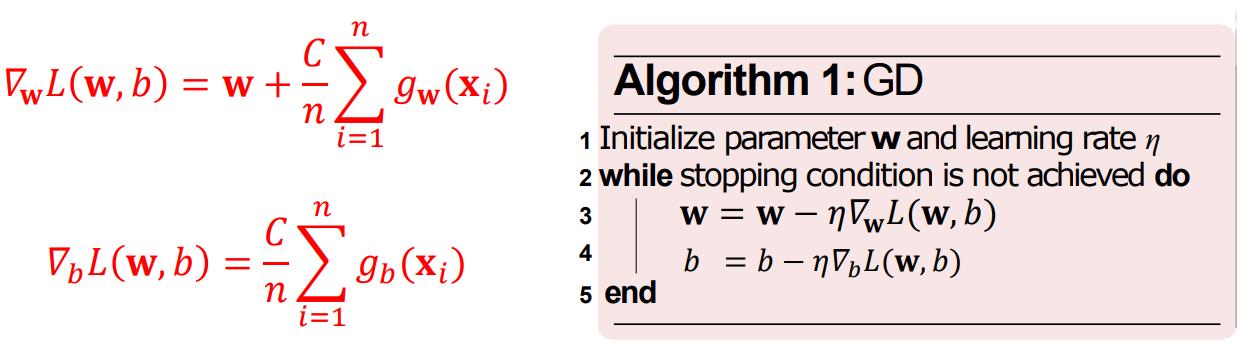
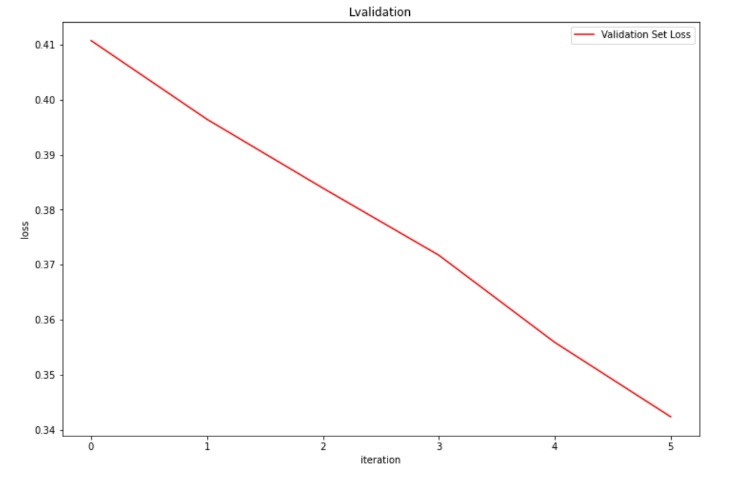
其优化：（就是在原优化函数的后面加了一截，这一截里面的C表示一个大于0的常数，损失函数取值为0或者1）

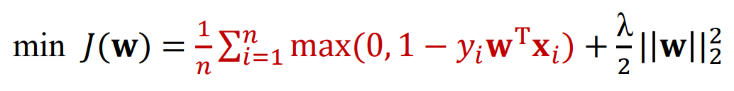


通过梯度下降，进行迭代

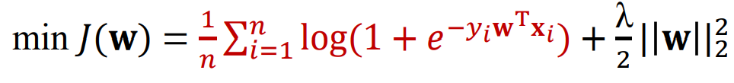


（本实验主要采用批量随机梯度下降）



1. 支持向量机（SVM）与逻辑回归的比较

支持向量机：

逻辑回归：

## **实验**

### 数据集

实验使用的是LIBSVM Data的中的a9a数据，包含32561/16281(testing)个样本，每个样本123/123 (testing)个属性。在读取数据时可能会出现维度不对的问题，是因为数据最后列全为零而被忽略，可以在下载的数据集文件后面自行添加后再读取，也可在读取数据集时指定n\_features=123来解决。

### 实现

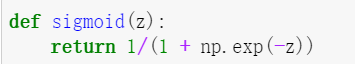
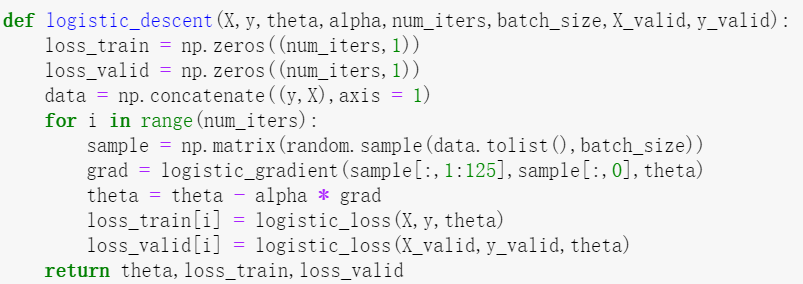
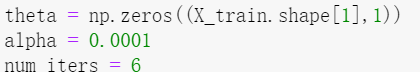
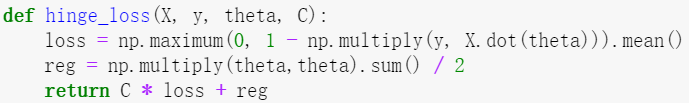
1. 对数据的处理（导入包的依赖，读取数据并切分为实验的训练集和验证集，进一步对数据进行类型的转换）
2. 图1.定义sigmoid函数
3. 图2.定义逻辑回归的损失函数
4. 图3.定义逻辑回归的批量随机梯度下降函数
5. 图4.确定batch\_size的大小，进行参数的初始化
6. 后执行梯度下降，进行6次迭代求解得到测试集的Lvalidation，其结果展示如下图5

图5. Lvalidation迭代结果图

1. 选取hinge作为损失函数

图6.Hinge loss的定义

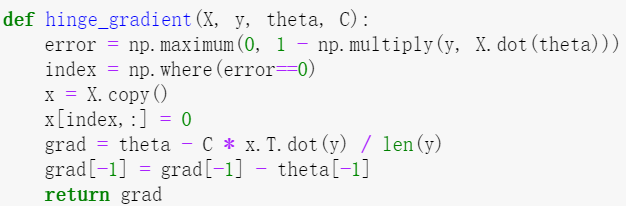
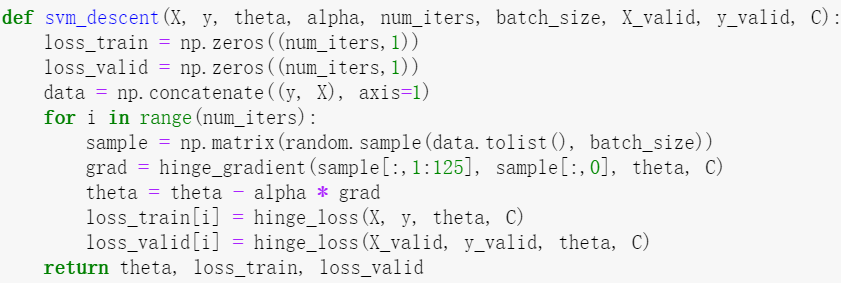
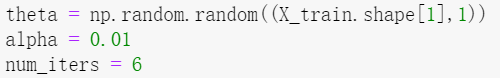
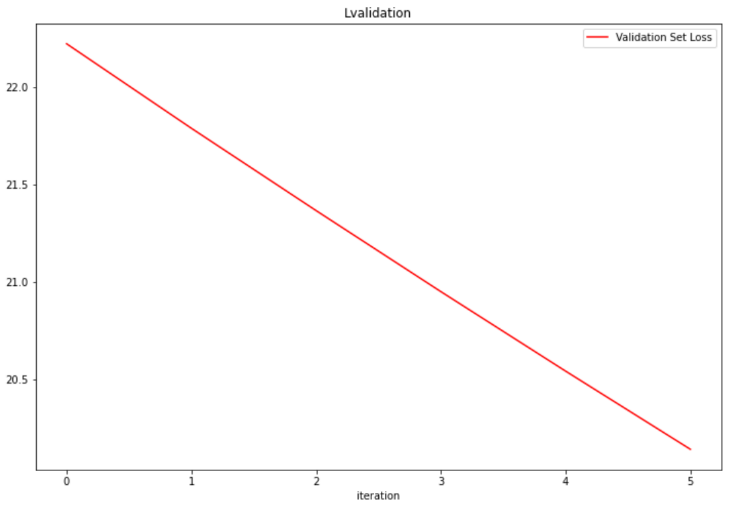
1. 图7.hinge回归函数的定义
2. 图8.支持向量机的梯度下降定义
3. 图9.进行模型参数的初始化
4. 调用SVM的梯度下降求解Lvalidation，迭代6次的结果如图10所示

图10. Lvalidation迭代结果

## **结论**

通过本次实验，我对于线性回归，线性分类有了更加明确的区分，对于逻辑回归以及支持向量机的原理有了更进一步的了解；

对于梯度下降的了解：批量梯度下降有准确度高但训练速度慢的特点，而随机梯度下降虽然准确度不高但速度快，本次实验也有才用到了二者的结合——批量随机梯度下降；

在此次实验中，也接触到了梯度上升，同时也Adam、SGD、Mini-Batch等算法，有一些虽然照着代码打哩一遍，但仍然似懂非懂，课后需要再花时间继续了解。