**使用 Python 实现对数**

**几率回归模型**

**课题报告**

姓名：苏宁

班级：计算机1604

时间：2018年10月11日

目录

[一、问题描述 2](#_Toc527058143)

[二、数据集描述： 3](#_Toc527058144)

[三、实验结果图： 4](#_Toc527058145)

[四、实验结果分析： 8](#_Toc527058146)

# 一、问题描述

编程实现对数几率回归模型，并对 Iris 数据集进行分类以验证模型的效能：

(1). 将数据集的 50%作为训练集，50%作为测试集，检验模型在测试集上的分类正确率

(2). 将数据集的 70%作为训练集，30%作为测试集，检验模型在测试集上的分类正确率

(3). 将数据集的 90%作为训练集，10%作为测试集，检验模型在测试集上的分类正确率

# 二、数据集描述：

数据集名称：Iris

实例数：150

属性数量：4

1. sepal length in cm    
2. sepal width in cm   
3. petal length in cm   
4. petal width in cm

5.类：   
- Iris Setosa   
- Iris Versicolour   
- Iris Virginica

PS：这里只采用了Setosa和Versicolour两个类的100个实例，每个实例取sepal length in cm sepal width in cm 两个属性。

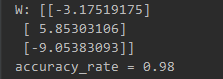
为了方便编程，在设计时，手动将两个label（Setosa和Versicolour）分别置为0和1

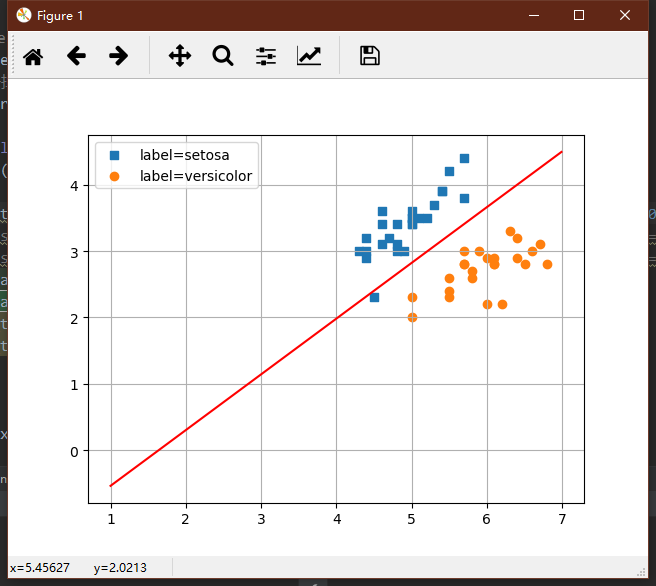
# 三、实验结果图：

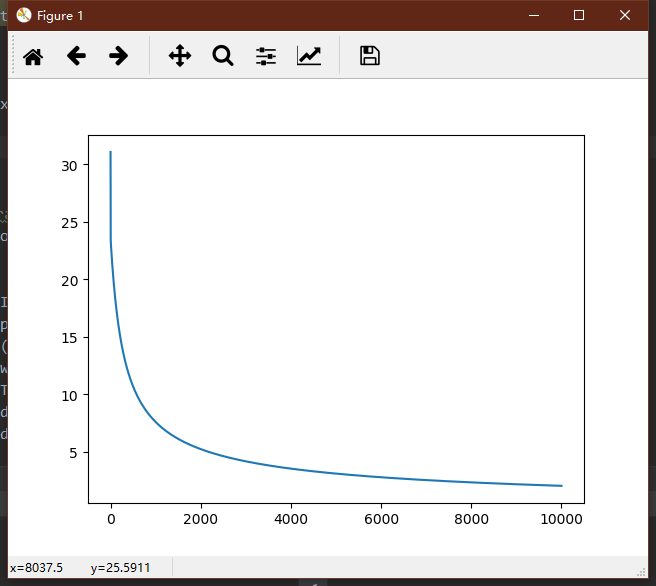
这里使用了第三方库sklearn来划分数据集为训练集和测试集

学习率alpha=0.001，最大迭代次数maxIter=10001

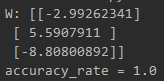
(1). 将数据集的 50%作为训练集，50%作为测试集.

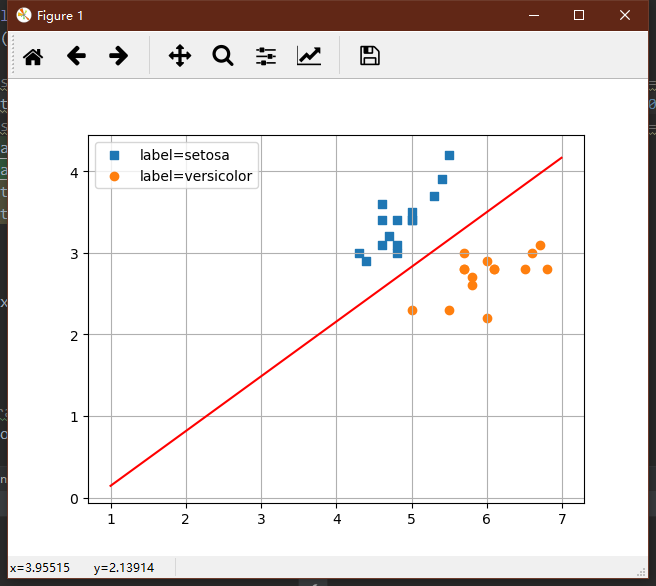


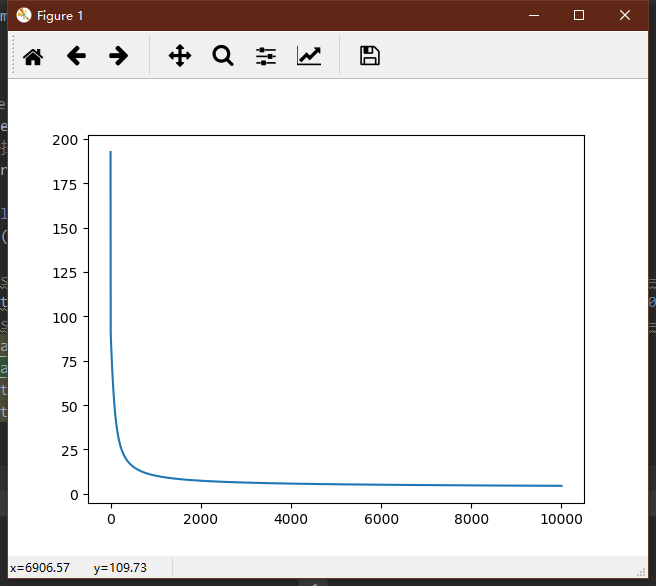


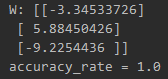


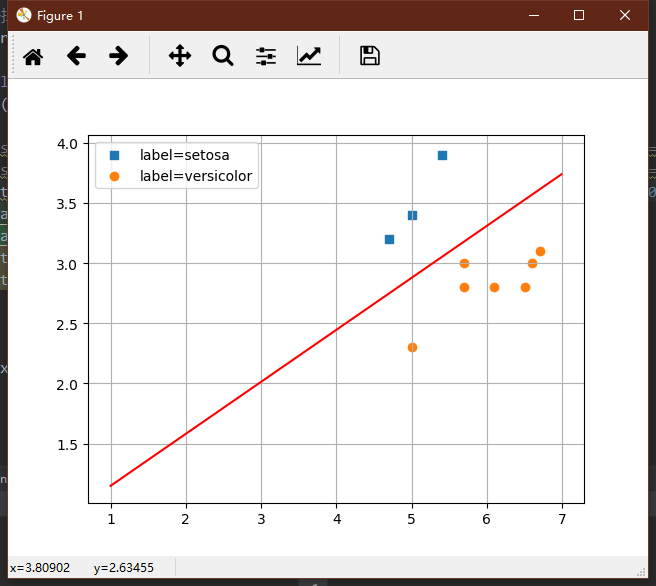
(2). 将数据集的 70%作为训练集，30%作为测试集.

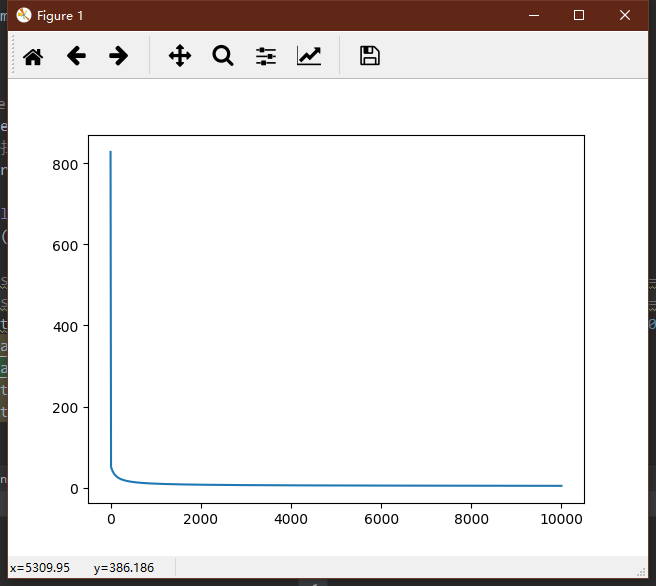






(3). 将数据集的 90%作为训练集，10%作为测试集. 





# 四、实验结果分析：

1.实现了利用多次迭代求解决策边界和损失函数的思想，利用相关库进行可视化方便直观的进行观察，并完成了对实验数据的划分和正确率的预测，基本实现作业要求。

2.由于数据总量并不大，经多次重复实验得，训练集为50%时，正确率在85%-98%浮动，训练集为70%以上时，测试时均达到100%正确率。

3.cost function图像与预期相符，随迭代次数逐渐下降，切线斜率由大变小趋近于0。