# 使用python实现对数几率回归模型

计算机1603 1611640310 姜浩东

### 一、问题描述

1、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

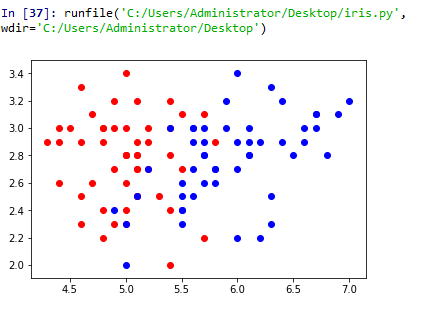
(3). 将数据集的90%作为训练集，10%作为测试集，检验模型在测试集上的分类正确率

2、鸢尾花可分为三个种类，可以通过花萼长度、宽度和花瓣长度、宽度的不同从而区分开来本次实验中，我们选择数据集中的两种进行实验。

### 二、数据集描述

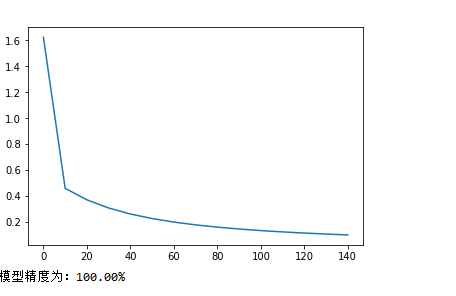
Iris也称鸢尾花卉数据集，是一类多重变量分析的数据集。通过花萼长度，花萼宽度，花瓣长度，花瓣宽度4个属性预测鸢尾花卉属于（Setosa，Versicolour，Virginica）三个种类中的哪一类。数据来源：<http://archive.ics.uci.edu/ml/datasets/Iris>

共150个数据，我们选择前一百个数据进行实现。我选择了两类花的前两种属性得来的散点图

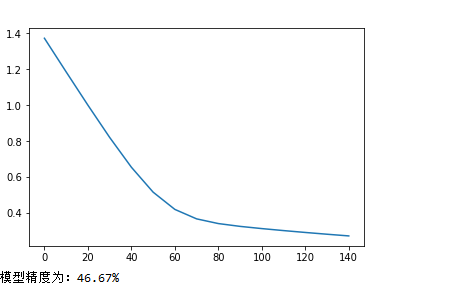


### 实验结果图

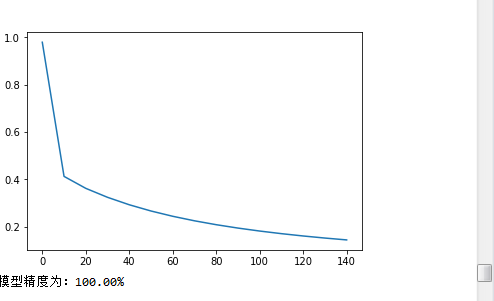
数据集50%作为训练，50%作为测试的损失函数变化曲线图



数据集70%作为训练，30%作为测试的损失函数变化曲线图



数据集90%作为训练，10%作为测试的损失函数变化曲线图



### 四、实验结果分析

一下为三种情况下的数据分析：

|  |  |  |  |
| --- | --- | --- | --- |
| 训练集个数 | 测试集个数 | 测试成功个数 | 精度 |
| 50 | 50 | 50 | 100% |
| 70 | 30 | 13 | 46.77% |
| 90 | 10 | 10 | 100% |

三次实验最后精度有很大差异，我认为是自己的迭代次数及学习率设置的不恰当，但是多次试验后还是得不到一个满意数据，可能是因为我的能力有限，不能设计出合理的回归模型。