**使用python实现对数几率回归**

计算机1601 常雨昂 1611640119

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 问题目标利用对数几率回归实现二分类

数据集描述 摘自archive.ics.uci.edu/ml/datasets/Iris

Data Set Information:

This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda & Hart, for example.) The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.   
  
Predicted attribute: class of iris plant.   
  
This is an exceedingly simple domain.   
  
This data differs from the data presented in Fishers article (identified by Steve Chadwick, spchadwick **'@'** espeedaz.net ). The 35th sample should be: 4.9,3.1,1.5,0.2,"Iris-setosa" where the error is in the fourth feature. The 38th sample: 4.9,3.6,1.4,0.1,"Iris-setosa" where the errors are in the second and third features.

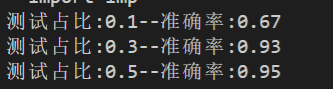
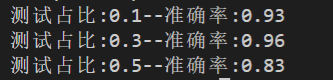
Arrtibute Information:

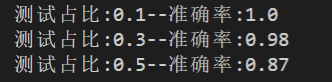
1. sepal length in cm   
2. sepal width in cm   
3. petal length in cm   
4. petal width in cm   
5. class:   
-- Iris Setosa   
-- Iris Versicolour   
-- Iris Virginica

来源：archive.ics.uci.edu/ml/datasets/Iris

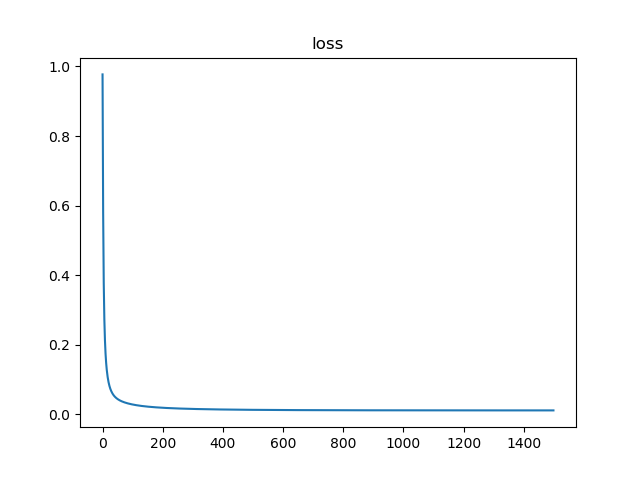
实验结果图

（注：红色（训练集）和品红（测试集）代表iriis-setosa 绿色（训练集）和青色（测试集）代表iris-versicolor）

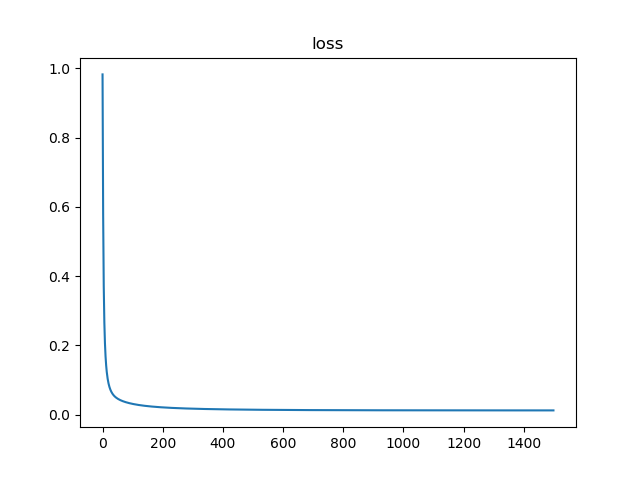




一、90%训练 10%测试

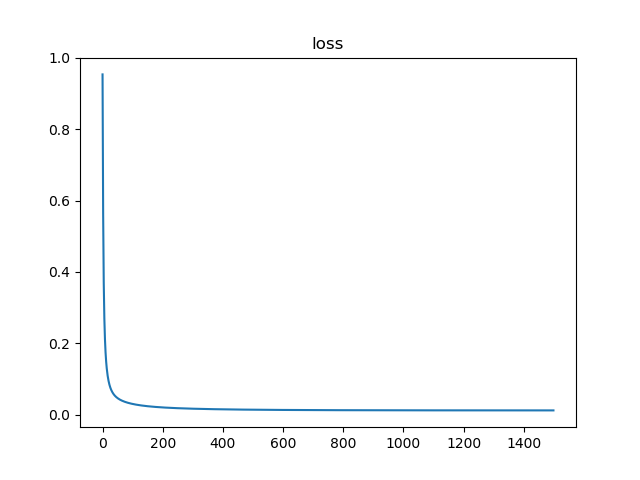
**第一次正确率为：93.0%**

二、70%训练 30%测试

第一次正确率为：96.0%

三、50%训练50%测试

第一次正确率为：83.0%



实验结果分析

通过本次实验，可以看出Iris数据集花的特征差异很大，相对容易分辨，随着训练次数增加，准确率会逐步提高