|  |
| --- |
|  |
| 第一次课题报告 |
|  |

|  |
| --- |
| 徐钰东 计算机1604 1611640409 |

1. 题目

使用Python实现对数几率回归模型

1. 问题描述

利用python编程实现对数几率回归模型，将鸢尾花数据集（Iris）中的两种花Iris-setosa和Iris-versicolor的二分类模型训练出来（通过给定的属性来预测花的种类）并验证模型的效能：

(1). 将数据集的 50%作为训练集，50%作为测试集

(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

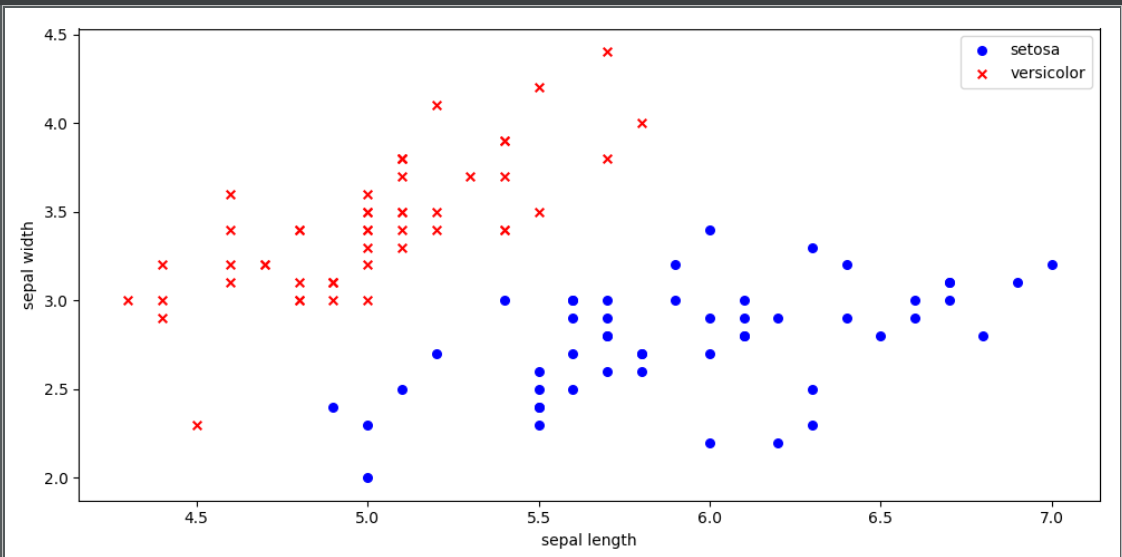
并检验模型在测试集上的分类正确率

1. 数据集描述

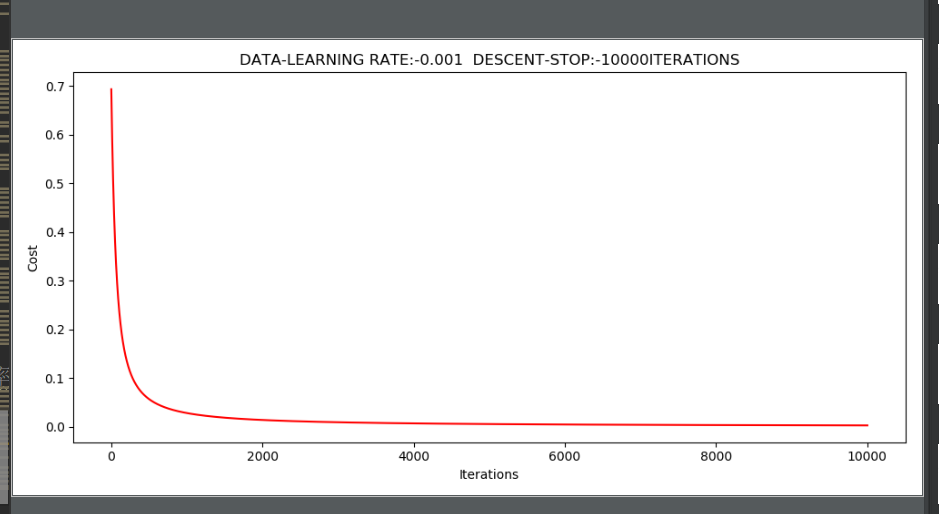
鸢尾花数据集（Iris）的两种花Iris-setosa和Iris-versicolor各包含50组数据，每组数据有四个属性项(sepal length,sepal width,petal length,petal width)，一个种类项（kind）.在训练模型中为了计算将Iris-setosa表示为0，将Iris-versicolor表示为1。

1. 实验结果图，包括随着训练次数的增加，模型损失函数的变化曲线图

两种花（以sepal length,sepal width两种属性）的散点图

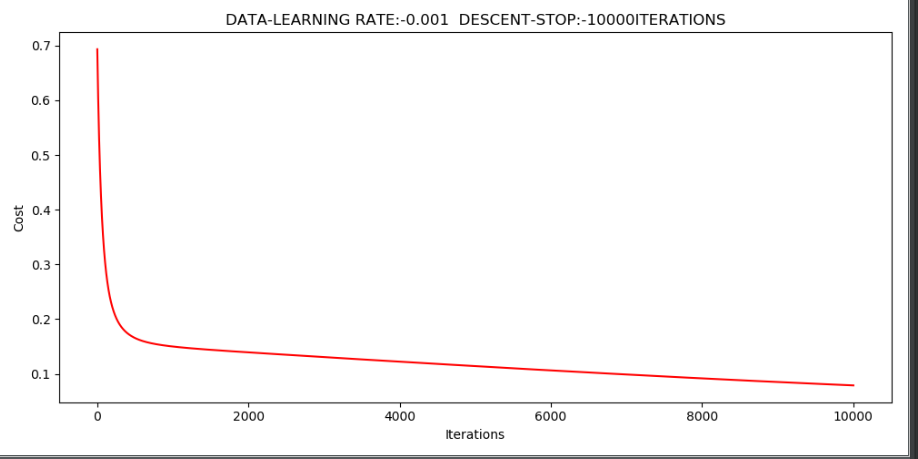


(1). 将数据集的 50%作为训练集，50%作为测试集



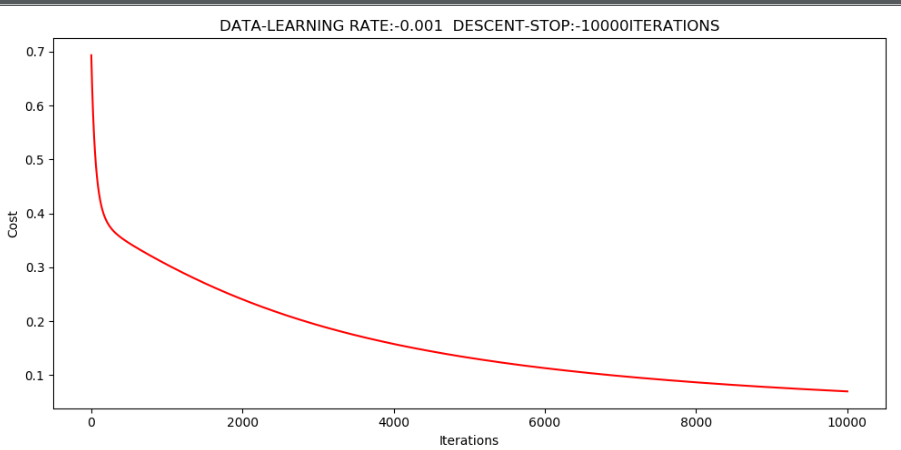
截图的该次训练的Accuracy=72%

(2). 将数据集的 70%作为训练集，30%作为测试集



截图的该次训练的Accuracy=80%

(3). 将数据集的 90%作为训练集，10%作为测试集



截图的该次训练的Accuracy=100%

1. 实验结果的分析:

通过对数几率回归模型的训练（设置学习率为0.001，迭代次数为10000次），得到了能预测鸢尾花的类别的模型，将数据集按比例分为训练集和测试集的不同情况下，发现训练集占比越大，得到的模型在测试时的准确度越高，泛化能力相对较强。在训练过程中随着训练次数的增加，得到模型的损失函数是在不断下降的。当然，本次实验的数据集比较小，训练得到的模型的泛化能力相对会较差一些。