课题报告

计算机1601 张博 1610830313

1. **题目**：用对数几率回归模型实现鸢尾花数据集分类

**二、问题描述**

用python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：分别将数据集的50%、70%、90%作为训练集，剩下的作为测试集，检验模型在测试集上的分类正确率。

**三、数据集描述**

Iris数据集是常用的分类实验数据集，由Fisher, 1936收集整理。Iris也称鸢尾花卉数据集，是一类多重变量分析的数据集。数据集包含150个数据集，分为3类，每类50个数据，每个数据包含4个属性。可通过花萼长度，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。

iris以鸢尾花的特征作为数据来源，常用在分类操作中。该数据集由3种不同类型的鸢尾花的50个样本数据构成。其中的一个种类与另外两个种类是线性可分离的，后两个种类是非线性可分离的。该数据集包含了5个属性：

& Sepal.Length（花萼长度），单位是cm;

& Sepal.Width（花萼宽度），单位是cm;

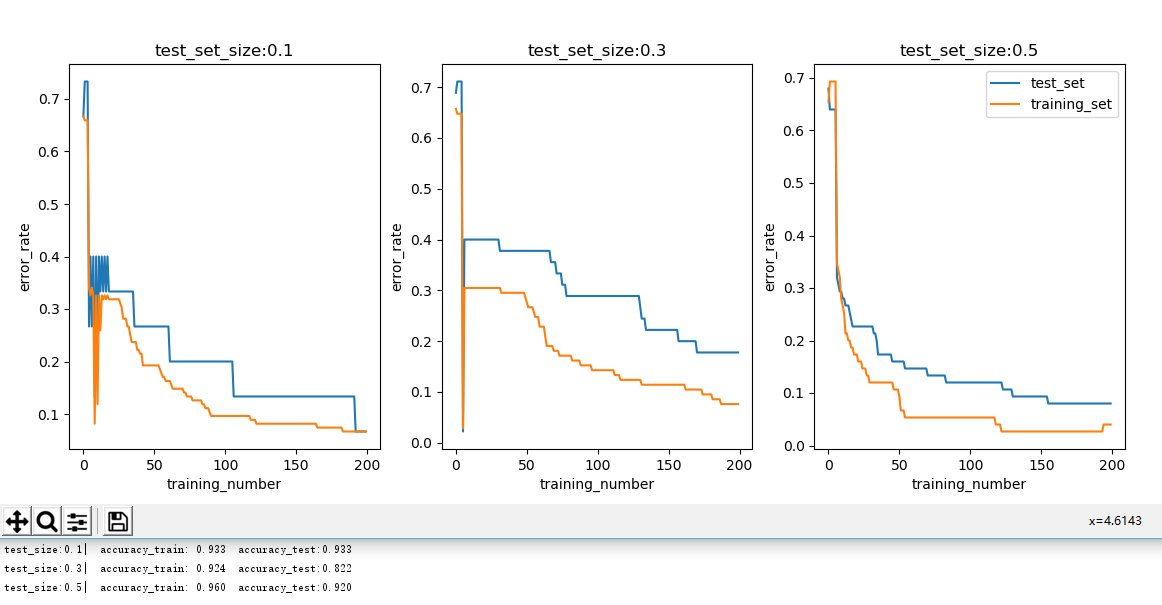
& Petal.Length（花瓣长度），单位是cm;

& Petal.Width（花瓣宽度），单位是cm;

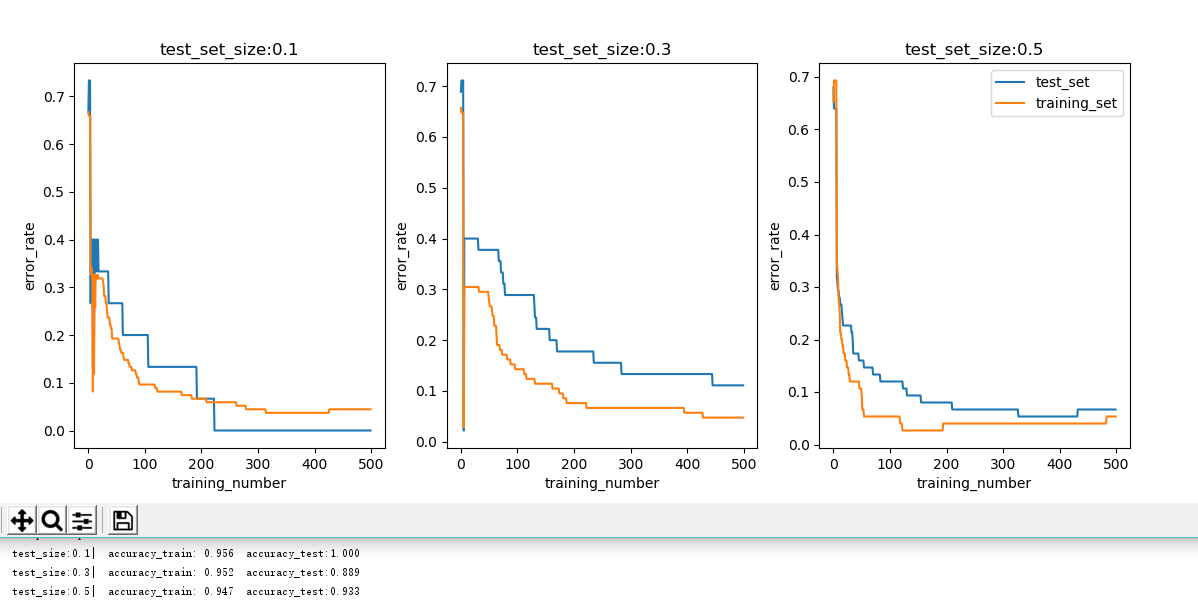
& 种类：Iris Setosa（山鸢尾）、Iris Versicolour（杂色鸢尾），以及Iris Virginica（维吉尼亚鸢尾）。

1. **实验结果图**

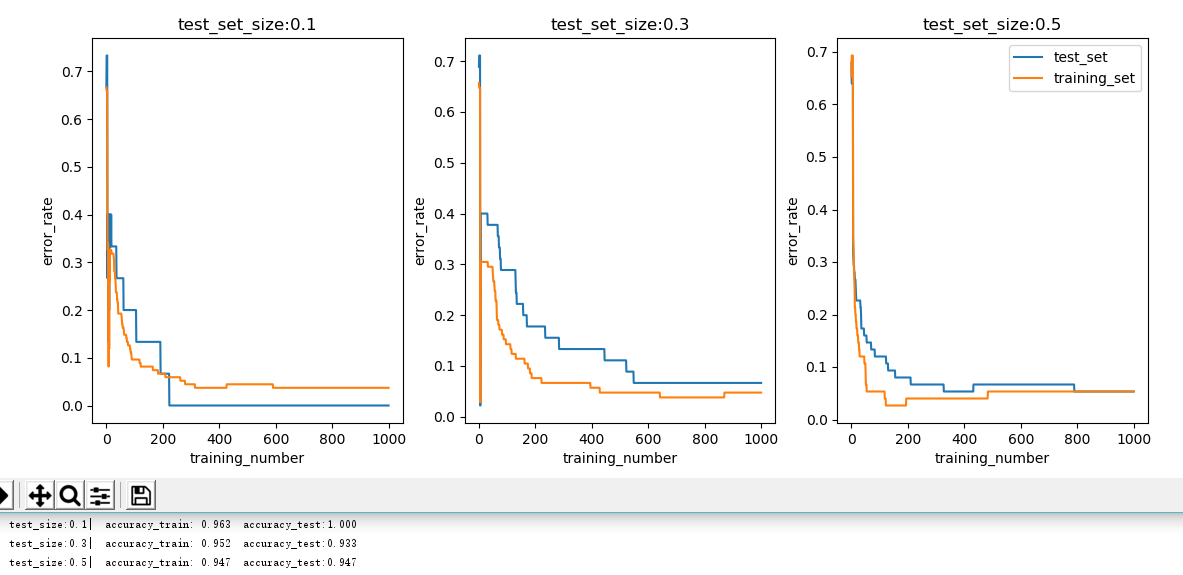
训练次数为200时：



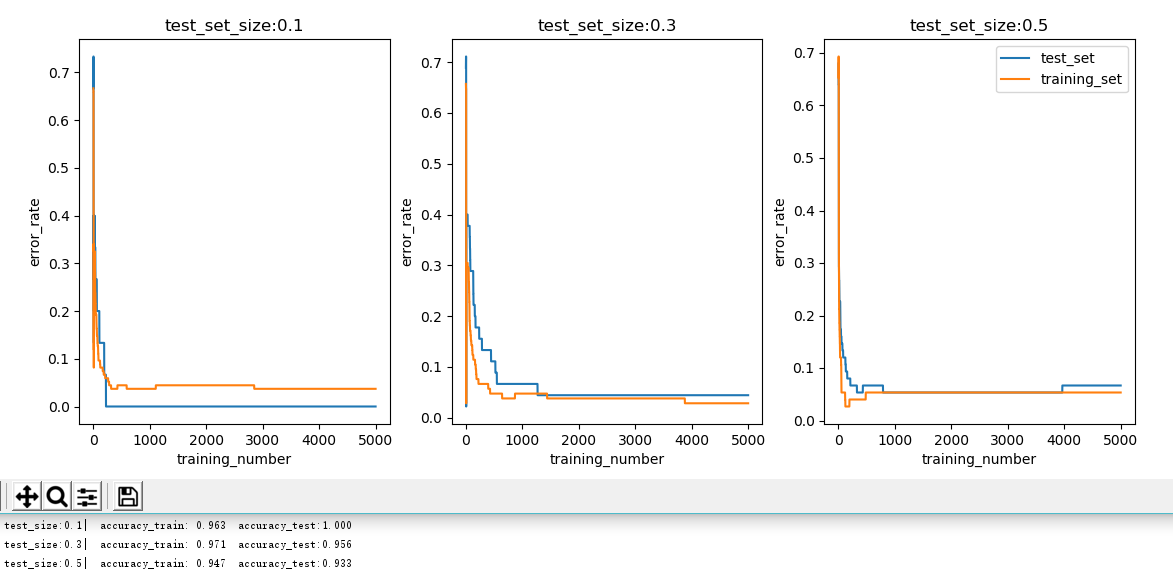
训练次数为500时：



训练次数为1000时：



训练次数为5000时：



1. **实验结果分析**

随着训练次数的增加，预测的正确率越来越高，训练集和测试集的曲线拟合度越高。