**Python实现鸢尾花数据集的对数几率回归模型**

**班级：计算机1603**

**学号：1611640320**

**姓名：全力**

一**、问题描述**

鸢尾花分为三个种类，分别为山鸢尾、杂色鸢尾、维吉尼亚鸢尾。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。利用python来构建对数几率数学模型来实现对鸢尾花的分类，在这里只选用前两种种类的鸢尾花来进行分类，即实现二分类。

**二、数据集描述**

该数据集由3种不同类型的鸢尾花（每种50个样本），共150个样本数据构成。

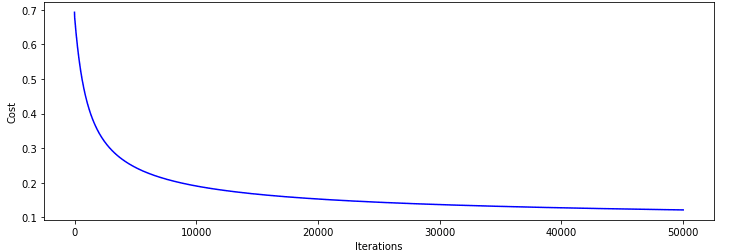
该数据集包含了4个属性：

Sepal.Length（花萼长度）、Sepal.Width（花萼宽度）、Petal.Length（花瓣长度）、Petal.Width（花瓣宽度）；

由于本次实验目的在于训练一个最基本的机器学习模型，在此只选用了鸢尾花前两种特征进行训练和分类。

**三、 实验结果图**

**1.损失函数变化曲线图**



1. **选取50%测试集的分类结果**

IMG_256

**3.选取30%测试集的分类结果**

IMG_256

**4.选取10%测试集的分类结果**

YC0GVY5([~Y_PX1SL]W{F]E

**四、实验结果分析**

以下为相同迭代次数和学习率下的三次分类试验结果表：

|  |  |  |  |
| --- | --- | --- | --- |
| 测试集比例 | 测试样本数 | 测试成功个数 | 分类正确率 |
| 50% | 50 | 47 | 92.00% |
| 30% | 30 | 29 | 96.67% |
| 10% | 10 | 10 | 100.00% |

从上表可以看出，训练数据多的模型的泛化性能好，预测结果准确率高。

在实验中发现模型优化的快慢和好坏与迭代次数和学习率的设置，发现梯度下降迭代50000次并将学习率设为0.05之后的分类效果更加良好。同时也可以以损失作为判断是否停止迭代的依据，且效果比以迭代次数作为依据的效果更好。