一、

题目：使用 Python 实现对数几率回归模型

二、

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

三、

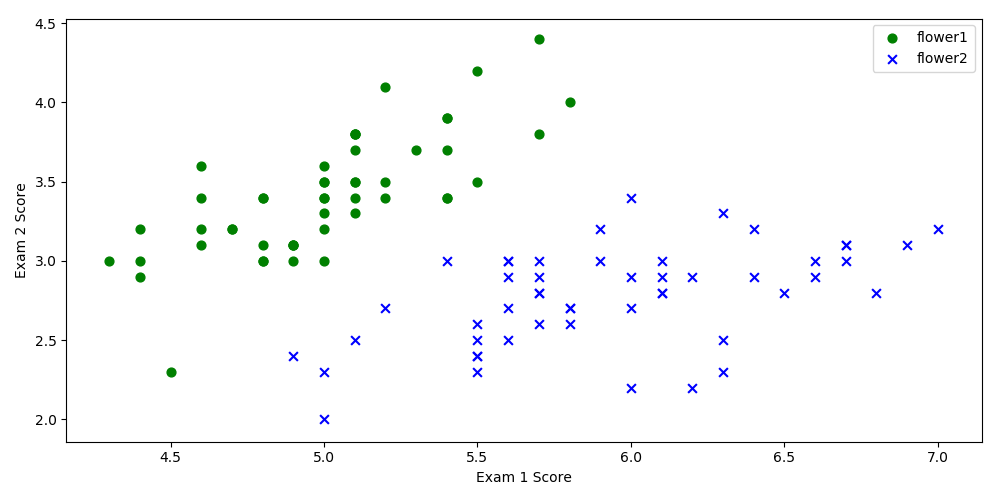
数据集：

取两类Iris-setosa，Iris-versicolor各50个

将Iris-setosa 取为1，Iris-versicolor取为0

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 5.1 | 3.5 | 1.4 | 0.2 | 1 |
| 4.9 | 3 | 1.4 | 0.2 | 1 |
| 4.7 | 3.2 | 1.3 | 0.2 | 1 |
| 4.6 | 3.1 | 1.5 | 0.2 | 1 |
| 5 | 3.6 | 1.4 | 0.2 | 1 |
| 5.4 | 3.9 | 1.7 | 0.4 | 1 |
| 7 | 3.2 | 4.7 | 1.4 | 0 |
| 6.4 | 3.2 | 4.5 | 1.5 | 0 |
| 6.9 | 3.1 | 4.9 | 1.5 | 0 |
| 5.5 | 2.3 | 4 | 1.3 | 0 |
| S1 | S2 | S3 | S4 | SS |

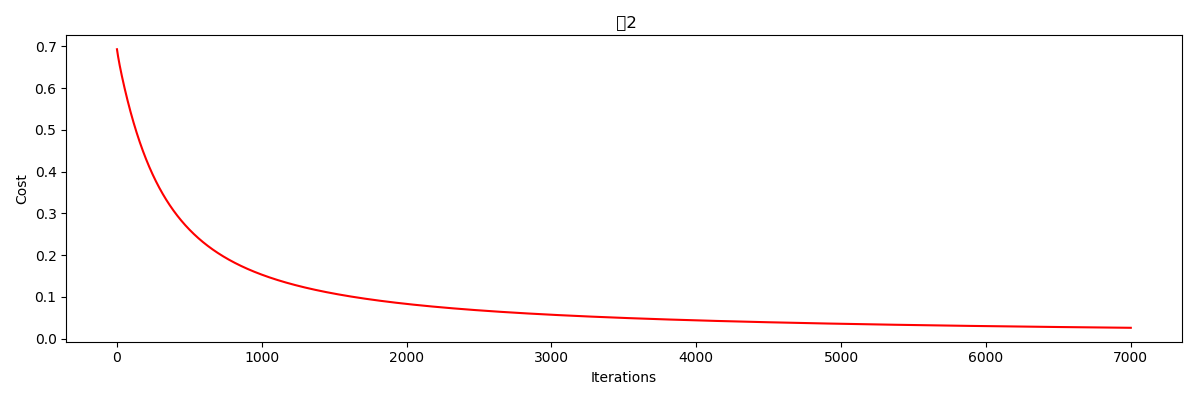
数据散点图:



四、

损失函数图：

a=0.004



五、

实验结果分析：

训练集50个数据，测试集50个数据 注：训练集测试集数据都是两类数据各一半

1. theta=[[ 0.23034881 0.36691169 1.49035704 -2.23096065 -0.9740326 ]]
2. 正确率100%

训练集70个数据，测试集30个数据

1. theta=[[ 0.25449761 0.37211719 1.49070639 -2.20285813 -0.96426419]]
2. 正确率100%

训练集90个数据，测试集10个数据

1）theta=[[ 0.26340734 0.40358758 1.46607403 -2.23152093 -0.95894687]]

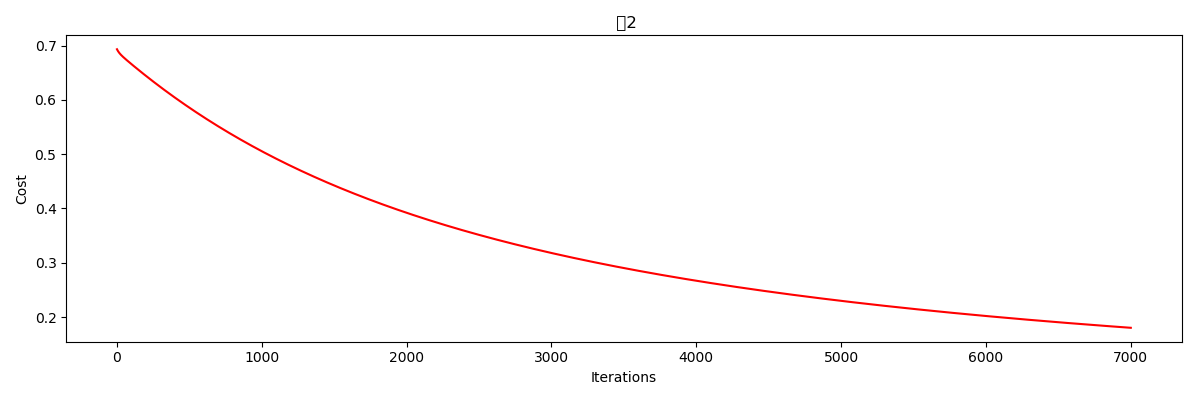
2）正确率100%

随着迭代次数的增加，损失值渐渐减小

为了查看训练查看数据集是四个属性与两个属性的差距，我算了一下两个属性的

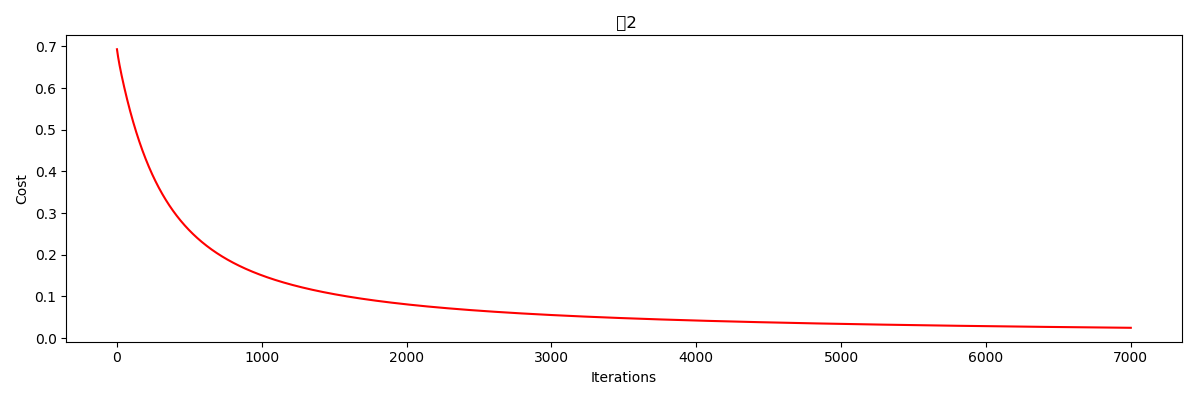
损失图像（两个属性）：

a=0.004 ，步数 7000



损失图像（四个属性）：

a=0.004 ,步数7000



很明显用四个属性的损失函数图像的值下降地比较快，说明四个属性，梯度的方向更准确，所以下降地更准确。