使用 Python 实现对数几率回归模型

1. 问题描述：

要训练一个函数，通过花的四个属性估测花为哪个种类的概率。控制数据集、训练集的比例，分析结果

1. 数据集描述

样本有四个属性，共有两种花。花的种类标记为0，1.属性与种类构成了100X5的DATA矩阵。实验要求数据集：训练集分别为1：1、7：3、9：1。提取数据，要在两类花中等比例提取。比如1：1时，1类花要25训练、25测试，0类花要25训练、25测试

三、思路

要学习的参数：β = （w；b） 5X1

βx尖：参数对某个样本的预测值。训练集中每个样本有其真实的标记（0或1），β对每个样本提供其属于其真实标记的概率

1.要求对数回归模型的最大值，等价于求另一个较好求的函数l（β）的最小值。不方便求其解析解，所以运用梯度下降法，不断优化β。

2.任取一点β0=（0；0；0；0；0），对l（β0）求梯度，返回一个方向。函数在这个方向相 反的方向下降最快。所以β0 = β0 – 步长step\*梯度l（β0），得到新的“落脚点”，继续求梯度。以此来寻找函数的极小值。

3.设定步长、训练次数。

4.在数据的三种划分下分别实验，分析结果

四、结果

正确率：

1：1 正确率：0.9

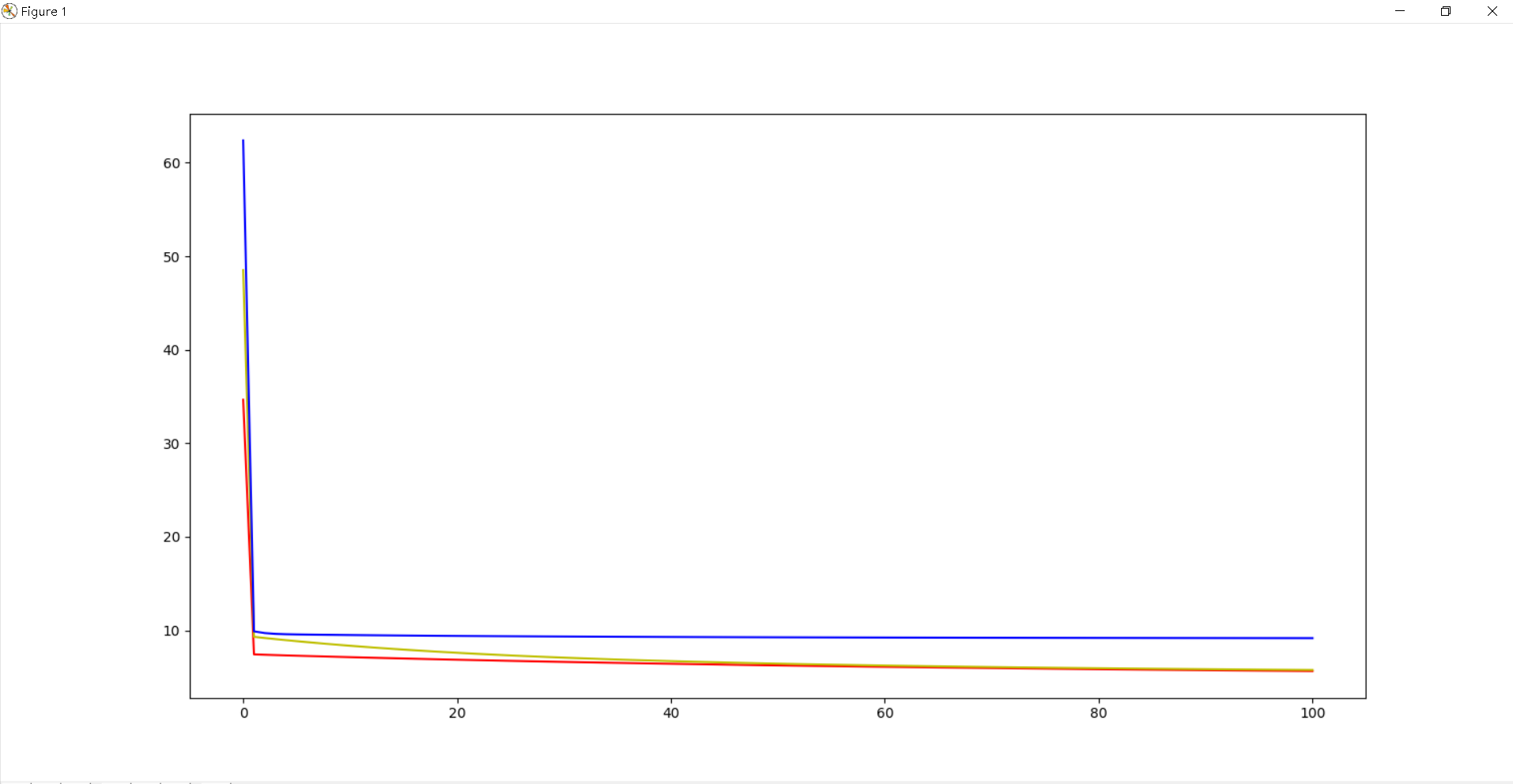
7：3正确率：0.866

9：1正确率：0.7

图为三次实验的损失函数：

红线为1：1 黄线为7：3 蓝线为 9：1

横轴为训练次数（每训练20次取样），纵轴为损失函数值



五、结果分析：

训练次数多的正确率反而更小。推测可能性：没有对β取不同初值进行梯度下降。没有多次划分不同的训练集测试集求解。结果可能不平均。也有可能是过多次的训练造成过拟合。还有可能是测试集太少，结果不平均。