目录

[多层感知机 1](#_Toc484973854)

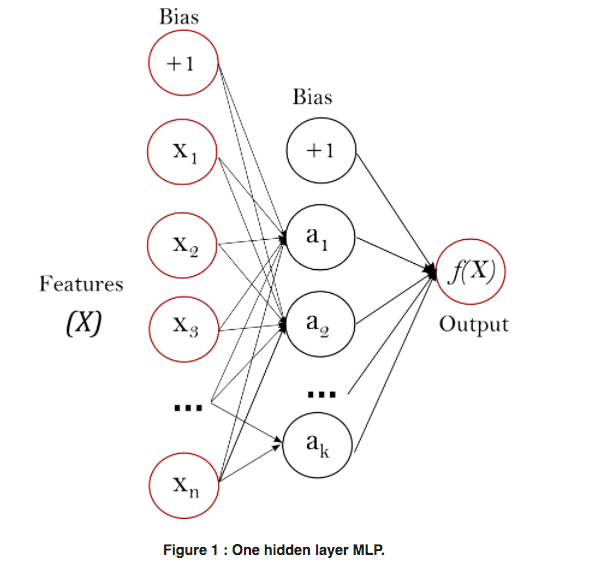
[类别不平衡问题解决 4](#_Toc484973855)

[欠采样 4](#_Toc484973856)

[调整阈值 6](#_Toc484973857)

# 多层感知机

多层感知机的基本模型图如下：

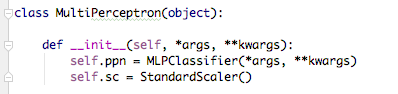


在本系统中，多层感知机使用基于后向传播算法的梯度下降进行训练。同时，使用softmax函数作为输出函数。

本模型的基本配置信息如下：

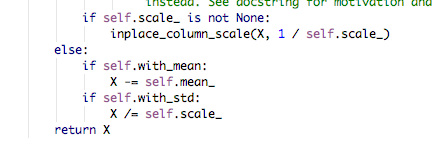


神经网络包含2个隐含层，其中每个隐含层有五个神经元，学习速率设定为0.001，迭代次数设定为50次。



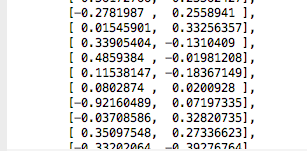
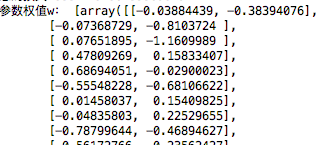
然后对数据进行均值归一与特征缩放：

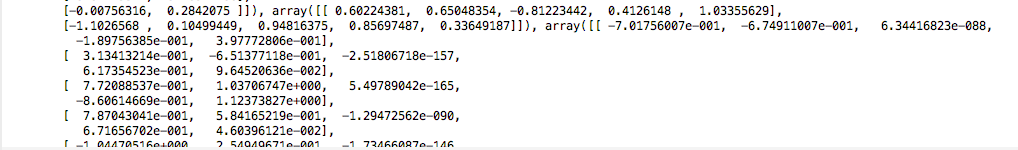




使用后向传播算法进行训练，训练出的模型结果为：





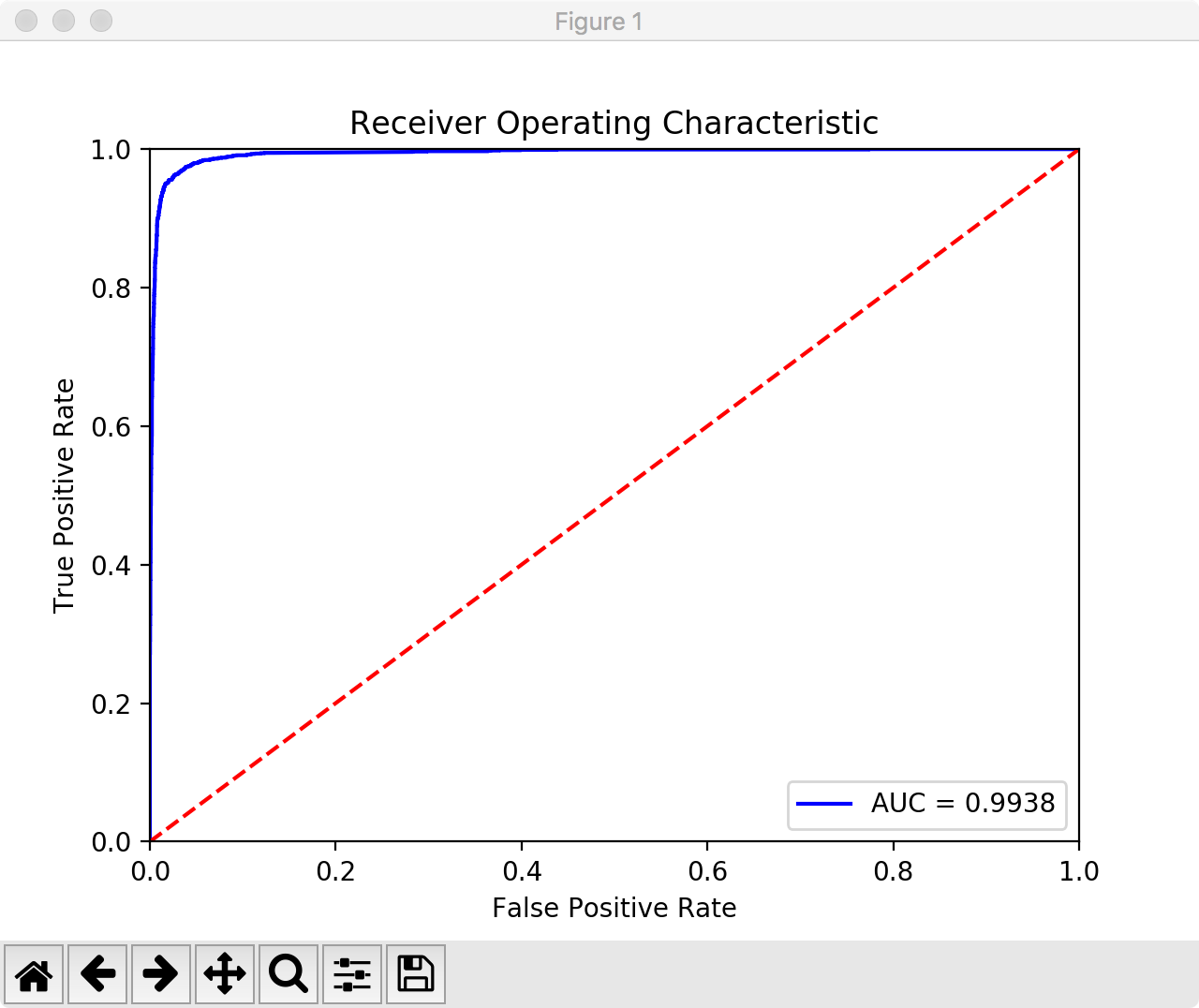




然后对测试数据进行预测，结果为：



绘制出ROC曲线：



ROC曲线是根据softmax函数的结果（两个类的概率）绘制出的：



# 类别不平衡问题解决

在本数据集中，正类与负类的比例是1：4，这就导致即使我们把所有的数据都判别为负类，对于测试集而言，就已经有个80%的正确率。这显然与训练模型的初衷不符合。所以我们要首先保证正类尽可能少的错分。

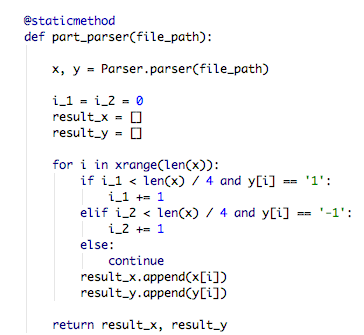
解决类别不平衡有一些策略：

1. 对正类过采样，就是创造更多的正类
2. 对负类欠采样，删去部分负类
3. 调整决策的阈值
4. 调整类的权重

这里选用其中一些策略进行测试。

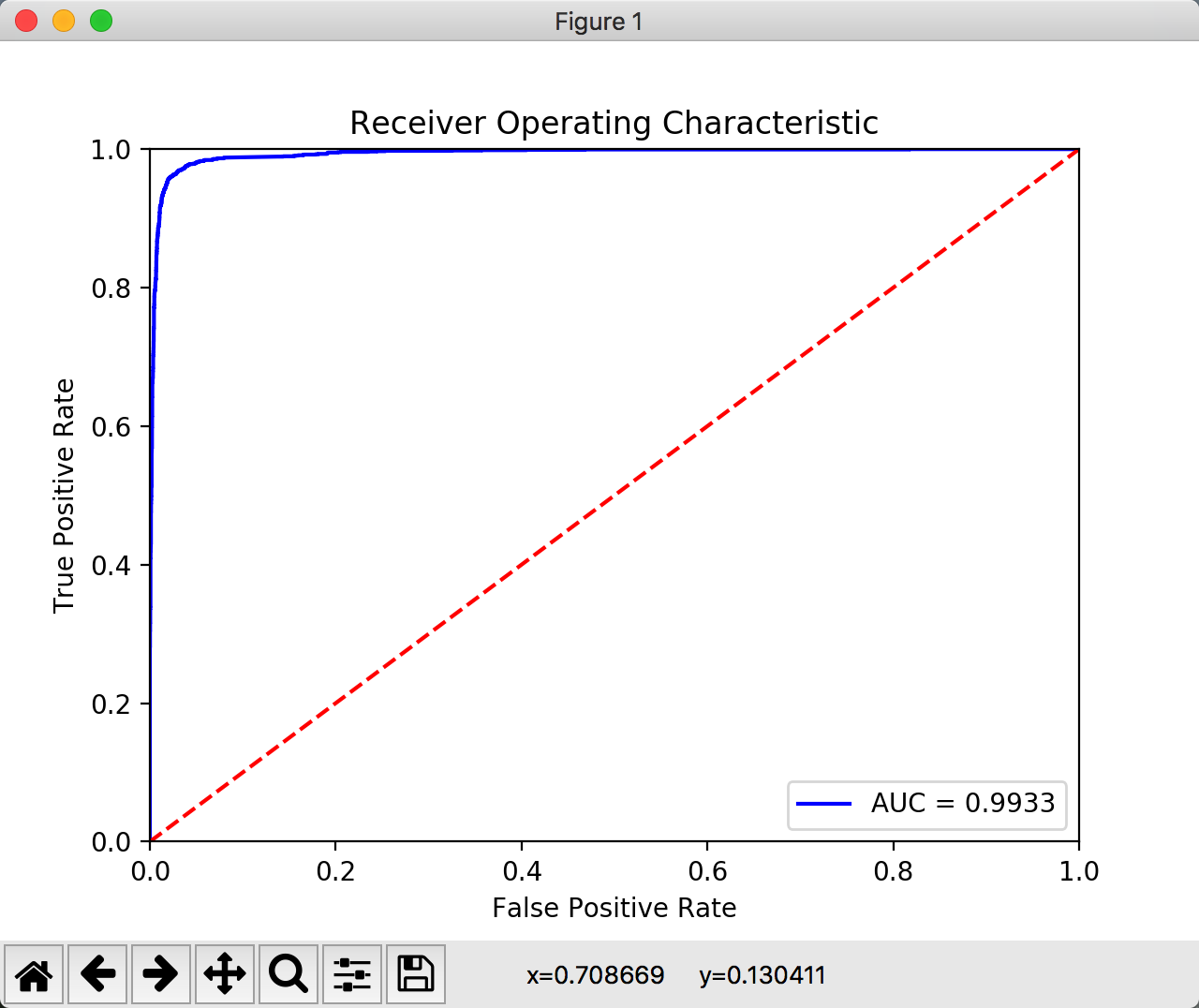
## 欠采样

由于正：负=1：4，所以我们删除负类中多余的“3”部分：

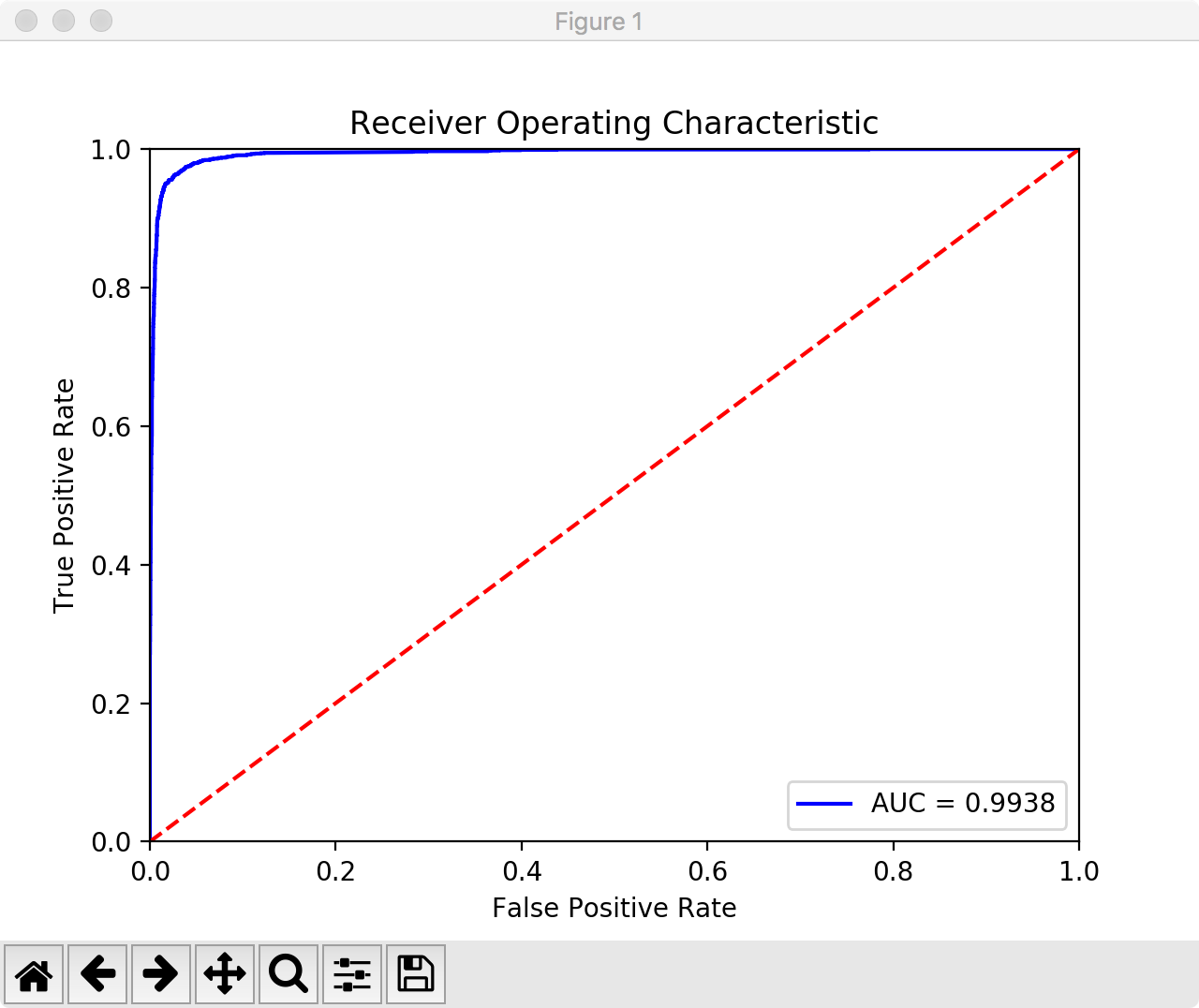


绘制出ROC图像：

由于之前拟合效果已经很好，所以这里效果并不明显。



## 调整阈值



以上为之前训练出模型的ROC图，是通过改变阈值来瞄点绘制出的。为了满足错误的正类比（FPR）尽量少的原则，我们可以在贴近y轴的地方选择某个点，然后选择这个点对应的阈值，这样优先保证正类不判错。