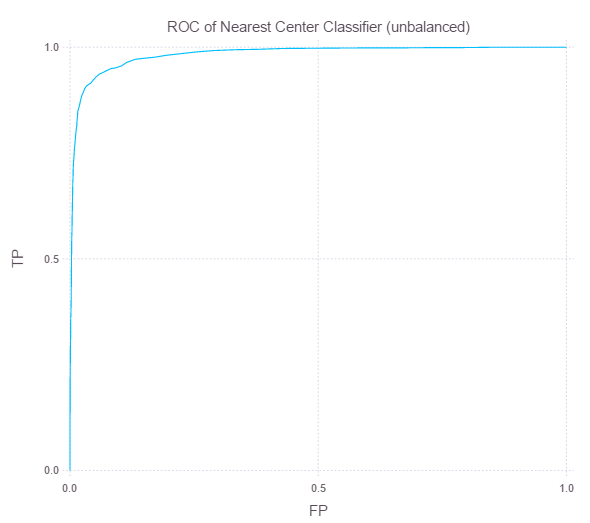
作业3.1 – 1133730117 - 张实唯

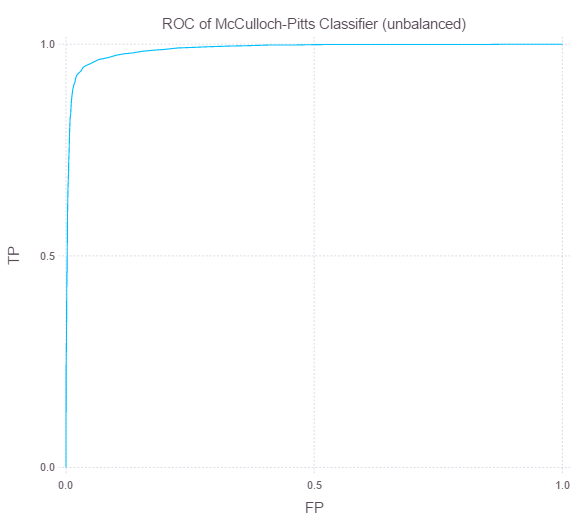
## 实验结果

1. **NearestCenterClassifier, 平衡前**



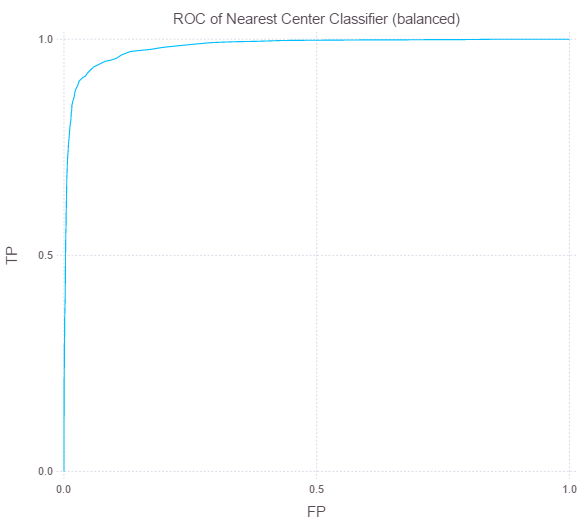
auc: 0.9831

1. **McCullochPittsClassifier, 平衡前**



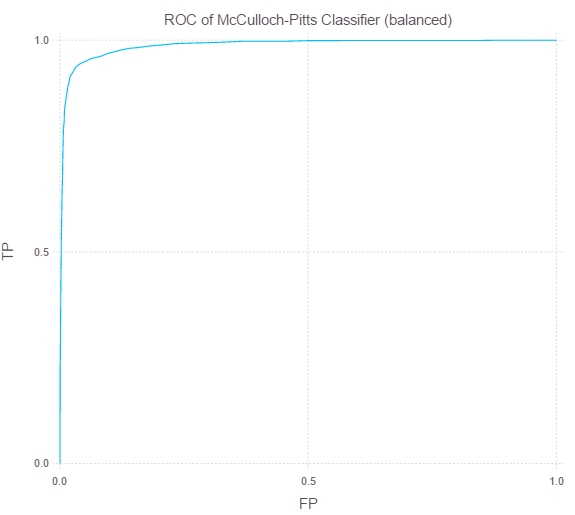
acu: 0.9878

1. **NearestCenterClassifier, 平衡后**



auc: 0.9830

1. **McCullochPittsClassifier, 平衡后**

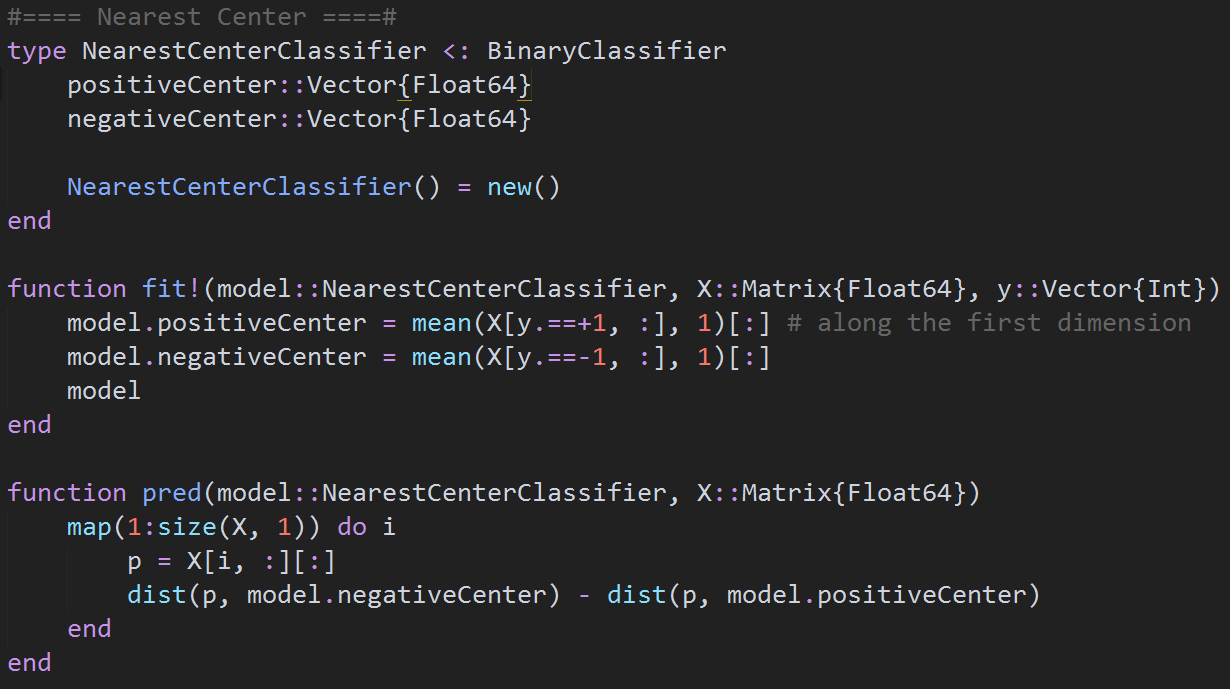


auc: 0.9885

## 代码解释

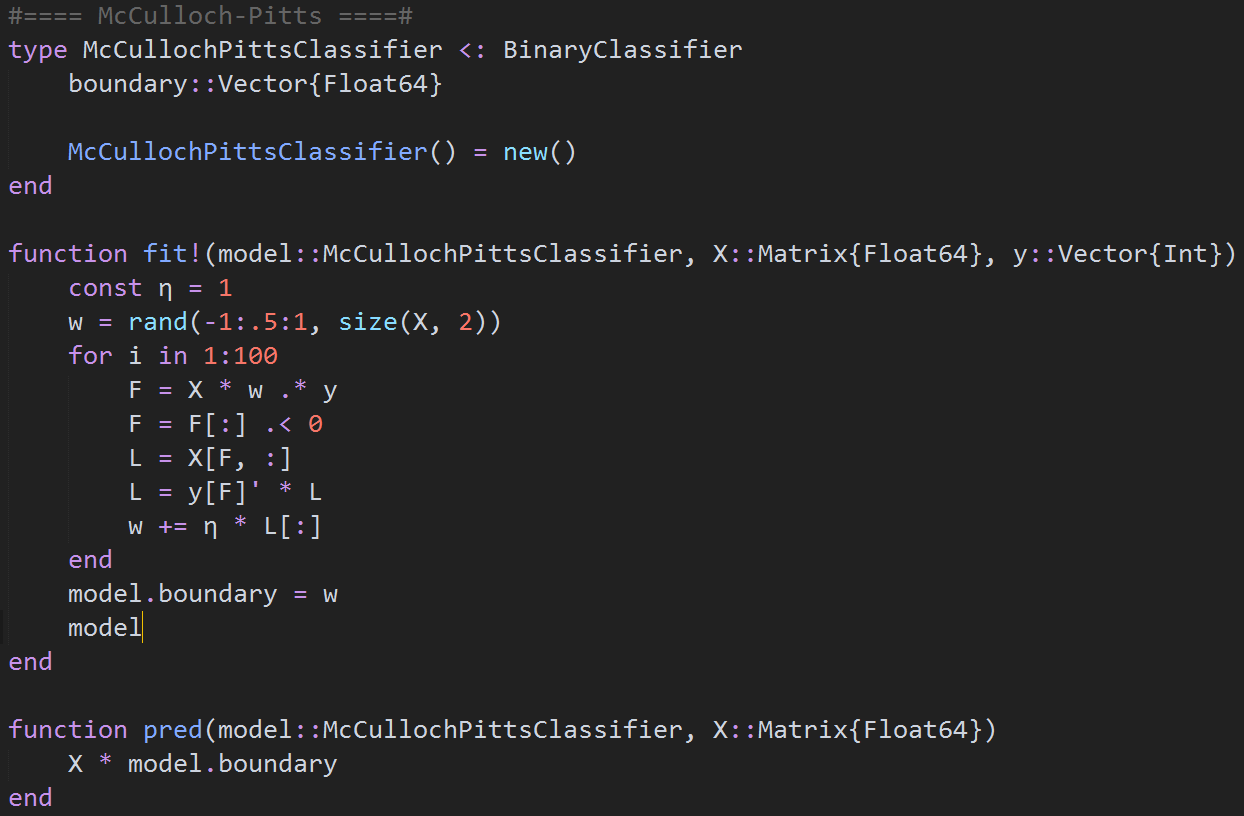
1. **NearestCenterClassifier模型**

* 训练方法： 筛选两个类别的所有记录，计算中心点(每个特征取平均)。
* 预测方法： 计算目标点到两个中心点的欧式距离，采用离负中心的距离减去正中心的距离再取sigmoid作为预测概率。



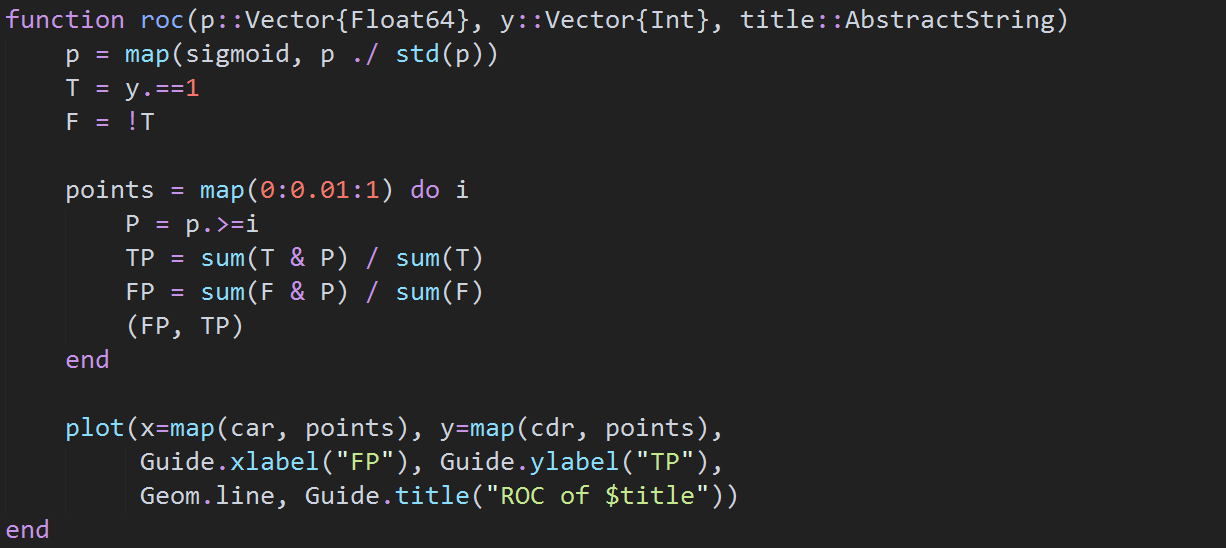
1. **McCullochPittsClassifier模型**

* 训练方法：
* 首先随机生成初始参数
* 每轮筛选出所有预测错误的记录，将其乘以预先定义的学习率再加和到一起
* 用上面得到的向量修正参数，再进行同样的步骤
* 预测方法： 将每条记录乘以参数向量即可。
* 每个过程都使用向量化的操作，将所有记录和他们的特征一起作为矩阵进行运算，这样一步就可以完成一轮所有记录的优化。



1. **ROC绘制方法**

计算TP和FP，等距离取阈值描点绘制曲线



auc的计算方法与此类似，在获得不同阈值下的点之后计算面积作为积分的估计值即可。

1. **平衡数据集方法**

计算正类样本缺少的数量，在已有的正类样本里随机过取样补齐即可。

