ROC 曲线的求解

接受者操作特性曲线（receiver operating characteristic curve，简称ROC曲线），又称为感受性曲线（sensitivity curve）。得此名的原因在于曲线上各点反映着相同的感受性，它们都是对同一信号刺激的反应，只不过是在几种不同的判定标准下所得的结果而已。

为了绘制 ROC 曲线，则分类器应该能输出连续的值，比如在逻辑回归分类器中，其以概率的形式输出，可以设定阈值大于 0.5 为正样本，否则为负样本。因此设置不同的阈值就可以得到不同的 ROC 曲线中的点，具体实现步骤如下：

(1)假定为正类定义了连续值输出，对检验记录按它们的输出值递增排序。

(2)选择秩最低的检验记录（即输出值最低的记录)，把选择的记录以及那些秩高于它的记录指派为正类。这种方法等价于把所有的检验实例都分为正类。因为所有的正检验实例都被正确分类，而所有的负测试实例都被误分，因此TPR=FPR=1。

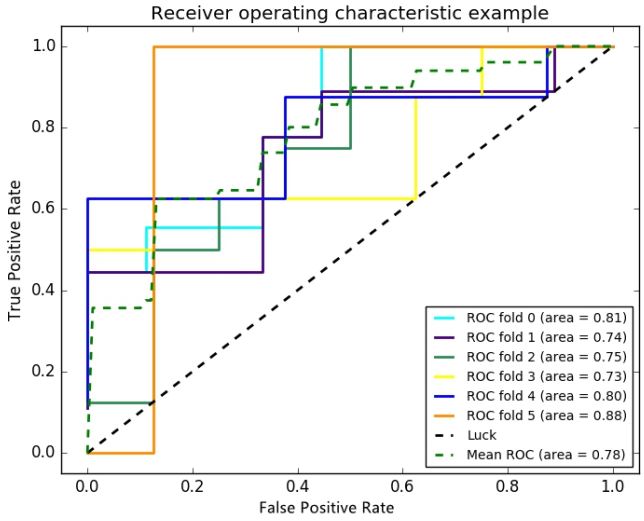
(3)从排序列表中选择下一个检验记录，把选择的记录以及那些秩高于它的记录指派为正类，而把那些秩低于它的记录指派为负类。通过考察前面选择的记录的实际类标号来更新 TP 和 FP 计数。如果前面选择的记录为正类，则 TP 计数减少而 FP 计数不变。如果前面选择的记录为负类，则 FP 计数减少而 TP 计数不变。

(4)重复步骤 3 并相应地更新 TP 和 FP 计数，直到最高秩的记录被选择。

(5)根据分类器的 FPR 画出 TPR 曲线。

代码为 sklearn 中的实现，具体代码请见 ROC 曲线的求解.py 文件

代码输出结果如下：



参考资料：

1.《机器学习》 [中]周志华

2.《数据挖掘导论》 [美]Pang-Ning Tan