

# 1、ROC曲线

ROC全称受试者工作特征，是用来研究学习器泛化性能的有力工具。ROC曲线横轴是假正利率FPR，纵轴是TPR，曲线的绘制过程是：根据学习器的预测结果对样例进行排序，按此顺序逐个把样例作为正例进行预测，计算出TPR和FPR，绘制ROC图。

## • ROC VS P-R

	ROC	PR
适用情况	测试样本正负分布均匀	测试样本正负分布不均匀（有明显差别）
对样本分布的敏感程度	不敏感	敏感
优点	能够衡量一个模型本身的预测能力	能够看出学习器随样本比例的变化的效果

「注」:先观察ROC曲线再观察PR曲线决定学习器的效果

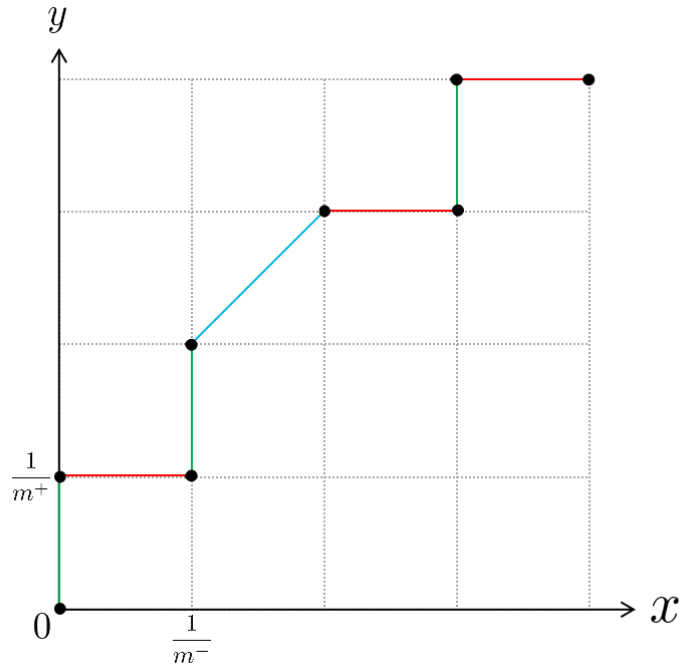
# 2、AUC=1- $\ell_{rank}$

## • ROC曲线绘制实例

已知一个学习器有以下预测结果：

s1,0.77,+	s2,0.62,-	s3,0.58,+	s4,0.47,+
s5,0.47,-	s6,0.33,-	s7,0.23,+	s8,0.15,-

ROC曲线横轴为 $FPR = \frac{FP}{FP+TN}$ ,纵轴为 $TPR = \frac{TP}{TP+FN}$ ，横轴纵轴的分母都是定值，分别为m-(反例个数),m<sup>+</sup>(正例个数),根据ROC的绘制过程可以看出每新增一个正例，ROC曲线水平上升一个单位，每新增一个反例，ROC曲线水平向右增加一个单位。最特殊的是上面例子中选定0.47作为正例阈值时，正例和反例各增加了一个，因此线段斜向上倾斜。



绘制过程如上图所示。

## • 证明 $AUC=1-\ell_{rank}$

现在已知排序损失定义为：

$$\ell_{rank} = \frac{1}{m^+m^-} \sum_{\mathbf{x}^+ \in D^+} \sum_{\mathbf{x}^- \in D^-} (\mathbb{I}(f(\mathbf{x}^+) < f(\mathbf{x}^-)) + \frac{1}{2} \mathbb{I}(f(\mathbf{x}^+) = f(\mathbf{x}^-)))$$

试证明： $AUC=1-\ell_{rank}$ ，也就是说明上述式子的含义是曲线上方的面积。

证明：

$$\begin{aligned} \ell_{rank} &= \frac{1}{m^+m^-} \sum_{\mathbf{x}^+ \in D^+} \sum_{\mathbf{x}^- \in D^-} ((f(\mathbf{x}^+) < f(\mathbf{x}^-)) + \frac{1}{2}(f(\mathbf{x}^+) = f(\mathbf{x}^-))) \\ &= \sum_{\mathbf{x}^+ \in D^+} \frac{1}{m^+} \left( \frac{1}{m^-} \sum_{\mathbf{x}^- \in D^-} (f(\mathbf{x}^+) < f(\mathbf{x}^-)) \right) + \frac{1}{2} \sum_{\mathbf{x}^+ \in D^+} \frac{1}{m^+} \left( \frac{1}{m^-} \sum_{\mathbf{x}^- \in D^-} (f(\mathbf{x}^+) = f(\mathbf{x}^-)) \right) \end{aligned}$$

- 上面式子第一项解释：

针对每个被选入当作正例的正点，在它被选入之前，那些预测值大于它的反点已被选入正例，也就是ROC曲线上已经向右平移 $\frac{1}{m^-} \sum_{\mathbf{x}^- \in D^-} (f(\mathbf{x}^+) < f(\mathbf{x}^-))$ 个单位，再乘于高 $\frac{1}{m^+}$ 便得到**每条绿色线段左边矩形的面积**。

- 上面式子第二项解释：

针对每个被选入当作正例的正点，当它选入时，可能会有一些反点和它预测值相同而被当作正例看待，因此ROC曲线上会同时向上、向右分别平移**1个单位**， $\sum_{\mathbf{x}^- \in D^-} (f(\mathbf{x}^+) = f(\mathbf{x}^-))$ **单位**，因此整体线段

会斜向上延伸，如上图蓝色线段所示，多的三角形面积计算公式便为：

$$\frac{1}{2} \sum_{\mathbf{x}^+ \in D^+} \frac{1}{m^+} \left( \frac{1}{m^-} \sum_{\mathbf{x}^- \in D^-} (f(\mathbf{x}^+) = f(\mathbf{x}^-)) \right)。$$

因此： $\ell_{rank}=1-AUC$

「参考」：

1. <https://www.jianshu.com/p/9d70c26b73a2>
2. <https://github.com/datawhalechina/pumpkin-book>