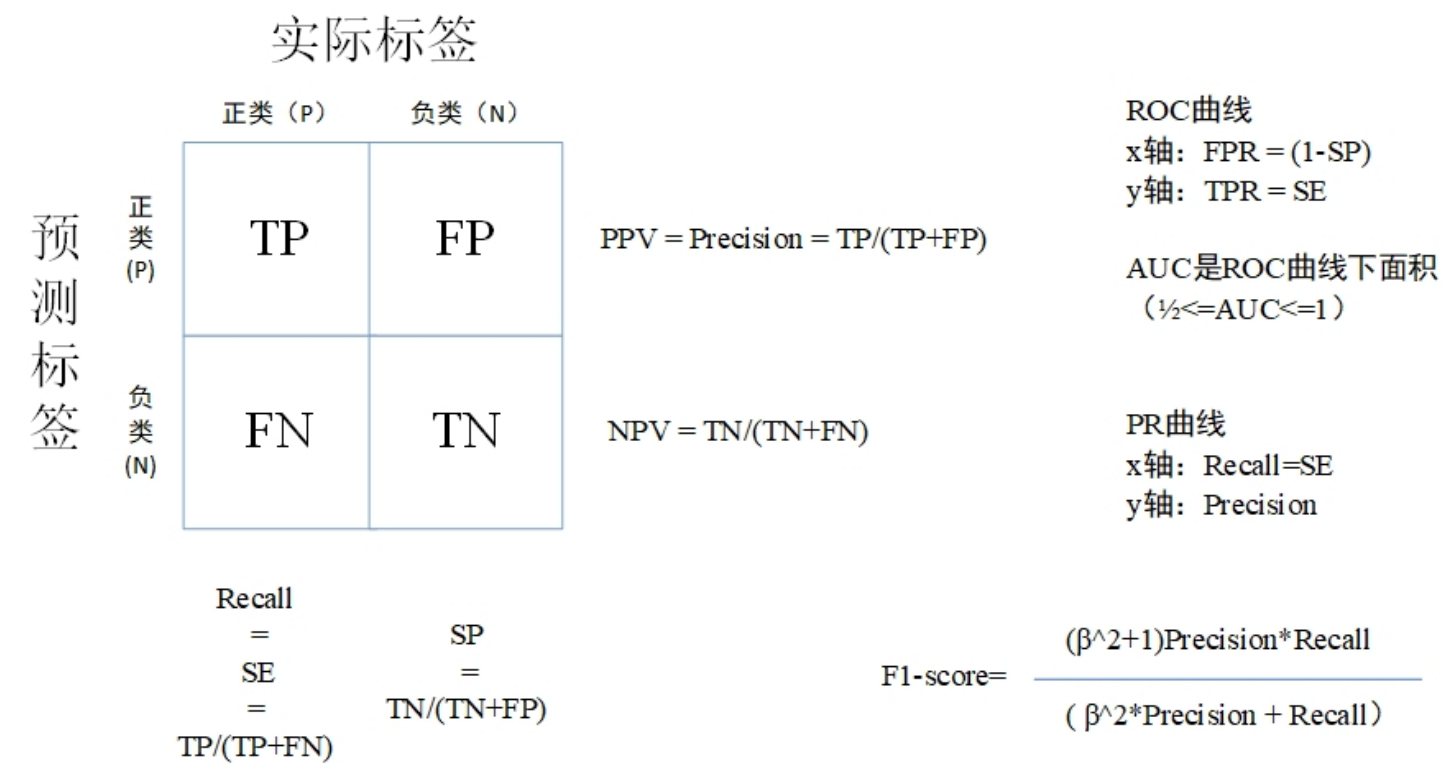


二分类的评价指标



单个指标

上图所示的多个评价指标一般用于二分类，对二分类来说，对于某些领域，如生物医学领域会区分，阳性样本(患病)、阴性样本（不患病）。

很多评价指标有多个名字，比如灵敏度(Sensitivity, SE)和召回率(Recall),两者都反应的是

$$SE = Recall = \frac{\text{预测正确的阳性样本}}{\text{所有的阳性样本}}$$

与SE对应的是特异性(Specificity, SP)，SP和SE很像，可以说是说是阴性样本的“SE”，

$$SP = \frac{\text{预测正确的阴性样本}}{\text{所有的阴性样本}}$$

与Recall对应的是精确度(Precision), 也叫阳性预测率(Positive Prediction Rate, PPV)，可以表述为

$$PPV = \frac{\text{预测正确的阳性样本}}{\text{所有的预测为阳性的样本}}$$

当然，对应PPV也有NPV，类似地，有阴性预测率(Negative Prediction Rate, NPV)

$$NPV = \frac{\text{预测正确的阴性样本}}{\text{所有的预测为阴性的样本}}$$

总结一下，观察混淆矩阵，可以发现(SE,SP,PPV,NPV)四兄弟是对混淆矩阵完备的描述，额外地，Recall = SE，Precision=PPV。（名字太多，吐槽一下）

ROC曲线

复合指标，接收器工作特性曲线(receiver operating characteristic curve, ROC)。

ROC空间里的单点，是给定分类模型且给定阈值后得出的。但同一个二元分类模型的阈值可能设定为高或低，每种阈值的设定会得出不同的(1-SP)和SE。将同一模型每个阈值的 (1-SP, SE) 座标都画在ROC空间里，就成为特定模型的ROC曲线。

1. 由于每个不同的分类器（诊断工具、侦测工具）有各自的测量标准和测量值的单位（标示为：“健康人 - 病人分布图”的横轴），所以不同分类器的“健康人 - 病人分布图”都长得不一样。
2. 比较不同分类器时，ROC曲线的实际形状，便视两个实际分布的重叠范围而定，没有规律可循。
3. 但在同一个分类器之内，阈值的不同设定对ROC曲线的影响，仍有一些规律可循：
 - 当阈值设定为最高时，亦即所有样本都被预测为阴性，没有样本被预测为阳性，此时(1-SP)为0，SE为0。必得出ROC座标系左下角的点 (0, 0)。
 - 当阈值设定为最低时，亦即所有样本都被预测为阳性，没有样本被预测为阴性，SE为100%，(1-SP)为100%。当阈值设定为最低时，必得出ROC座标系右上角的点 (1, 1)。
 - 阈值调低，ROC点 往右上（或右 / 或上）移动，或不动

ROC的作用

1. 选择最佳的信号侦测模型、舍弃次佳的模型。
2. 在同一模型中设定最佳阈值。

AUC为ROC曲线下面积，越大越好。

PR曲线

PR曲线就是Recall为x轴、Precision为y轴，绘制的曲线。类似ROC也是选取不同的阈值。样本不均衡，且正类很重要，PR表现的效果会更好。

F1-score

F1-score可以解释为Precision和Recall的加权平均值