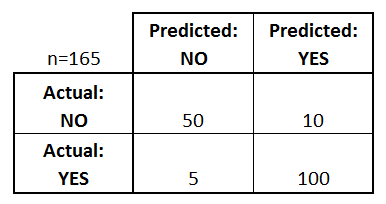
混淆矩阵术语的简单指南

混淆矩阵是一种表，通常用于描述分类模型（或“分类器”）在一组测试数据上的性能，其中真值是已知的。 混淆矩阵本身相对易于理解，但相关术语可能令人困惑。

我想为混淆矩阵术语创建一个“快速参考指南”，因为我找不到符合我要求的现有资源：紧凑的演示，使用数字而不是任意变量，并用公式和句子来解释。

让我们从二进制分类器的示例混淆矩阵开始（尽管它可以很容易地扩展到两个以上类的情况）：



我们可以从这个矩阵中学到什么？

有两种可能的预测类：“是”和“否”。例如，如果我们预测疾病的存在，“是”意味着他们患有疾病，“否”意味着他们没有疾病。

分类器总共进行了165次预测（例如，165名患者正在测试该疾病的存在）。

在这165例中，分类器预测“是”110次，“否”预测55次。

实际上，样本中的105名患者患有该疾病，而60名患者没有。

我们现在定义最基本的术语，即整数（不是费率）：

真阳性（TP）：这些是我们预测是（他们患有疾病）的病例，他们确实患有这种疾病。

真阴性（TN）：我们预测不，他们没有这种疾病。

误报（FP）：我们预测是，但他们实际上没有这种疾病。 （也称为“I型错误。”）

假阴性（FN）：我们预测不，但他们确实患有这种疾病。 （也称为“II型错误”。）

我已将这些术语添加到混淆矩阵中，并添加了行和列总计：



这是一个通常根据二元分类器的混淆矩阵计算的速率列表：

准确性：总体而言，分类器的使用频率是否正确？

（TP + TN）/总=（100 + 50）/ 165 = 0.91

错误分类率：总的来说，错误的频率是多少？

（FP + FN）/总=（10 + 5）/ 165 = 0.09

相当于1减去精度

也被称为“错误率”

真阳性率：当它实际上是肯定的时候，它有多少次预测是？

TP /实际是= 100/105 = 0.95

也被称为“灵敏度”或“召回”

误报率：当它实际上没有时，它多久预测一次？

FP /实际编号= 10/60 = 0.17

真正的负面率：当它实际上没有时，它多久预测一次？

TN /实际编号= 50/60 = 0.83

相当于1减去误报率

也被称为“特异性”

精度：当它预测是，它的频率是多少？

TP /预测是= 100/110 = 0.91

患病率：我们样本中的“是”条件实际发生的频率是多少？

实际是/总= 105/165 = 0.64

其他几个术语也值得一提：

空错误率：如果你总是预测多数班级，这就是你错误的频率。 （在我们的例子中，空错误率将是60/165 = 0.36，因为如果你总是预测是的话，那么对于60“无”的情况你只会出错。）这可以是一个有用的基线指标来比较你的分类器。但是，特定应用程序的最佳分类器有时会比空误差率具有更高的错误率，正如Accuracy Paradox所证明的那样。

Cohen的Kappa：这基本上衡量了分类器的表现与其完全按机会表现的程度相比。换句话说，如果准确度和空误差率之间存在很大差异，则模型将具有高Kappa分数。 （有关科恩卡帕的更多细节。）

F分数：这是真实阳性率（召回）和精确度的加权平均值。 （关于F分数的更多细节。）

ROC曲线：这是一个常用的图表，它总结了分类器在所有可能阈值上的性能。当您改变将观察值分配给给定类的阈值时，通过绘制真阳性率（y轴）与假阳性率（x轴）来生成它。 （关于ROC曲线的更多细节。）

最后，对于来自贝叶斯统计世界的人们，这里是应用预测建模的这些术语的快速摘要：

关于贝叶斯统计，灵敏度和特异性是条件概率，流行率是先验，正/负预测值是后验概率。