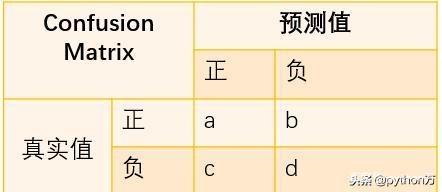
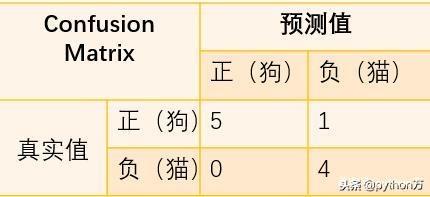
初步认识混淆矩阵

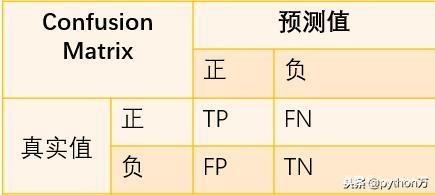
**混淆矩阵**是机器学习中总结分类模型预测结果的情形分析表，以矩阵形式将数据集中的记录按照真实的类别与分类模型预测的类别判断两个标准进行汇总。其中矩阵的行表示真实值，矩阵的列表示预测值，下面我们先以二分类为例，看下矩阵表现形式，如下：

二分类混淆矩阵

现在我们举个列子，并画出混淆矩阵表，假如宠物店有10只动物，其中6只狗，4只猫，现在有一个分类器将这10只动物进行分类，分类结果为5只狗，5只猫，那么我们画出分类结果混淆矩阵，并进行分析，如下（我们把狗作为正类）：

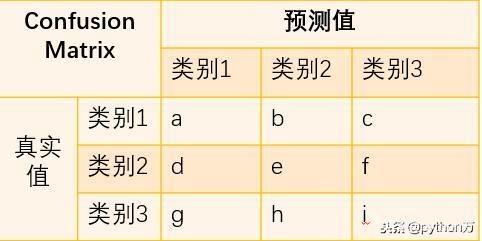
猫狗分类混淆矩阵

通过混淆矩阵我们可以轻松算的真实值狗的数量（行数量相加）为6=5+1，分类得到狗的数量（列数量相加）为5=5+0,真实猫的数量为4=0+4，分类得到猫的数量为5=1+4。同时，我们不难发现，对于二分类问题，矩阵中的4个元素刚好表示TP,TN,FP,TN这四个符号量，如下图：



那么对于二分类问题来说，精确率Precision=a/(a+c)=TP/(TP+FP),召回率recall=a/(a+b)=TP/(TP+FN),准确率accuracy=(a+d)/(a+b+c+d)=(TP+FN+FP+TN),可以看到准确率中的分子值就是矩阵对角线上的值。

刚才分析的是二分类问题，那么对于多分类问题，混淆矩阵表示的含义也基本相同，这里我们以三类问题为例，看看如何根据混淆矩阵计算各指标值。

多分类混淆矩阵

与二分类混淆矩阵一样，矩阵行数据相加是真实值类别数，列数据相加是分类后的类别数，那么相应的就有以下计算公式;

精确率\_类别1=a/(a+d+g)召回率\_类别1=a/(a+b+c)

**Python中的sklearn库提供了相应的方法来输出矩阵数据，非常方便，函数如下：**

**sklearn.metrics.confusion\_matrix(y\_true,y\_pred,labels=None,sample\_weight=None)**

**其中，y\_true:是样本真实分类结果，y\_pred 是样本预测分类结果 ，labels是所给出的类别，通过这个可对类别进行选择 ，sample\_weight 是样本权重。**