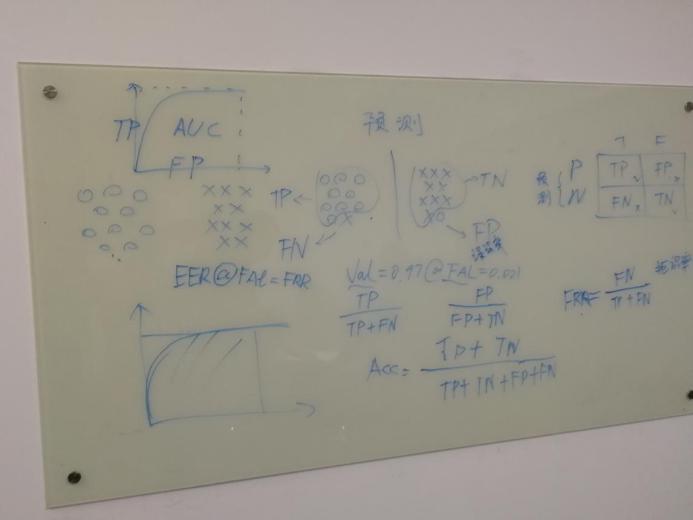
# 评价标准



围绕着混淆矩阵推导，True/False代表预测是否正确，Positive/Negative代表预测的结果。

True Positive(真正，TP)：将正类预测为正类数；

True Negative(真负，TN)：将负类预测为负类数；

False Positive(假正，FP)：将负类预测为正类数 → 误报 (Type I error).

False Negative(假负，FN)：将正类预测为负类数 →漏报 (Type II error).

具体例子:

病者癌症为良性，检测结果为恶性，则为False Positive，假阳性

病者癌症为良性，检测结果为良性，则为True Positive，真阳性

病者癌症为恶性，检测结果为恶性，则为True Negative，真阴性

病者癌症为恶性，检测结果为良性，则为False Negative，假阴性

Accuracy=准确率=(TP+TN)/(TP+TN+FP+FN)

P=precision=Precision Rate=查准率=精度=TP/(TP+FP)

R=recall=Recall Rate=查全率=召回率=真正例率（True Postive Rate，TPR）=validation rate(VR)=TP/(TP+FN)。代表分类器预测的正类中实际正实例占所有正实例的比例（将正例分对的概率）。

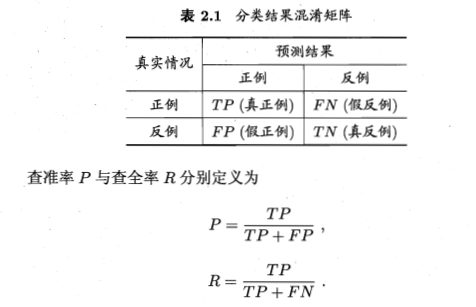
False Postive Rate=FPR=负正类率=FP/(FP+TN)。代表分类器预测的正类中实际负实例占所有负实例的比例(将负例错分为正例的概率)。

miss rate=漏检率=FN/(TP+FN)

误检率=FP/(TP+FP)

## PR曲线

一般用于二分类的模式识别中（如人脸验证）。是查准率（precision）和查全率（recall）的曲线。



1）P=precision=Precision Rate=查准率=精度=TP/(TP+FP)

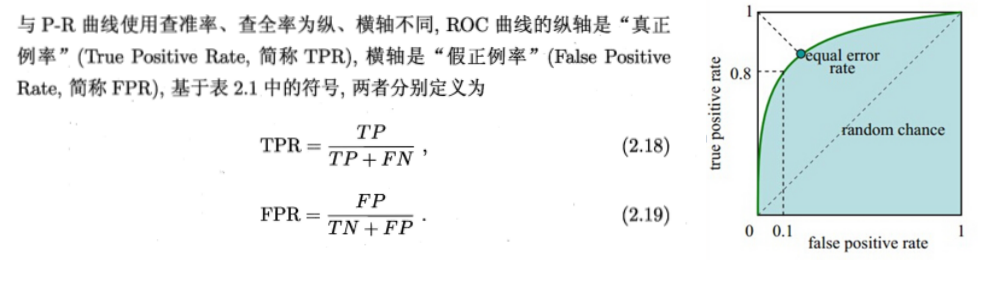
2）R=recall=Recall Rate=查全率=召回率=真正例率（True Postive Rate，TPR）=validation rate(VR)=TP/(TP+FN)。代表分类器预测的正类中实际正实例占所有正实例的比例（将正例分对的概率）。

3）F1值是Precision和Recall的调和平均。Recall和Precision大，F1值也大。公式是2/(F1)=1/P+1/R，即F1=(2\*precision\*recall)/(precision+recall)=2\*TP/(2\*TP+FP+FN)。

## ROC曲线

ROC曲线，即受试者工作特征曲线，Receiver Operating Characteristic curve。

ROC曲线越靠近左上角越好。



1）横轴：True Postive Rate=TPR=真正类率：TP/(TP+FN)，代表分类器预测的正类中实际正实例占所有正实例的比例（将正例分对的概率）。

2）纵轴：False Postive Rate=FPR=负正类率：FP/(FP+TN)，代表分类器预测的正类中实际负实例占所有负实例的比例(将负例错分为正例的概率)。

3）等错误率（EER Equal-Error Rate）是拒识率和误识率的一个平衡点，等错误率能够取到的值越低，表示算法的性能越好。

4）AUC（曲线下面积）：一种考虑到所有可能的分类阈值的评估标准。ROC 曲线下面积代表分类器随机预测真正类（Ture Positives）要比假正类（False Positives）概率大的确信度。

## PR曲线与ROC曲线的比较

当正负样本数量差不多时，PR和ROC表现都能正常反应模型的性能。

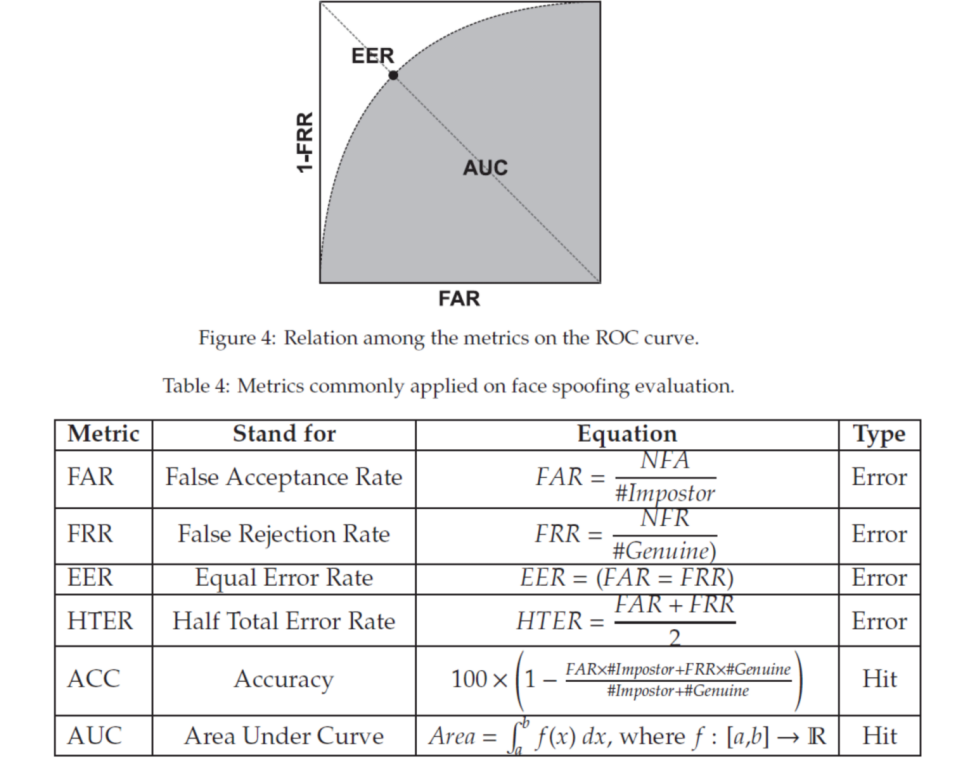
当正负样本数量差距很大时，即数据不平衡时，即负样本数量远多于正样本时，PR曲线会随着正负样本比例的变化，发生强烈的变化；而ROC曲线基本不变。故当数据不平衡时，应该使用PR曲线。

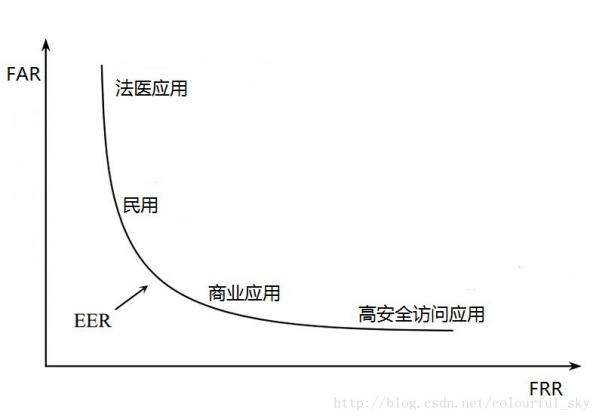
从公式上看，当负样本数量远多于正样本时，对于ROC曲线而言，使用了FPR（FPR=FP/N），N很大导致FPR基本不变；对于PR曲线而言，使用了TP/(TP+FN)，能很好地体现正负样本数量的差异。

实际应用中，正负样本的比例往往比较不平衡，所以ROC应用的场景会更多。

## FRR与FAR（另一种ROC）

ROC曲线（Receiver Operator characteristic Curve）是一种已经被广泛接受的系统匹配算法测试指标，它是匹配分数阈值、误识率（FRR）以及拒识率（FAR）之间的一种关系。它反映了识别算法在不同阈值上，拒识率和误识率的平衡关系。





1）FRR（False Reject Rate），即错误拒绝率，拒识率。指将身份相同的两张照片，判别为不同身份（将指定人员误作其它人员的概率）。数值越低越好。

2）FAR（False Accept Rate），B/(B+D)，即错误接受率，误识率。指将身份不同的两张照片，判别为相同身份的概率（将其他人误作指定人员的概率）。数值越低越好，比FRR更加重要。

这两个指标有明确的物理意义，FAR决定了系统的安全性，FRR决定了系统的易用程度，在实际中，FAR对应的风险远远高于FRR，因此，生物识别系统中，会将FAR设置为一个非常低的范围，如万分之一甚至百万分之一，在FAR固定的条件下，FRR低于5%，这样的系统才有实用价值。

　　计算机在判别时采用的阈值不同，这两个指标也不同。一般情况下，误识率FAR，随阈值的增大(放宽条件)而增大；拒识率FRR，随阈值的增大而减小。因此，可以采用等错误率(Equal Error Rate，ERR)作为性能指标，它表示使这FAR和FRR两个指标相等时的FAR（或FRR）。

例如：腾讯的人脸识别算法在实际互联网金融场景测试，自拍-身份证缩略图对比达到0.01%错误率，通过率95%速度指标，支持最小人脸尺寸64x64，人脸特征尺寸1-2KB，1：1的人脸对为500ms。

# 参考资料

https://blog.csdn.net/aws3217150/article/details/50479457