

机器学习中的评估指标

Classification Problem 分类问题

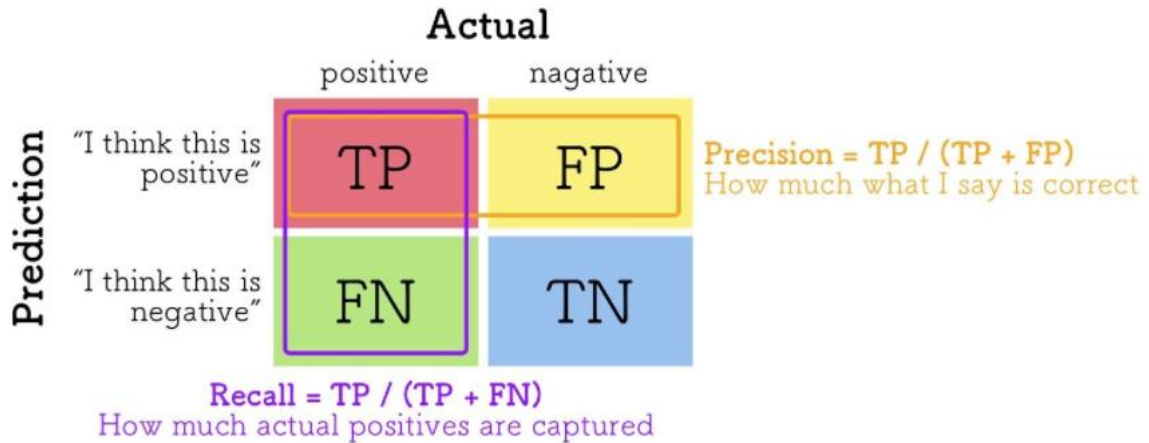
		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

TP (True Positive)	模型给出的答案是 Positive，正确的答案也是 Positive。 Example: 模型说这个人的病，确实这个人是有病
TN (True Negative)	模型给出的答案是 Negative，正确的答案也是 Negative。 Example: 模型说这个人没病，确实这个人没病
FP (False Positive)	模型给出的答案是 Positive，正确的答案也是 Negative。 Example: 模型说这个人有病，其实这个人没病
FN (False Negative)	模型给出的答案是 Negative，正确的答案也是 Positive。 Example: 模型说这个人没病，其实这个人有病

Accuracy 准确率

模型给出的所有预测里， 模型答对了多少。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$



### Precision 精确率 (Positive Predictive Value)

只是针对预测正确的正样本而不是所有预测正确的样本。正确预测的正例数 / 预测正例总数

$$Precision = \frac{TP}{TP + FP}$$

### Recall 召回率 (Sensitivity, Hit Rate or True Positive Rate)

表现在实际正样本中，分类器能预测出多少，即正确预测的正例数 / 实际正例总数

$$Recall = \frac{TP}{TP + FN}$$

### F1 Score

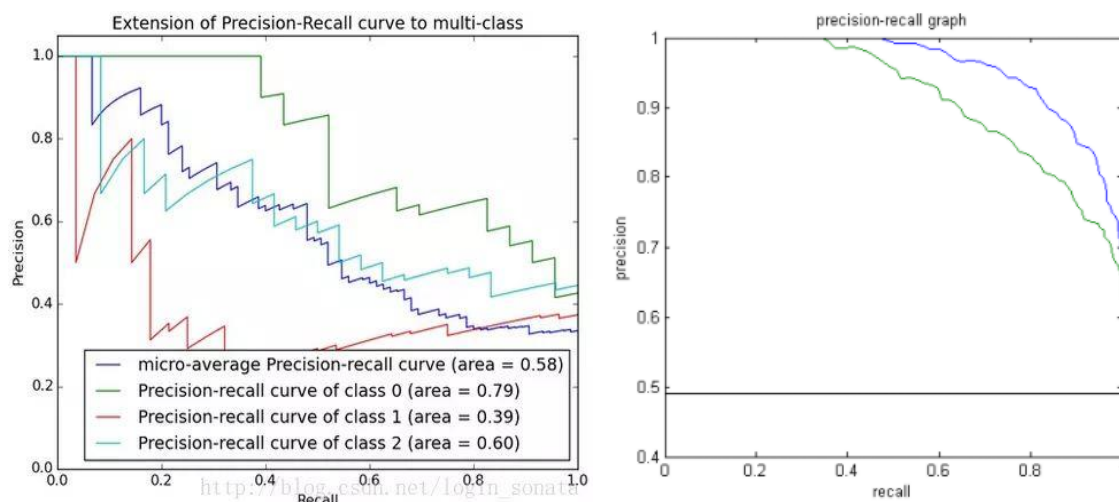
Precision 和 Recall 的调和平均值，更接近于两者中较小的那个。用于判断模型是好还是不好。

$$F1 = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

TPR (True Positive Rate) 真正率	$Recall = Sensitivity = TPR = \frac{TP}{TP + FN}$ 多少正确的 Positive 被预测出来
TNR (True Negative Rate) 真负率	$Specificity = TNR = \frac{TN}{TN + FP}$ 多少正确的 Negative 被预测出来
FPR (False Positive Rate) 假正率	$FPR = 1 - Specificity = \frac{FP}{TN + FP}$ 应该要是 Negative 的，模型预测了多少%是 Positive
FNR (False Negative Rate) 假负率	$FNR = \frac{FN}{TP + FN}$ 应该要是 Positive 的，模型预测了多少%是 Negative

## PR (Precision-Recall) Curve

PR Curve with Precision values on the y-axis and Recall values on the x-axis.

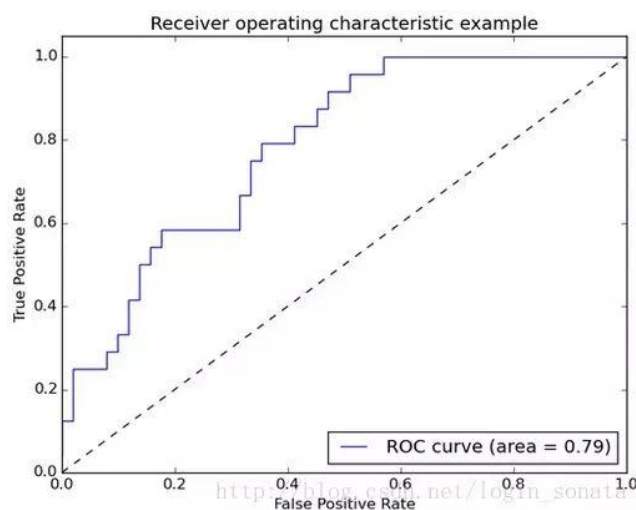


显然精确率(Precision) 和召回率(Recall) 都是越大越好，即在**上方的曲线比下方的曲线好**。当精确率(Precision) 和召回率(Recall)接近时 F1 Score 最大，因此也可以直接找 F1score 的最大点 (只需要画直线  $y=x$ )

Most ML Algorithms often involve a trade-off between Precision and Recall. A good PR Curve has greater AUC (Area Under Curve).

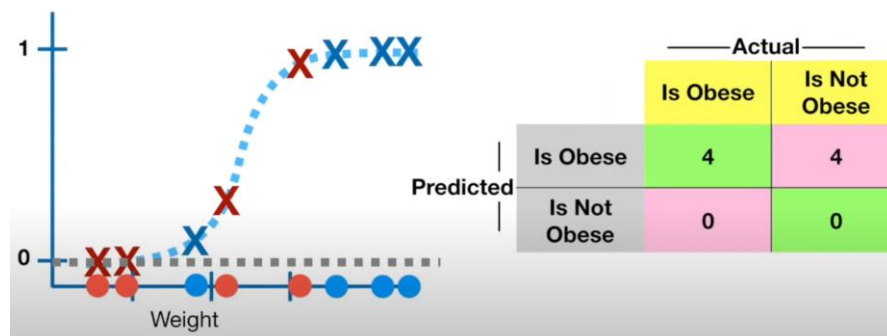
## ROC (Receiver Operating Characteristic) Curve

ROC Curve with True Positive Rate (多少正确的 Positive 被预测出来) on the y-axis and False Positive Rate (应该要是 Negative 的，模型预测了多少%是 Positive) on the x-axis

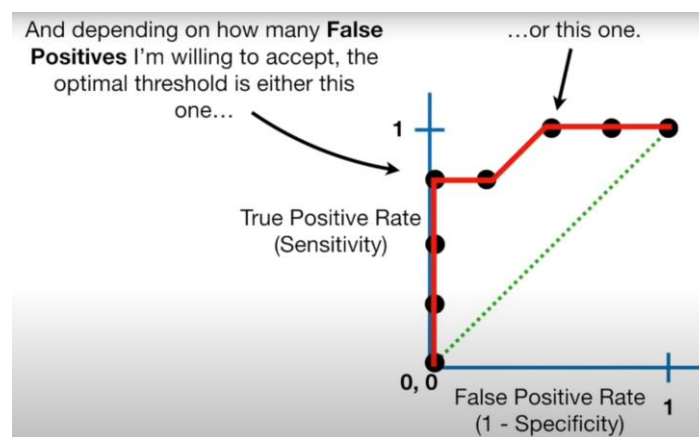


ROC Curves make it easy to identify the best threshold for making a decision. AUC help to decide which categorization method is better.

设置不同的 threshold 然后根据模型的 prediction 来算出 TPR 和 FPR 然后来 Plot ROC Curve. Threshold 越低, FPR 就会越高因为大概率都会识别成 Positive。Y-axis 越高越好, X-axis 越低越好。Plot 完 ROC 需要有 trade-off 来选择一个适合的 threshold.



蓝色是 Positive, 红色是 Negative。不同的 threshold, TPR 和 FPR 的概率不同。上图的 threshold 为 0, 所有的 prediction 都为 Positive。



图上 2 个点都是不错的选择, 根据需求来决定 threshold。是要更高的准确率还是需要保存更多的 Positive。

## AUC (Area Under Curve)

AUC 越大, 模型越好。当  $AUC = 0.5$ , 则表示为随机猜测, 没有预测价值。当  $AUC > 0.5$ , 优于随机猜测。当  $AUC < 0.5$ , 比随机猜测还差, 但只要反预测便可优于随机猜测。AUC 的物理意义为, 任取一对 (正、负) 样本, Positive 的 Confidence 大于 Negative 的 Confidence 的 Probability。

## 代价敏感

不同的错误具有不同的权重, 将 Positive 预测为 Negative 的后果与将 Negative 预测为 Positive 的后果不一致。将生病的情况预测为健康的危害大于将健康的情况预测为生病, 体检的时候如果把正确率直接设置为 100%, 则与实际情况很接近, 但这并不是我们想要的, 我们需要的是找出其中的病人, 因此需要将这两种情况赋予不同的权重。