

机器学习的效果评价指标

原创 老居 老居搞机 3月11日

在有监督机器学习中, 会采用不同的算法, 不同的参数来训练出多个模型, 如何评价这些训练出来的模型好坏呢? 今天我们看看常用的模型评价指标有哪些

一、分类

- 对于二分类模型我们可以看: 精确率、准确率、召回率、F1值、ROC-AUC、PRC、混淆矩阵
- 假如有一批二分类样本结果P代表二分类结果为positive(判断属于该分类), N代表结果是Negative(判断不属于该分类). 而True(T)与False(F)分别代表了预测结果与实际结果是否一致. 因此**矩阵对角线上都是预测正确的结果**
- 例如:

样本	预测为正例	预测为负例
真实为正例	TP	FN
真实为负例	FP	TN

- TP: 为真实是正样本预测为正
- FN: 为真实是正样本预测为负
- FP: 为真实是负样本预测为负
- TN: 为真实是负样本预测为负

1.精确率(Precision)

- 精确率 = $TP / (TP + FP)$
- 精确率反映的是模型判定的精确度, 预测为预测正样本里面有多少是真实的正样本

2.召回率(Recoll)

- 召回率 = $TP / (TP + FN)$
- 召回率反映的是模型的查全率, 意味着是真实的正样本有多少被找出来了

3.准确率(Accuracy)

- 准确率 = $(TP + TN) / (TP + TN + FP + FN)$

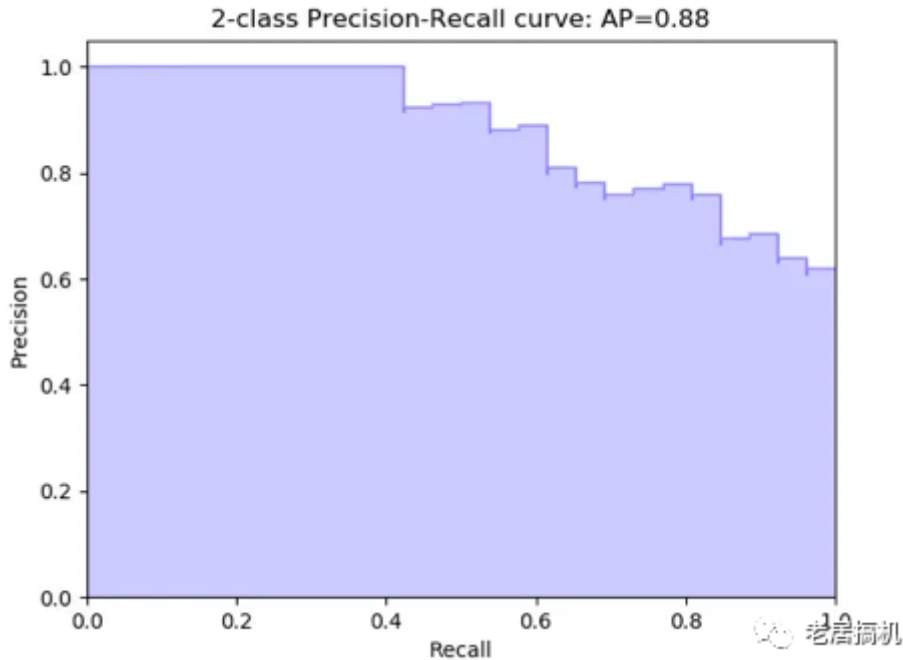
4.F1值

- $F1 = 2TP / (2TP + FP + FN)$

- F1值为精确率和召回率的调和均值, 反映的是模型的综合精度/召回能力

5.PR(Precision Recall)曲线

- Precision和Recall存在互相冲突的影响(追求精度越高, 召回会降低, 反过来追求更高的召回, 精度也会相应降低), 一般我们模型刚上线的时候为了不造成负面的问题影响后续的产品推广, 会追求更高的精度(查准), 而产品成熟以后会追求更高的召回, 以提高经济效益(查全)
- 可以利用PR曲线找到满足精度和召回要求的平衡点

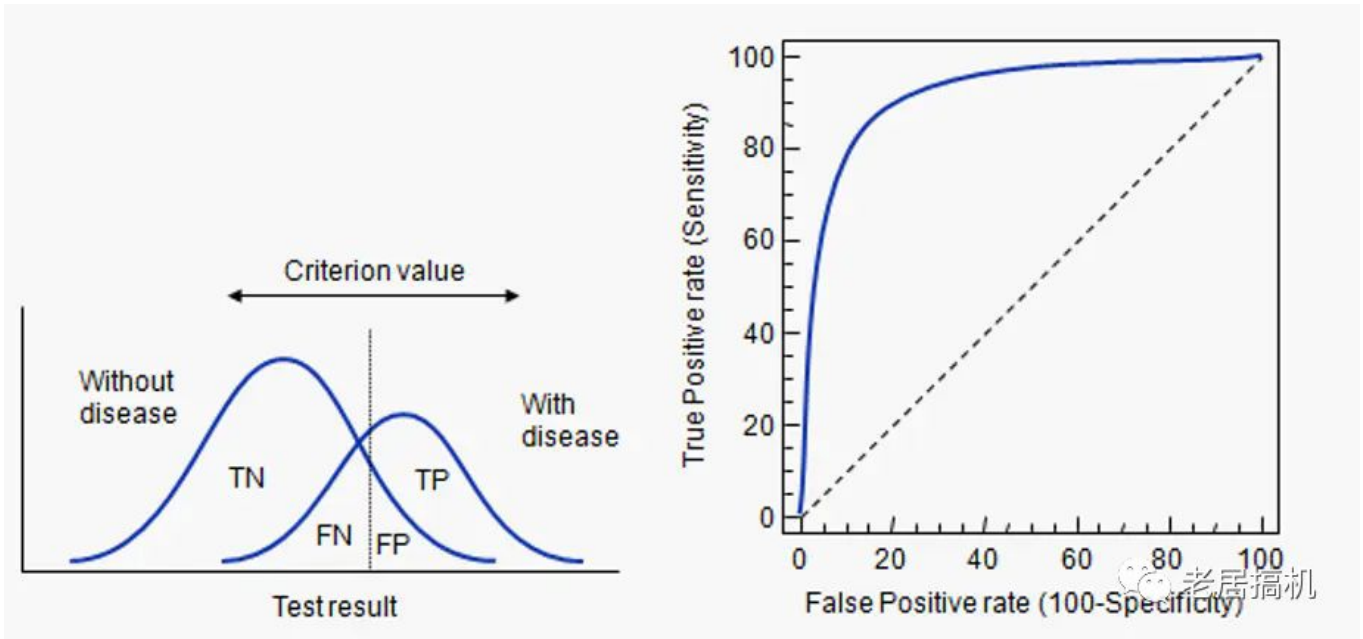


6.ROC曲线(Receiver Operating Characteristic Curve)

- ROC曲线是根据TPR和FPR画出来的曲线

- 1 True Positive Rate: $TPR = TP / (TP + FN)$ --> 将正例分对的概率
- 2 Fales Positive Rate: $FPR = FP / (FP + TN)$ --> 将负例错分为正例的概率

- 在 ROC 空间中, 每个点的横坐标是 FPR, 纵坐标是 TPR, 这也就描绘了分类器在 TP (真正率) 和 FP (假正率) 间的 权衡



7.AUC(Area Under Curve)

- AUC定义为ROC曲线下的面积(积分), 这个面积越大, 用来衡量
- AUC的计算: 正样本数 m , 负样本数 n , 组合成 $m*n$ 的组合, 组合中正样本预测为正样本score大于负样本预测为正样本score总数 / $(m*n)$ 组合数

8.混淆矩阵

- 混淆矩阵(Confusion Matrix)又称为错误矩阵, 通过它可以可视化算法的效果, 它的每一行是样本的真实分类, 每一列是样本预测的分类, 反映了分类结果的混淆程度

样本	预测分类1	预测分类2	预测分类3	预测分类4
真实分类1	80	10	8	2
真实分类2	7	70	20	3
真实分类3	9	11	80	0
真实分类4	5	4	1	90

二、回归

- 对于回归模型我们可以看: 平均绝对误差、均方根误差等相关的指标

1.平均绝对误差

- 平均绝对误差MAE(Mean Absolute Error) 又被称为L1范数损失(l1-norm loss)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|$$

2.均方根误差

- 均方根误差RMSE(Root Mean Squared Error) 又被称为L2范数损失(l2-norm loss)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$$

参考:

[1]李航.统计学习方法

[2]<https://www.jianshu.com/p/bf8927202d39>

[3]<https://www.jianshu.com/p/b960305718f1>

[4]<https://www.cnblogs.com/guoyaohua/p/classification-metrics.html>

长按以下二维码, 关注本公众号

