

评价指标 已购

来自【机器学习面试题汇总与解析（蒋豆芽面试题总结）】 | 56 浏览 | 0 回复 | 2021-05-20



蒋豆芽

[+关注](#)

机器学习面试题汇总与解析——评价指标

1. 说说机器学习评价指标☆☆☆☆☆
2. AUC是什么？AUC是否对正负样本比例敏感？☆☆☆☆☆
3. 分类模型如何评价☆☆☆☆☆
4. 准确率与精准率的区别☆☆☆☆☆
5. AUC的意义和两种计算方法☆☆☆☆☆
6. 讲讲分类，回归，推荐，搜索的评价指标☆☆☆☆☆
7. AB test的原理☆☆☆☆☆

- =====
- 本专栏适合于Python已经入门的学生或人士，有一定的编程基础。
 - 本专栏适合于算法工程师、机器学习、图像处理求职的学生或人士。
 - 本专栏针对面试题答案进行了优化，尽量做到好记、言简意赅。这才是一份面试题总结的正确打开方式。这样才方便背诵
 - 如专栏内容有错漏，欢迎在评论区指出或私聊我更改，一起学习，共同进步。
 - 相信大家都有着高尚的灵魂，请尊重我的知识产权，未经允许严禁各类机构和个人转载、传阅本专栏的内容。
- =====

介绍一下机器学习中的评价指标

先说一下含义：TP、FP、FN、TN分别为真阳性、假阳性和假阴性、真阴性。

简单来说，就是：

- 真阳性 (True Positive, TP)：标签为正类，预测为正类；
- 假阳性 (False Positive, FP)：标签为负类，预测为正类；
- 假阴性 (False Negative, FN)：标签为正类，预测为负类；
- 真阴性 (True Negative, TN)：标签为负类，预测为负类；

用混淆矩阵表示为：

蒋豆芽

真实情况	预测结果	
	正例	反例
正例	TP	FN
反例	FP	TN

牛客@蒋豆芽

1. 准确率 (Accuracy)

定义：

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

(.)

即所有分类正确的样本占全部样本的比例

2. 精确率/查准率 (Precision)

定义：

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

(.)

即预测是正例的结果中，确实是正例的比例。Precision同样是衡量误检

3. 查全率 (Recall)

定义：

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

(.)

即所有正例的样本中，被找出的比例。Recall同样是衡量漏检

Precision和Recall的应用场景：

地震的预测 对于地震的预测，我们希望的是Recall非常高，也就是说每次地震我们都希望预测出来。这个时候我们可以牺牲Precision。情愿发出1000次警报，把10次地震都预测正确了；也不要预测100次对了8次漏了两次。

“宁错拿一万，不放过一个”，分类阈值较低

嫌疑人定罪基于不错怪一个好人的原则，对于嫌疑人的定罪我们希望是非常准确的。即使有时候放过了一些罪犯，但也是值得的。因此我们希望有较高的Precision值，可以合理地牺牲Recall。

“宁放过一万，不错拿一个”，“疑罪从无”，分类阈值较高

4. F1-score

定义:

$$F1 - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (.)$$

衡量Precision 和 Recall 之间的联系。

5. ROC 和 AUC

AUC是一种模型分类指标，且仅仅是二分类模型的评价指标。**AUC**是Area Under Curve的简称，那么Curve就是 **ROC** (Receiver Operating Characteristic)，翻译为"接受者操作特性曲线"。也就是说**ROC**是一条曲线，**AUC**是一个面积值。

ROC曲线为 FPR 与 TPR 之间的关系曲线，这个组合以 FPR 对 TPR，即是以代价 (costs) 对收益 (benefits)，显然收益越高，代价越低，模型的性能就越好。

- x 轴为**假阳性率 (FPR)**：在所有的负样本中，分类器预测错误的比例

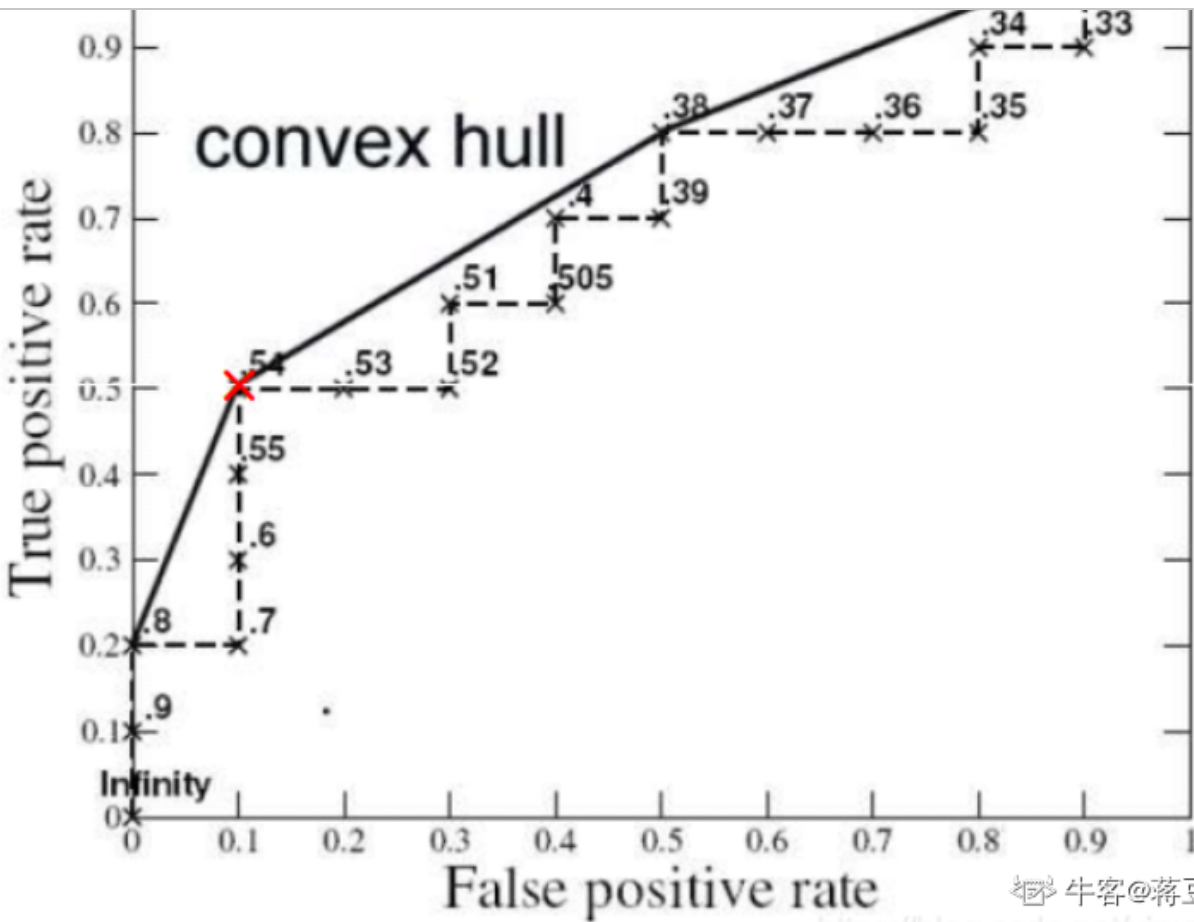
$$FPR = \frac{FP}{FP + TN} \quad (.)$$

- y 轴为**真阳性率 (TPR)**：在所有的正样本中，分类器预测正确的比例（等于Recall）

$$TPR = \frac{TP}{TP + FN} \quad (.)$$

那么每一个分类阈值都可以得到一组(FPR,TPR)，以此画出ROC曲线：

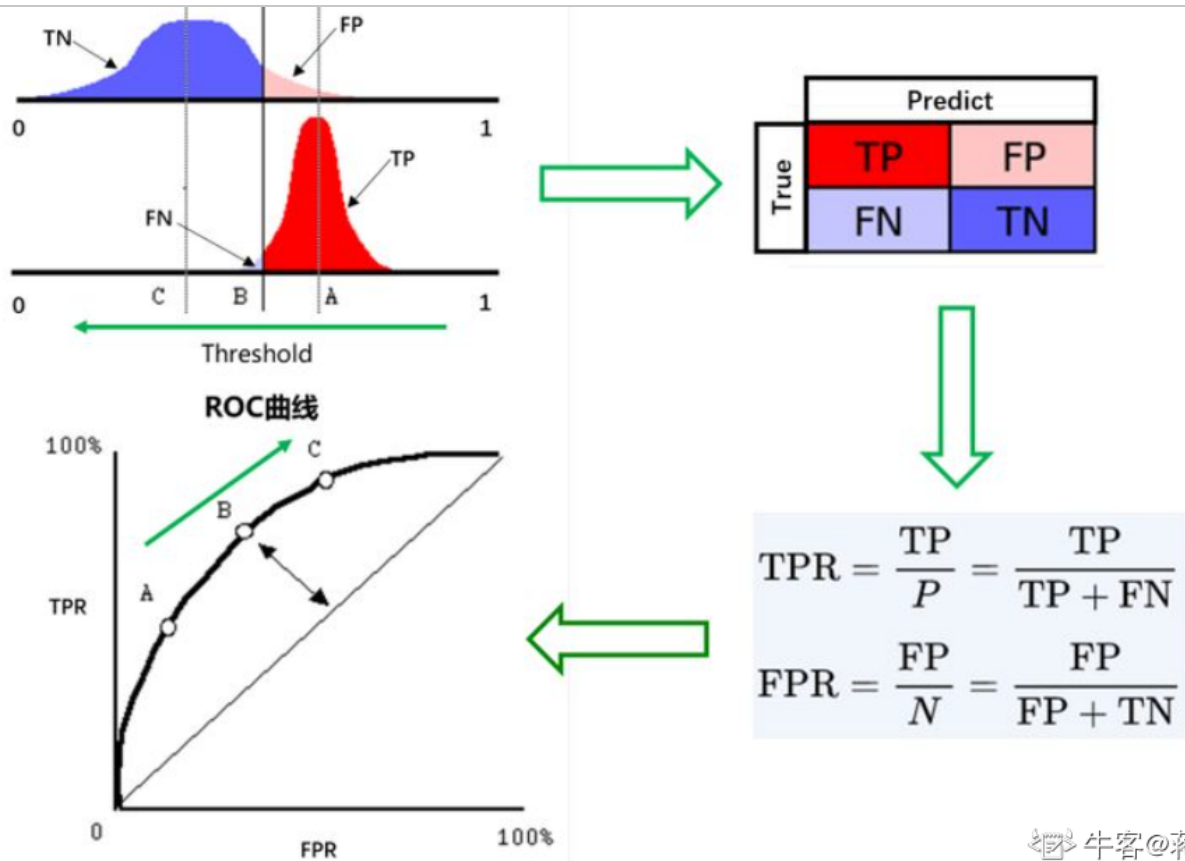
☰ 蒋豆芽



牛客@蒋豆芽

整个流程如下：

蒋豆芽



牛客@蒋豆芽

AUC定义： AUC 值为 ROC 曲线所覆盖的区域面积，显然，AUC越大，分类器分类效果越好。

- AUC = 1，是完美分类器。
- $0.5 < AUC < 1$ ，优于随机猜测。有预测价值。
- AUC = 0.5，跟随机猜测一样（例：丢铜板），没有预测价值。
- AUC < 0.5，比随机猜测还差；但只要总是反预测而行，就优于随机猜测。

AUC=1

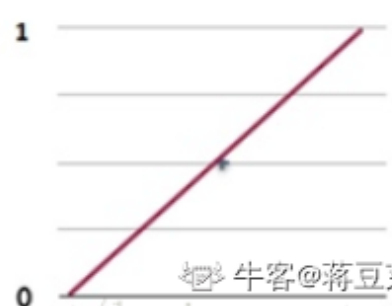
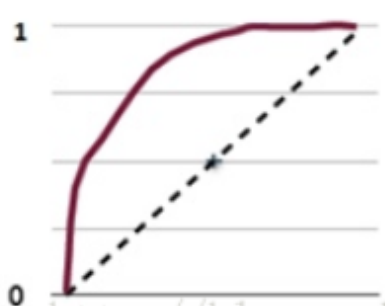
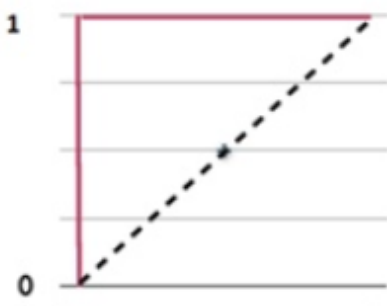
AUC=0,8

AUC=0,5

+ valor diagnóstico perfecto

+ valor diagnóstico

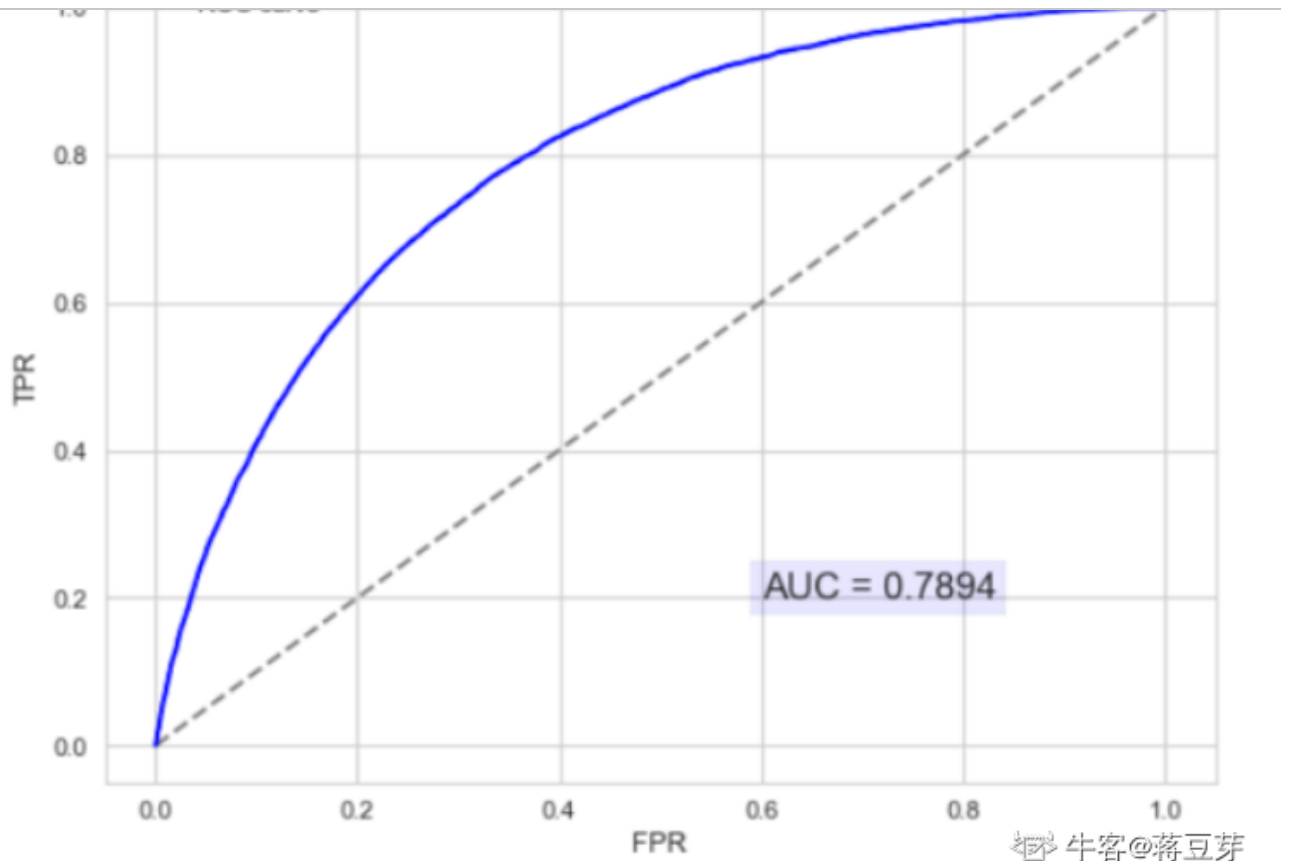
+ sin valor diagnóstico



牛客@蒋豆芽

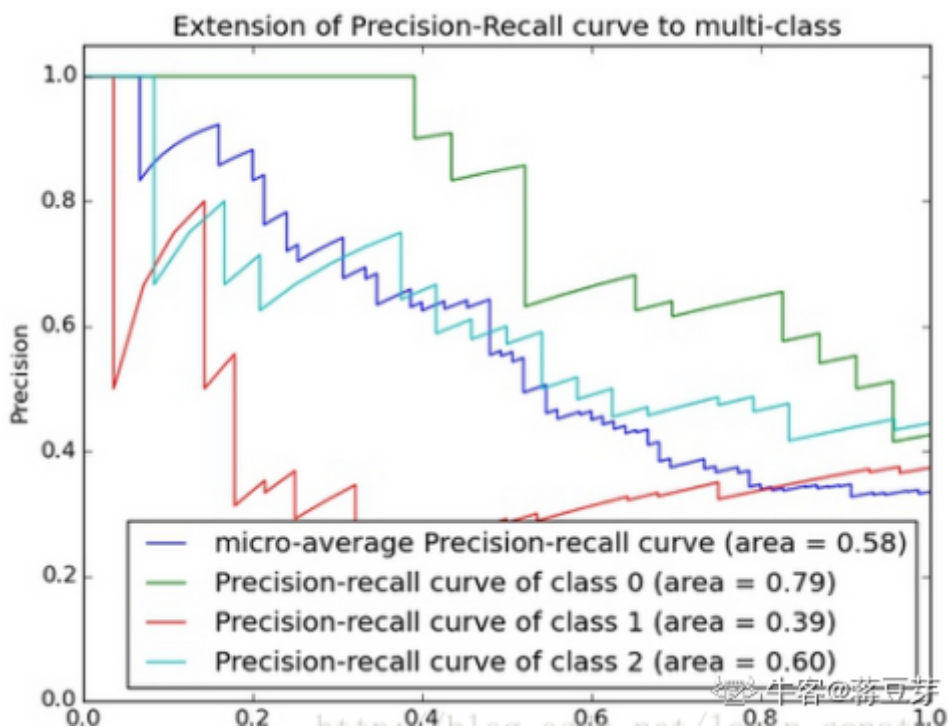
以下为ROC曲线和AUC值的实例：

蒋豆芽



6. Precision-Recall曲线

PR曲线的横坐标是**精确率P**，纵坐标是**召回率R**。评价标准和ROC一样，先看平滑不平滑（蓝线明显好些）。一般来说，在同一测试集，上面的比下面的好（绿线比红线好）。当P和R的值接近时，F1值最大，此时画连接(0,0)和(1,1)的线，线和PRC重合的地方的F1是这条线最大的F1（光滑的情况下），此时的F1对于PRC就好像AUC对于ROC一样。一个数字比一条线更方便调型。



此时曲线上的点就对应**F1**。**P-R曲线**同样可以用AUC衡量，**AUC**大的曲线越好。

8. AP和mAP

AP (average Precision) 和**mAP** (mean average Precision) 常用于目标检测任务中。**AP**就是每一类的Precision的平均值。而**mAP**是所有类的AP的平均值。

1. 说说机器学习评价指标☆☆☆☆☆

参考回答

答案参考上面。

答案解析

无。

类似的问题还有：

2. AUC是什么？AUC是否对正负样本比例敏感？☆☆☆☆☆

参考回答

AUC定义：**AUC** 值为 **ROC** 曲线所覆盖的区域面积，显然，AUC越大，分类器分类效果越好。

AUC还有另一个意义：分别随机从正负样本集中抽取一个正样本，一个负样本，正样本的预测值大于负样本的概率。

AUC对正负样本比例不敏感

答案解析

无。

3. 分类模型如何评价☆☆☆☆☆

参考回答

有几个指标：

1. **Accuracy**：即所有分类正确的样本占全部样本的比例
2. **Precision**：查准率。
3. **Recall**：查全率
4. **F1-score**：衡量Precision和Recall之间的联系
5. **AUC** 值为 ROC 曲线所覆盖的区域面积，显然，AUC越大，分类器分类效果越好。

蒋豆芽

无。

4. 准确率与精准率的区别 ☆ ☆ ☆ ☆ ☆

参考回答

1. 指标不一样：准确率是**Accuracy**；精准率是**Precision**
2. 定义不一样：准确率是所有分类正确的样本占全部样本的比例；精准率是预测是正例的结果中，确实是正例的比例。
3. 公式不一样：

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (.)$$

$$Precision = \frac{TP}{TP + FP} \quad (.)$$

答案解析

无。

5. AUC的意义和两种计算方法 ☆ ☆ ☆ ☆ ☆

参考回答

AUC 值为 ROC 曲线所覆盖的区域面积，显然，AUC越大，分类器分类效果越好。

AUC还有另一个意义：分别随机从正负样本集中抽取一个正样本，一个负样本，正样本的预测值大于负样本的概率。

计算方式一，计算ROC曲线覆盖的区域面积

计算方式二：根据定义：分别随机从正负样本集中抽取一个正样本，一个负样本，正样本的预测值大于负样本的概率。根据古典概率模型

$$AUC = \frac{\sum(pred_{pos} > pred_{neg})}{positiveNum * negativeNum} \quad (.)$$

分母是正负样本总的组合数，分子是正样本大于负样本的组合数

答案解析

无。

6. 讲讲分类，回归，推荐，搜索的评价指标 ☆ ☆ ☆ ☆ ☆

参考回答

1. **Accuracy**: 即所有分类正确的样本占全部样本的比例
 2. **Precision**: 查准率。
 3. **Recall**: 查全率
 4. **F1-score**: 衡量Precision和Recall之间的联系
 5. **AUC** 值为 ROC 曲线所覆盖的区域面积, 显然, AUC越大, 分类器分类效果越好。
2. 回归指标 (我们在讲损失函数的时候讲过, 不再赘述)

有几个指标:

1. **均方误差 (MSE)** :
 2. **均方根误差 (RMSE)**
 3. **MAE(平均绝对误差)**
 4. **R Squared**
3. 推荐任务评价指标:

离线评估 offline evaluation

1. 评分预测

对于评分预测模型: 训练数据集训练好数据, 测试数据集预测用户对物品的评分。评价指标: MAE和RMSE

2. 对于Top N模型: 对排名进行评估。评价指标: 准确率, 召回率, F1

在线评估 online evaluation

1. A/B test

将用户划分为A,B两组, A实验组用户, 接受所设计的推荐算法推荐的商品, B对照组用户, 接受基线方法推荐的商品。通过对比两组用户的行为来评估推荐算法的性能。

CTR:用户点击率, 通过该算法计算出的被点击的项目占推荐项目总数的百分比

CR:用户转化率, 用户购买的项目占被点击的项目的比率。

4. 搜索任务评价指标

Accuracy、Precision、Recall、AUC、P-R曲线

答案解析

无。

7. AB test的原理☆☆☆☆☆

 蒋豆芽

线方法推荐的商品。通过对比两组用户的行为来评估推荐算法的性能。

答案解析

ABTest，简单来说，就是为同一个产品目标制定两个方案（比如两个页面一个用红色的按钮、另一个用蓝色的按钮），让一部分用户使用A方案，另一部分用户使用B方案，然后通过日志记录用户的使用情况，并通过结构化的日志数据分析相关指标，如点击率、转化率等，从而得出那个方案更符合预期设计目标，并最终将全部流量切换至符合目标的方案。

资源分享

python

机器学习

算法工程师

软件工程师

春秋招

面试题

软件开发

举报

面经



收藏



赞

相关专栏



机器学习面试题汇总与解析（蒋豆芽面试题总结）

27篇文章 | 90订阅

已订阅

0条评论

 默认排序 ▾

没有回复

请留下你的观点吧~

发布

