

# 最浅显易懂的图解AUC和GAUC

这是一个快乐的人 搜索与推荐Wiki 昨天

点击标题下「[搜索与推荐Wiki](#)」可快速关注

▼ 相关推荐 ▼

- 1、谈一谈算法工程师的落地能力
- 2、转化率预估中的贝叶斯平滑

原文出处: [https://blog.csdn.net/zhaohang\\_1/article/details/92794489](https://blog.csdn.net/zhaohang_1/article/details/92794489)  
原作: 这是一个快乐的人  
编辑: 搜索与推荐Wiki

网上看过不少关于AUC和GAUC的材料，但是都是文字描述，一点也不直观，因此萌生了使用图解的方式详细剖析一下AUC和GAUC的想法，也希望以此能帮助大家快速理解概念。其中，说到AUC就不得不提ROC曲线，因此这里分三个部分来解读：ROC、AUC、GAUC。

## 一、ROC前身：通用的对分类模型的评价

步骤1：给定样本如图A：其中绿色为正样本，红色为负样本

步骤2：训练一个机器学习模型，然后用这个模型对每个样本计算得到一个预测概率，这个概率值在[0, 1]范围内，如图B

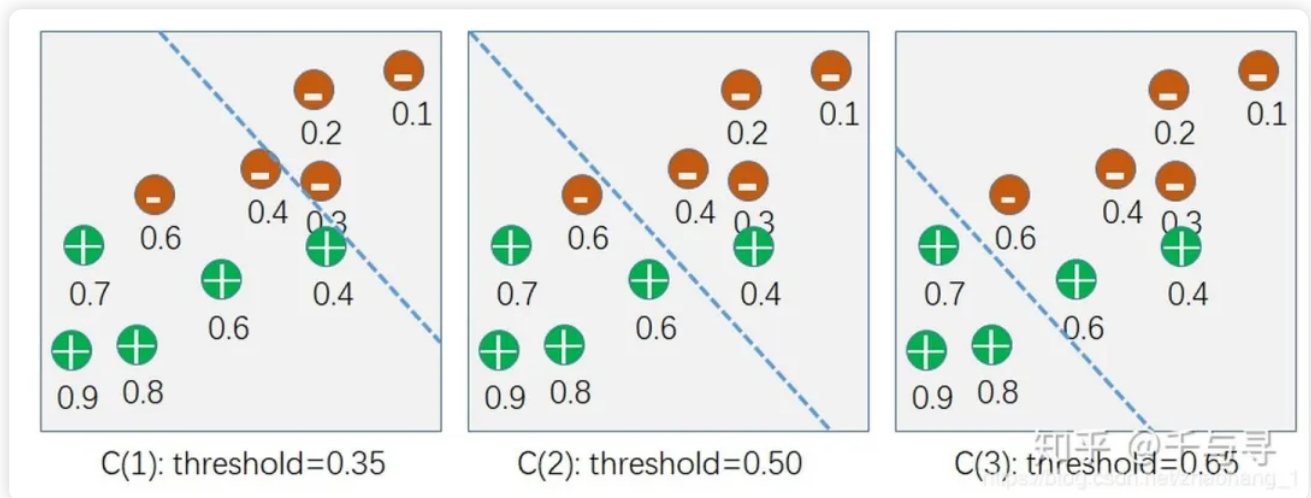


A

B 知乎 @千与寻

[https://blog.csdn.net/zhaohang\\_1](https://blog.csdn.net/zhaohang_1)

步骤3：此时，有一个参数是可以人为指定的：即阈值。当我们将阈值设为0.35、0.50、0.65的时候，会得到以下分类结果。



通常我们都是用取中间的值（0.50）作为阈值：大于0.50为正样本，小于0.50为负样本。但实际我们会发现，不同的阈值直接影响了最终的分类结果：

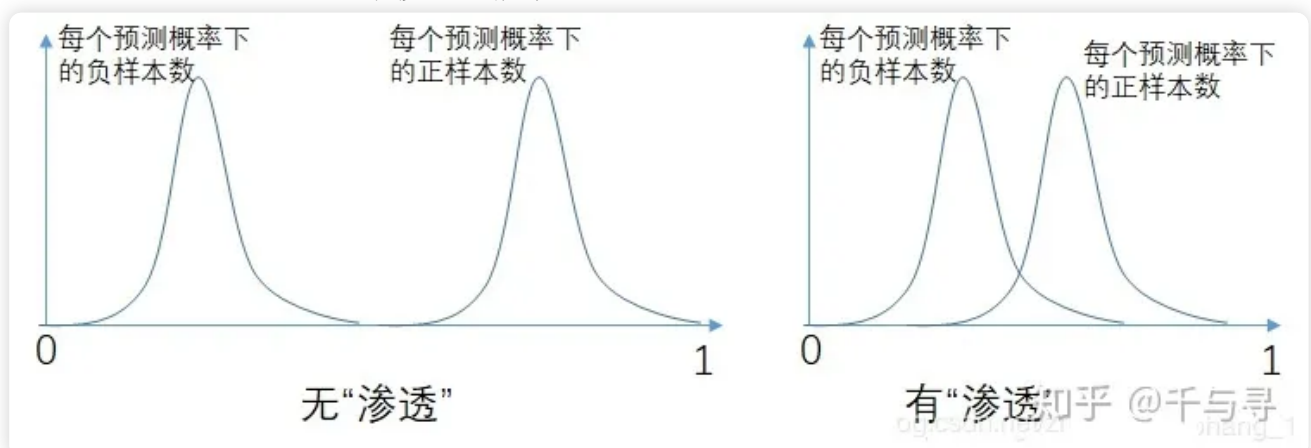
- 1. threshold=0.35：正样本全预测正确，负样本有两个预测错误
- 2. threshold=0.50：正样本有一个预测错误，负样本有一个预测错误
- 3. threshold=0.65：正样本有两个预测错误，负样本全预测正确

小归纳：我们动态地调节了阈值，并发现不同阈值下的分类结果会不同，这么做有什么用呢：实际这可以评估模型对所有样本的预测概率在宏观层面，是否足够好。

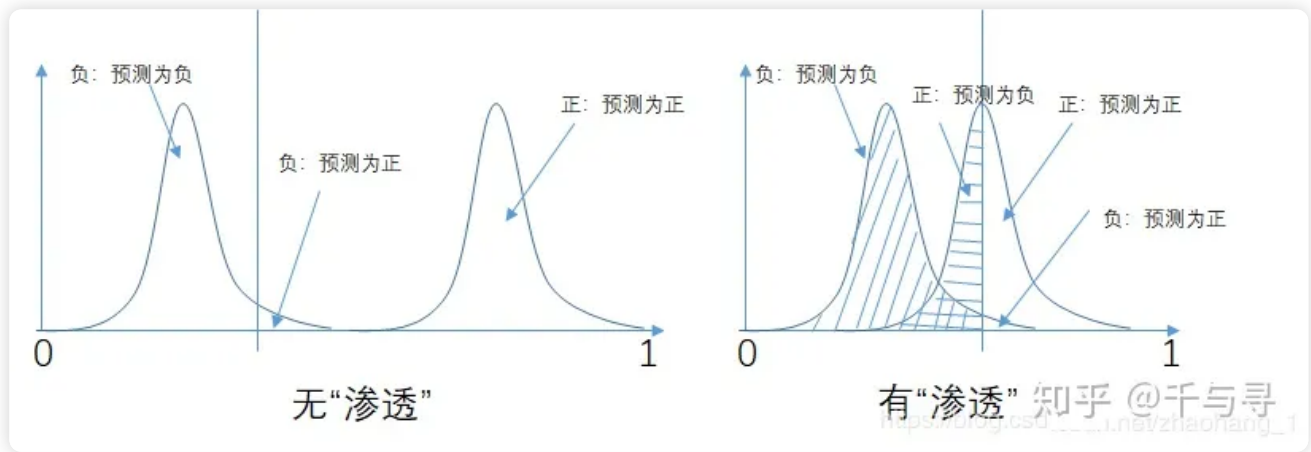
那什么叫预测的概率好呢？有点抽象。简单来说，就是正样本被预测的概率都集中在一起而且倾向于1，负样本被预测的概率都集中在一起而且倾向于0；正负样本被预测的概率尽量不存在互相渗透（正负样本预测的概率值相互揉杂在一起）的现象，实际上，上面的C(1)、C(2)、C(3)都有渗透。

那怎么理解渗透呢？

如下图所示，分别对应两个模型的概率预测结果：



当给两个模型随机指定一个阈值的时候：



直观上，很明显地，左图的预测结果，要比右图好，因为当左图的分割线恰好在两个分布中央的时候，正样本和负样本都完全被预测正确，而右图无论怎么找阈值，都找不到一个这样的阈值能够完全把两类样本分割开。

所以现在应该能体会到遍历所有阈值的好处了吧：

遍历所有阈值取看模型分类分类效。实际上是查看模型对两类数据的区分度是否够大，类间距是否更明显。ok，先理解到这里，接下来我们讨论roc曲线！

## 二、ROC曲线

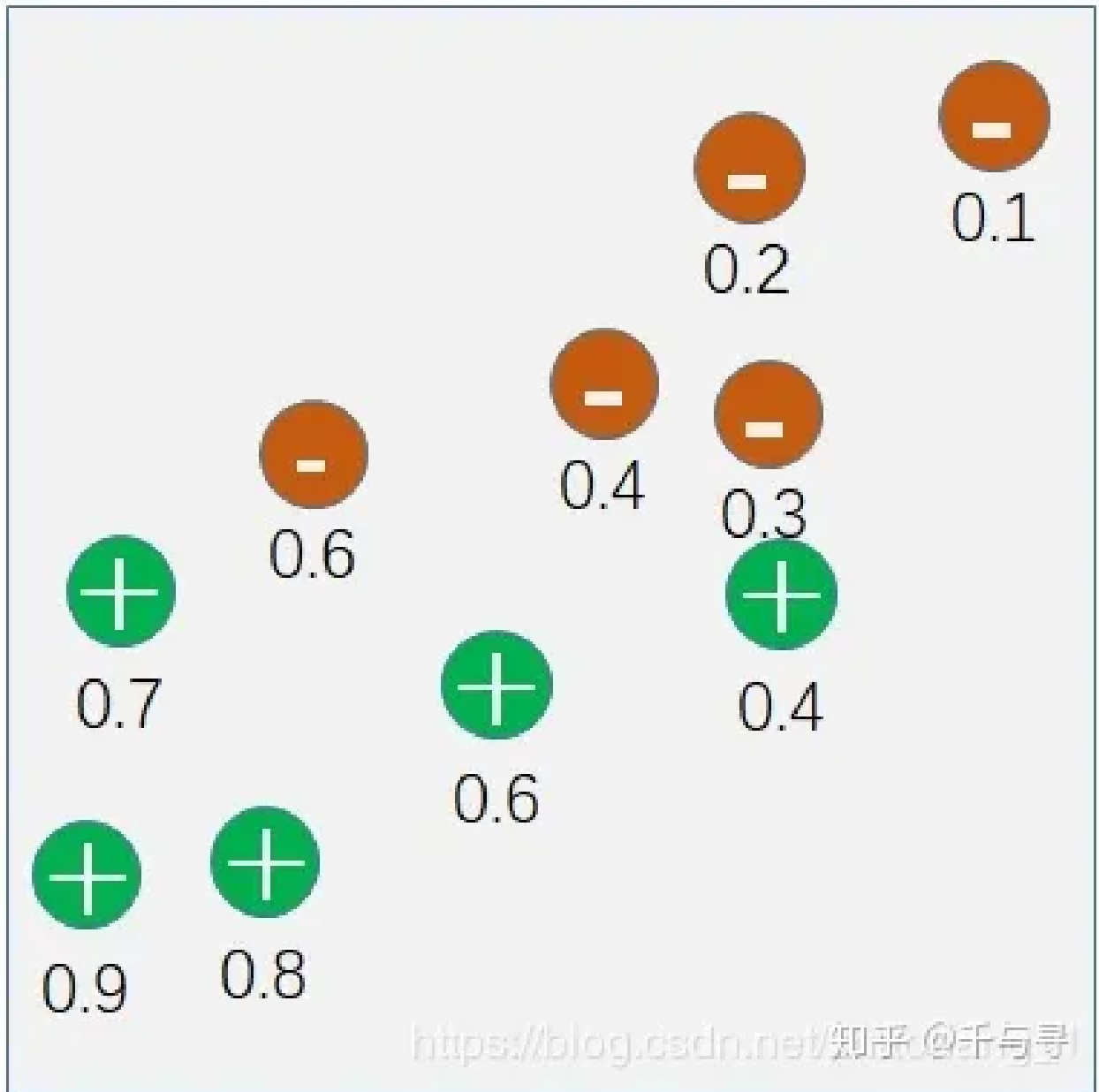
刚才，我们遍历所有阈值，并查看给定每个阈值的情况下的分类情况如何。但是最终我们要把所有的这些分类结果进行融合，如何做呢？大牛们就准备找个方式来度量这个融合，因此诞生了ROC。

由于预测概率值取值范围为： $\{0.9, 0.8, 0.7, 0.6, 0.4, 0.3, 0.2, 0.1\}$

因此我们可选的用于遍历的阈值也是： $\{0.9, 0.8, 0.7, 0.6, 0.4, 0.3, 0.2, 0.1\}$

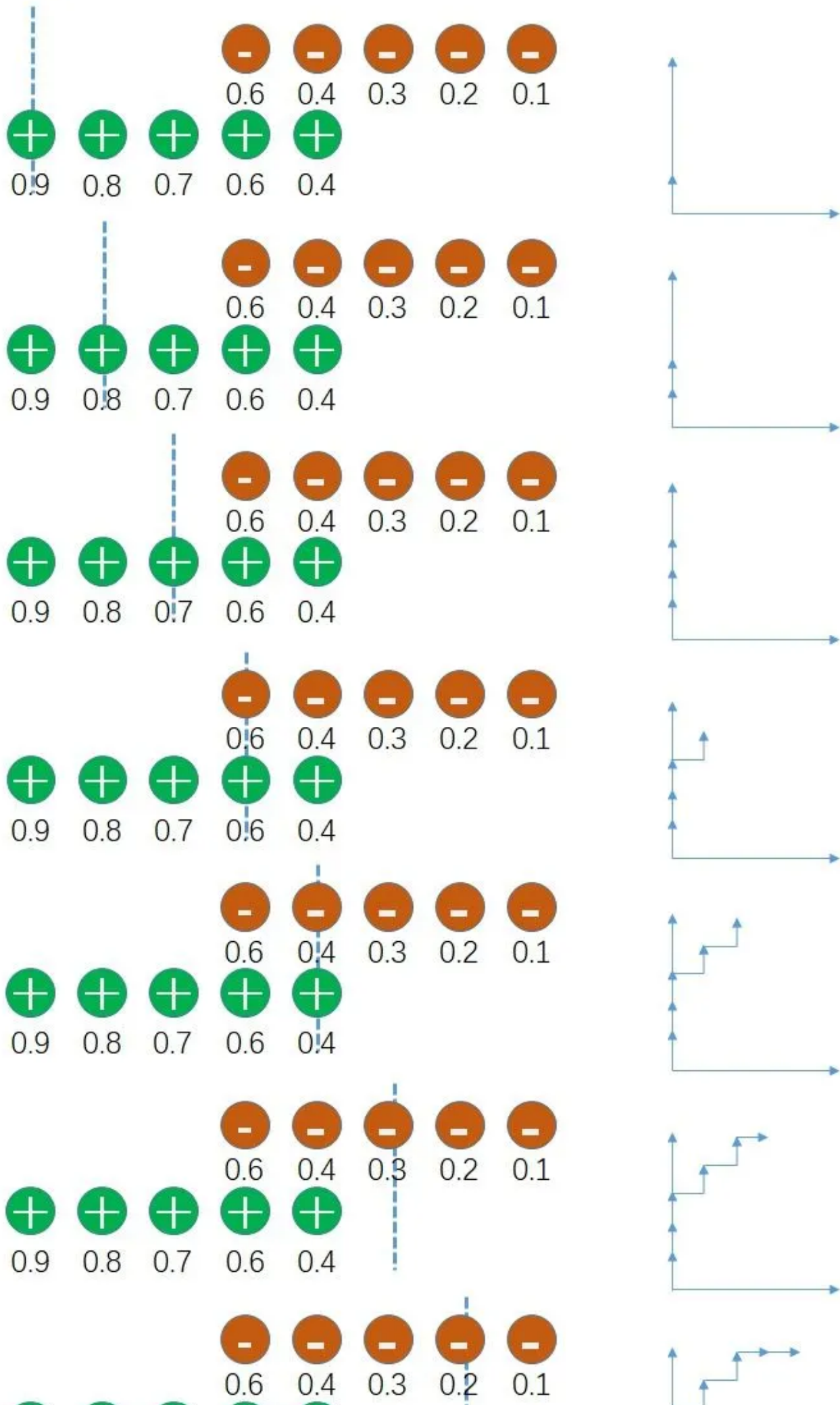
阈值的使用方法是：当预测概率大于等于阈值的时候，则预测结果正样本，否则预测为负样本。

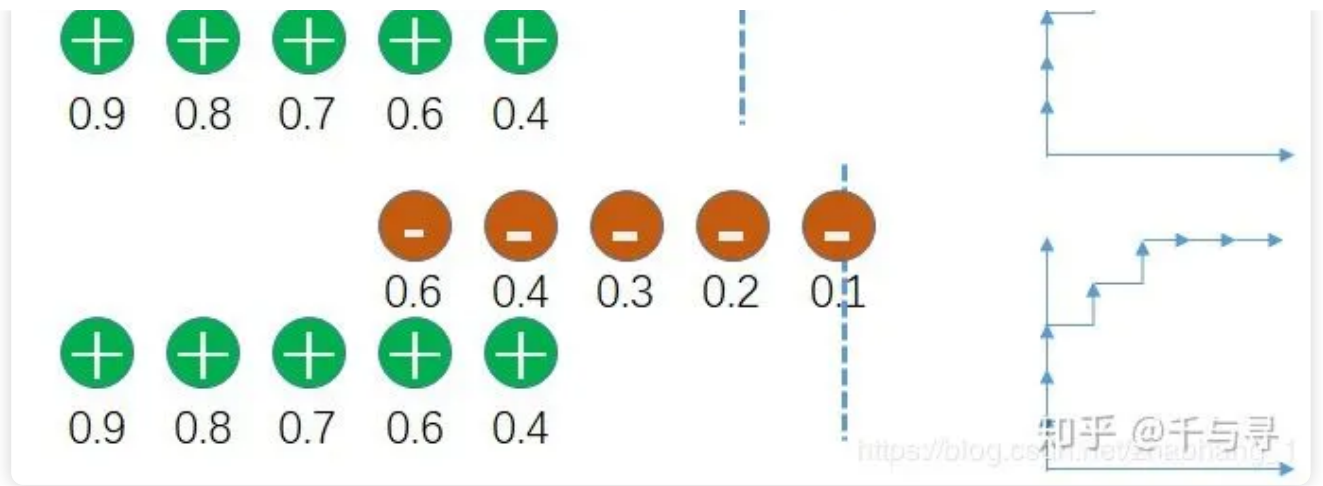
还是用最上面的样本集合，当模型对样本的预测概率如下图时：



我们看一下，在选用不同的阈值的时候，以下两种指标是如何变化的：

- 1.横坐标：假正数（红色的被预测为正样本的个数）
- 2.纵坐标：真正数（绿色的被预测为正样本的个数）





实际至此，我们就得到了ROC曲线D(1)，一般而言，得到了ROC就直接可以计算AUC的值了。通用的结论是

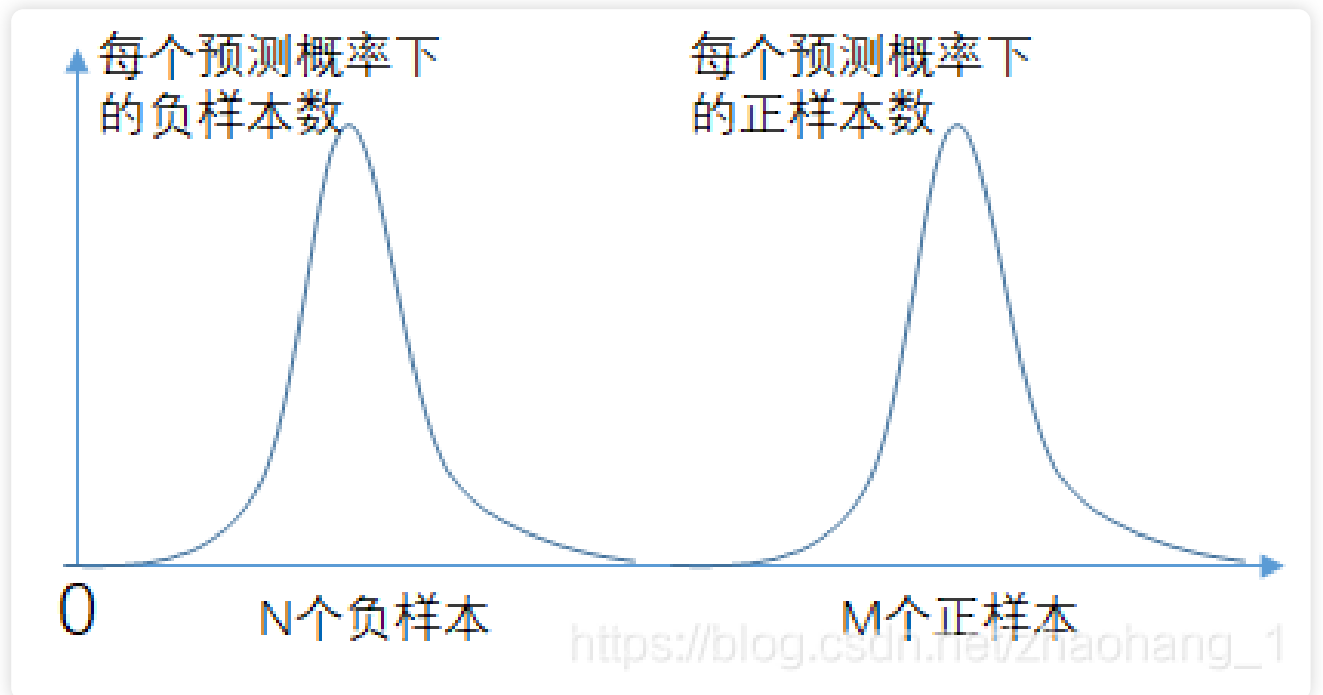
$$AUC = \frac{\text{ROC 曲线下的面积}}{\text{在 } x * y \text{ 区域的乘积}}$$

### 三、AUC

但是ROC曲线下的面积等于AUC，这句话怎么理解呢？（其实这里讲解了AUC的第二种计算方式）

实际上，这可以理解为一种积分过程，积分的内容是啥呢：每个预测为正的样本，能比多少个负样本大 积分所在的区域是啥呢？实际是正样本和负样本的交叉，也即 正样本数\*负样本数

这里我们可以设想一种理想状态：



正样本和负样本是两个互不纠缠的正态分布，其中有M个正样本，N个负样本。



阈值如果遍历所有正样本，则每个正样本都比N个负样本大，因此，积分下来，就是 $N+N+N+\dots+N = M * N$ ，而积分的区域是 $M * N$ ，因此，这种理想状态下，得到了AUC为

$$AUC = \frac{\sum_1^N N}{M * N} = 1$$

因此这里我们可用通过这种方式重新计算第一节中的AUC

第一节中，原始有五个正样本：

$p=0.9$ 的真实正样本，它在所有5个负样本前面，因此记为5

$p=0.8$ 的真实正样本，它在所有5个负样本前面，因此记为5

$p=0.7$ 的真实正样本，它在所有5个负样本前面，因此记为5

$p=0.6$ 的真实正样本，它在4个负样本前面，因此记为4

$p=0.4$ 的真实正样本，它在3个负样本前面，因此记为3

交叉区域记为 $5*5=25$

因此最终的AUC记为

$$AUC = \frac{5 + 5 + 5 + 4 + 3}{5 * 5} = 0.88$$

## 四、GAUC

auc在传统的机器学习二分类中还是很能打的，但是有一种场景，虽然是分类模型，但是却不适用auc，即广告推荐领域。推荐领域使用的CTR（点击率）来作为最终的商用指标，但是在训练推荐模型时，却不用这个指标，用的是GAUC，为什么呢，因为推荐模型目前比较成熟的模式是训练分类模型，这个分类模型的任务是预测用户是否会点击给定的商品，因此，推荐系统的核心，仍然是一个二分类问题，但是是更细力度的二分类。

总结：传统的AUC可以评判二分类，但是推荐领域要算的是对于每个人的二分类结果 给定如下情形：

	模型 $\alpha$		
	iphone（正样本）	华为（正样本）	小米（负样本）
用户A	0.5		0.3
用户B	0.4	0.2	

我们准备训练一个模型用来预测用户A和用户B购买iphone、华为和小米的可能性，我们训练了模型 $\alpha$

对于所有出现的概率值：我们可以计算得到AUC为  $\frac{2+2+1}{2*3} = 0.833$ ，好像预测效果不太好。但是，如果对每个用户查看AUC，则有：

用户A auc:  $\frac{1}{1*1} = 1$

用户B auc:  $\frac{1+1}{2*1} = 1$

也就是说这个模型其实预测地很完美！

所以传统的AUC在这种情况下失效了，由此引入了GAUC来从更细的力度上评估分类结果。

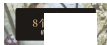
- END -

往期推荐👉



论文 | Item2vec中值得品味的8个经典tricks

2020-12-08



结合论文看Youtube推荐系统中召回和排序的演进之路（上）篇

2021-01-12



和过去的自己和2020好好道个别吧！

2021-01-01



聊一聊海量公众号下我是如何进行筛选和内容消费的

2020-12-06

