

机器学习（周志华）第2.3.4节中，代价曲线的理解？

你可以设定特殊规则或将知乎加入白名单，以便我们更好地提供服务。（为什么？）

[关注问题](#)
[写回答](#)
[机器学习](#)

机器学习（周志华）第2.3.4节中，代价曲线的理解？

[图片] 关于这张图，有几个地方想不通：1) 从ROC能得到FPR和FNR， $(0, \text{FPR})$ 和 $(1, \text{FNR})$ 也不错，但是为什么就能断定它是线段，不能是曲线吗？2) ...显示全部 ▾

[关注问题](#)
[写回答](#)
[邀请回答](#)
[2 条评论](#)
[分享](#)
[...](#)
[应用 · 工作 · 申请开通知乎机构号](#)
[侵权举报 · 网上有害信息举报专区](#)
[违法和不良信息举报者010-8271被浏览](#)
[儿童色情信息举报专区](#)
7,843
[电信与服务业务经营许可证](#)
[网络文化经营许可证](#)
[联系我们 © 2019 知乎](#)
5 个回答
[默认排序 ▾](#)

xf3227

计算生物，机器学习，深度学习

110 人赞同了该回答

注 1：感谢 @Sapaud 指正。西瓜书里假设 0 代表正例，1 反例；而这篇回答默认 0 代表反例，1 正例。很抱歉可能带来的误导，不过在多数语境下，0 反 1 正依旧是传统做法，所以就不更正了。以下是原文截图

数，而是希望最小化“总体代价”(total cost). 若将表 2.2 中的第 0 类作为正类、第 1 类作为反类，令 D^+ 与 D^- 分别代表样例集 D 的正例子集和反例子集，则“代价敏感”(cost-sensitive)错误率为

来源：p.36 机器学习（周志华）

好问题，谢邀~ 不过我不知道周老师说的学习机指的是啥，所以先答第一个问题，因为这个只要代数知识就好了。注意两点

- 1) 代价曲线，以下简称 CC，是参数曲线。这个从纵轴横轴定义就可以看出来了，而且参数是 p 。
- 2) 连接两点的线段如何用参数方程表示？假设两点分别是点 A 点 B ，如果想表示 AB 连线所有点的集合就用 $\lambda A + (1 - \lambda)B$ ， $\lambda \in [0, 1]$ 。通过变化 λ ，我们可以得到线段 AB 。

仔细对比一下 2) 中的式子和 CC 的定义，是不是有很多相通点？只不过 λ 变成了 p 罢了。

A 点呢，就是 $(0, \text{FPR})$ ，即 $p = 0$ 时对应的点。同理， B 点 $(1, \text{FNR})$ 是 $p = 1$ 所对应的。

题主试着变化一下 p 从 0 到 1，是不是这么一条线段呢？

3) 以上说的是不归一情况，必须为直线；如果归一情况，题主说对了~ 不一定是直线，而是一条凸或凹的曲线。

关键在于题主的问题 2 和 3，坐等贴出学习机的定义~

继续说，要理解代价曲线，得直接从期望总体代价的式子入手。


[下载知乎客户端](#)

与世界分享知识、经验和见解

相关问题

[《神经网络和机器学习》，PRML，周志华的《机器学习》和李航《统计机器学习》请各位推荐学习顺序？](#) 14 个回答
[周志华的《机器学习》这本书怎么来学习？](#) 20 个回答

[机器学习方向，南大周志华组lambda实验室vs申请国外cs？](#) 29 个回答

[如何评价周志华在微博对AlphaGo和机器学习技术的评论？](#) 36 个回答

[南大周志华老师的机器学习和深度学习「花书」这两本书的区别以及学习顺序可以给介绍一下吗？](#) 32 个回答

相关推荐



[翘臀不粗腿：成为办公室的共 30 节课](#)

[▶ 试听](#)


[我们一起开始机器学习吧](#)
 王晋东不在家

1967 人参与



[这年头，谁还没点儿心病](#)

我们知道，在二分类问题中，一个分类器对应一个阈值 η ，一个阈值又对应着 ROC 曲线上的一点，也就是 ROC 曲线上一个固定的 (**FPR**, **TPR**)

先不研究总体，只看单个分类器的期望代价，假设我们的分类器是

$$x \begin{cases} \text{decide } H_1 \\ \geqslant \eta \\ \text{decide } H_0 \end{cases}$$

于是该分类器的期望代价就是

$$\mathbb{E}[Cost] = \sum_{i=0}^1 \sum_{j=0}^1 c_{ij} \Pr[\text{decide } H_i | H_j] \Pr[H_j]$$

其中， c_{ij} 来自于我们定义的 cost matrix。

$\Pr[\text{decide } H_i | H_j]$ 是条件概率，又称 likelihood。题主不知道是否熟悉这一说法，但是换个称呼就一定知道了

$$\begin{aligned} \Pr[\text{decide } H_1 | H_0] &= \text{FPR} \\ \Pr[\text{decide } H_1 | H_1] &= \text{TPR} \\ \Pr[\text{decide } H_0 | H_1] &= \text{FNR} \\ \Pr[\text{decide } H_0 | H_0] &= \text{TNR} \end{aligned}$$

$\Pr[H_j]$ 是先验概率，在周老师的书里， $\Pr[H_1] = p$ ，所谓的正例概率。显然负例概率就是 $\Pr[H_0] = 1 - \Pr[H_1] = 1 - p$

根据周老师的语境，我们假设了 c_{00}, c_{11} 都是 0，所以

$$\mathbb{E}[Cost] = c_{10} \times (1 - p) \times \text{FPR} + c_{01} \times p \times \text{FNR}$$

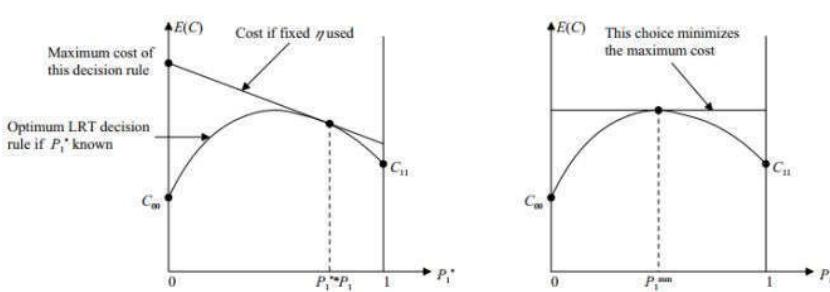
建议题主把上面这个式子的化简自己操作一下，很简单，但是可以加深印象：)

你看一下这个式子，是不是和书中 (2.25) 的分子一摸一样？分母不用管，分母存在只是为了归一化。

再次强调一下， $\mathbb{E}[Cost]$ 目前只是代表单个分类器（不是学习器）的期望代价。那么，连接 (0, FPR) 和 (1, FNR) 的这条线段是什么意思呢？就是 $c_{01} \times p$ （正例先验概率代价）和该分类器的期望代价的关系。比如说，我们看图中的这条粗直线：随着正例先验概率的提高，该分类器的期望代价也随之线性提高。

先看看题主还有什么疑惑，慢些再更~

我觉得理解的难度可能在于跳步太大，直接引入了归一化，不妨先考虑不归一的情况。



这张图的来源是 EC505 Class Note by Prof. D. Castanon & Prof. W. Clem Karl 中的 136 页。波士顿大学的 EC505 讲的是随机过程，不是专门的机器学习课程，但涵盖了详细的机器学习的理论基



赵文平

9人读过

阅读

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引
应用 · 工作 · 申请开通知乎机构号
侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
儿童色情信息举报专区
电信与服务业务经营许可证
网络文化经营许可证
联系我们 © 2019 知乎



下载知乎客户端

与世界分享知识、经验和见解

相关问题

《神经网络和机器学习》，PRML，周志华的《机器学习》和李航《统计机器学习》请各位推荐学习顺序？ 14 个回答

周志华的《机器学习》这本书怎么来学习？ 20 个回答

机器学习方向，南大周志华组lamda实验室vs申请国外cs？ 29 个回答

如何评价周志华在微博对 AlphaGo 和机器学习技术的评论？ 36 个回答

南大周志华老师的机器学习和深度学习「花书」这两本书的区别以及学习顺序可以给介绍一下吗？ 32 个回答

相关推荐



翘臀不粗腿：成为办公室的
共 30 节课

▶ 试听



我们一起开始机器学习吧
王晋东不在家

★★★★★ 1967 人参与



这年头，谁还没点儿心病

础，有兴趣作为理论知识的补充值得一看，网上可以搜到 PDF。

这幅图我觉得说的已是相当清楚，我就补充一下解释。

横轴是 P_1 ，就是正例先验概率，周老师用的符号是 p 。数字 1 就表示正，0 反之，这和编程的习惯是一样。

纵轴是 $E(C)$ ，代价的期望值，周老师还加上了归一，但是这里没有。

注意左图的那根切线。这根切线说的是，给定一个分类边界 η ，(η 是一个阈值，例如，输入大于 η 判断 + 类，小于则判断 - 类)，随着正例先验概率从 0 到 1 变化，期望代价从最高的左边下降到最低的右端。

我来举个例子来说明这个问题。比如我们用身高来判断一个人是否是职业篮球运动员，初定阈值 $\eta = 190$ ，单位当然是公分了，高于 190 就是篮球运动员（正例），矮于 190 就不是（负例）。现在我们看到一个人 180 公分高，根据该判断规则，他自然不是打篮球的，但是有可能判断错了不是？190 以下的职业篮球运动员比比皆是。

既然会分错，代价函数里就得记上一过，那记几分？在上次更新中，我有说

$$E[Cost] = c_{10} \times (1 - p) \times FPR + c_{01} \times p \times FNR$$

这一过就是由该式子后半部分 $c_{01} \times p \times FNR$ 表示的。

c_{01} 可以看作一个代价权重。如果我们的判断规则把一个打篮球的人判断成不打篮球的，这一过记多重？ c_{01} 说了算。

p 在这里，代表所有人口中职业篮球运动员的比例。

FNR 的具体式子是 $Pr[\text{decide } H_0 | H_1]$ 。这是一个条件概率，不熟悉的话，就再回到我们的例子。这个式子说的就是 $Pr[\text{判断规则说被测对象不是篮球运动员} | \text{被测对象是篮球运动员}]$ 。判断规则是什么？矮于 190！所以进一步，是不是等同于 $Pr[\text{被测对象矮于 190} | \text{被测对象是篮球运动员}]$ 。用人话说就是，所有篮球运动员中矮于 190 的比例（概率）。

所以， $p \times FNR$ 得到的是：所有人中既矮于 190 又同时是打篮球的概率。

碰上这种人，咱的 190 规则就会出错，所以 $p \times FNR$ 就是它的出错概率，配合着乘一个代价权重 c_{01} ，得出的结果就是这种错误的期望代价，并记入过簿 $E[Cost]$ 。

同理，也有那么一部分人，高于 190 但并不打篮球，我们的 190 规则也会出错。这种错误的期望代价就是 $c_{10} \times (1 - p) \times FPR$ ，也记入 $E[Cost]$ 。

最终，两种错误的期望代价加一起，就有了

$$E[Cost] = c_{10} \times (1 - p) \times FPR + c_{01} \times p \times FNR$$

这么个式子。（还有没有其他可能的错误？没有了 :))

回过头再去看图，能不能有更清晰的理解？通过滑动 p 我们是能改变 $E[Cost]$ 的，而且这种变化是线性的。关于为何是线性的解释我在回答的最开始已经说过。

现在可以说，图中那条曲线怎么来的了。

看周老师的图，从几何上，我们已经很清楚那条曲线的由来了。阴影部分就是所有可能的线段的下半部分的交集。

如果我们有两个决策阈值 η_1, η_2 ，分别对应图中这两条红色交叉线



赵文平
9 人读过

阅读

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引
应用 · 工作 · 申请开通知乎机构号
侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
儿童色情信息举报专区
电信与服务业务经营许可证
网络文化经营许可证
联系我们 © 2019 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

[《神经网络和机器学习》，PRML，周志华的《机器学习》和李航《统计机器学习》请各位推荐学习顺序？](#) 14 个回答

[周志华的《机器学习》这本书怎么来学习？](#) 20 个回答

[机器学习方向，南大周志华组lamda实验室vs申请国外cs？](#) 29 个回答

[如何评价周志华在微博对AlphaGo 和机器学习技术的评论？](#) 36 个回答

[南大周志华老师的机器学习和深度学习「花书」这两本书的区别以及学习顺序可以给介绍一下吗？](#) 32 个回答

相关推荐



[翘臀不粗腿：成为办公室的共 30 节课](#)

▶ 试听

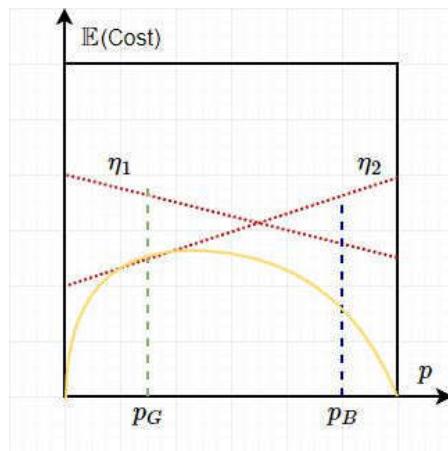


[我们一起开始机器学习吧](#)
王晋东不在家

★★★★★ 1967 人参与



[这年头，谁还没点儿心病](#)



看绿色虚线这里。显然，在正例概率是 p_G 的情况下， η_2 的代价期望更低，做得更好，所以我们取 η_2 。再看蓝色虚线这里，这时候反过来了， η_1 的代价期望更低，所以我们采用 η_1 的期望代价。

我们有许多可能的 η ，从负无穷到正无穷都可以。多画几条，再多画几条，面积越围越小，直到最后不怎么变了，形成了图中黄色这条曲线。现在，对每一个正例概率 p （横轴），这条黄色曲线所对应的纵轴值是什么？就是在所有的可能性中那最小的代价期望！

代价曲线的含义，一言蔽之，就是正例先验概率与最小代价期望的关系。

为了理解得再透彻一点，问自己一个问题：曲线两端为什么一定是从纵轴等于 0 开始的？想一想， $p = 0$ ，这世界上根本就没有篮球运动员？那我设 η 为 1000 公分，反正没有人 10 米高吧？所以 1000 规则见谁都会说不是打篮球的，那我这决策还能出错？不会啊，所以代价期望 $E[Cost]$ 在 $p = 0$ 处一定是 0。那 $p = 1$ 呢？芸芸众生全是职业篮球运动员！那我设 η 为 1，（拇指姑娘不算人 XD），是不是也能达到无错误，代价期望 $E[Cost]$ 为 0 呢？

时隔半月继续更~

既然要说到代价曲线包含的面积问题，先容我先强调几点。

- 上次更新强调过的：代价曲线可视化了**正例先验概率（横轴）和最小代价期望（纵轴）**的关系。而反观 ROC 曲线，它可视化的是 **FPR（横轴）和 TPR（纵轴）**，如果放在侦测理论的语境中，横纵轴又被分别称为误警率（False Alarm Rate）和侦测率（Detection Rate）。浓浓的战争氛围，没错，ROC 曲线最先就是用来评价雷达的好坏的，哪怕现在的教材，习题里没有一道雷达题压压阵也是不达标的。
- ROC 曲线坐标系（强调一下，是那个 1×1 的坐标系）中的一个点，可以对应代价曲线所在的坐标系中的一条线段。不严格点，咱可以把这种关系称为**对偶（Dual）**。强调这个概念是为了说明：这两类曲线是可以互相转换的。因为，一旦你获得了 ROC 曲线的所有信息，你就可以逐点画出另一个坐标系里对应的线段；反之，如果你有一条代价曲线，你就可以通过画切线，看切线左右的截距来得到 FPR 和 FNR。（如果不明白，可以看题主提供的图片）。而有了这两个数字，你就又可以在 ROC 曲线的坐标系中描出相应的点。
- 那他们不一样在哪里？周老师提到，**代价曲线包含了代价的权重 c_{01}, c_{10}** ，但依本人愚见，周老师还漏说了相当重要的一点，那就是**代价曲线同时考虑了先验概率 p 的影响**。在 ROC 曲线，咱有 FPR 和 TPR 就可以描点，不需要先验概率 p 的半点信息；反观代价曲线，整个横轴就是先验概率 p 啊！突然想到：**ROC 曲线所折射的方法论其实就是统计学的频率派**，因为人家不考虑先验啊，而近年（大概是 2006 吧）才出现的代价曲线，它所折射的恰是后来不断发展的贝叶斯派。知乎站内传送门：[贝叶斯学派与频率学派有何不同？](#)

现在说面积，第一个问题：**曲线下面积评价的到底是什么？以周老师教材的语境，我认为是学习器，不是分类器。**

简单来想，最终的代价曲线是无数不同的分类器所对应的线段围啊叠啊的产物，所以这个面积怎么可能也不可能是用来评价单个分类器的。[@petersg](#) 在评论区中提到了极限，很是启发~所以我觉得可



赵女士

9 人读过

阅读

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引
应用 · 工作 · 申请开通知乎机构号
侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
儿童色情信息举报专区
电信与服务业务经营许可证
网络文化经营许可证
联系我们 © 2019 知乎



下载知乎客户端

与世界分享知识、经验和见解

相关问题

[《神经网络和机器学习》，PRML，周志华的《机器学习》和李航《统计机器学习》请各位推荐学习顺序？](#) 14 个回答

[周志华的《机器学习》这本书怎么来学习？](#) 20 个回答

[机器学习方向，南大周志华组lamda实验室vs申请国外cs？](#) 29 个回答

[如何评价周志华在微博对AlphaGo和机器学习技术的评论？](#) 36 个回答

[南大周志华老师的机器学习和深度学习「花书」这两本书的区别以及学习顺序可以给介绍一下吗？](#) 32 个回答

相关推荐



翘臀不粗腿：成为办公室的
共 30 节课

▶ 试听



我们一起开始机器学习吧
王晋东不在家

★★★★★ 1967 人参与



这年头，谁还没点儿心病

以这么严格定义代价曲线（不归一情况）

$$C_l(p) = \inf_{f \in \mathcal{F}} \left\{ c_{01}p \times \text{FNR}_f + c_{10}(1-p) \times \text{FPR}_f \right\}, \quad p \in [0, 1]$$

其中，下标 l 想表示的是一个学习器（learner）； \mathcal{F} 代表所有分类器的全集； $\text{FNR}_f, \text{FPR}_f$ 分别是分类器 f 对应的 FNR 和 FPR。当然，我们从这个定义出发或许还能证明该曲线一定是凹函数，不过太烦了。

然后，曲线下面积可以表示为

$$A_l = \int_0^1 C_l(p) dp$$

所以面积 A_l 是，在假设先验概率 p 符合 0 到 1 均匀分布的前提下，学习器 l 对应的代价期望的总期望。不得不承认这句话相当拗口。。。后面的总期望是针对随机变量 p 而言的，如果要数学严格，我们可以这么看

令 $p \sim B(\alpha, \beta)$ ，则

$$\mathbb{E}_{p \sim B(\alpha, \beta)}[C_l(p)] = \int_0^1 B(p; \alpha, \beta) C_l(p) dp$$

其中， $B(\alpha, \beta)$ 是 Beta 分布的概率密度函数，大家常用它来描述概率的概率 XD。

现在，令 $\alpha = \beta = 1$ ，我们有

$$B(1, 1) = U(0, 1)$$

所以当我们用 0,1 均匀分布的概率密度函数 $U(0, 1)$ 替代原先的 Beta 分布后，就有

$$\mathbb{E}_{p \sim U(0,1)}[C_l(p)] = \int_0^1 U(p; 0, 1) C_l(p) dp = \int_0^1 C_l(p) dp = A_l$$

编辑于 2018-03-02

▲ 赞同 110 ▾ 73 条评论 分享 收藏 感谢

收起 ^

 大吉哥

大吉大力「活捉一只微信公号：大吉哥」

3 人赞同了该回答

我看到的回答，对 cost curve 的理解都有一些问题，所以忍不住来回答一发。

回答之前，必须先给童鞋们明确一下，虽然西瓜书里的表述比较简略，但是内容是完全没有错误的。（答主的回答默认 0 代表反例，1 正例。）

先解决一个简单的问题：(0,FPR) 和 (1,FNR) 的连线是直线。这个很好理解，你把 $p * \text{cost10} / (p * \text{cost10} + (1-p) * \text{cost01})$ 看作 x， $(\text{FNR} * \text{cost10} * p + \text{FPR} * \text{cost01} * (1-p)) / (p * \text{cost10} + (1-p) * \text{cost01})$ 看作 y，那么 (0,FPR) 和 (1,FNR) 不就经过直线 $y = \text{FNR} * x + \text{FPR} * (1-x)$ 吗？

下面理一理 cost curve 和 ROC 曲线的关系：

1. ROC 曲线有两个前提：

- 正反例的分布是确定的，也就是 p 是确定的，且 $p = m_+ / ((m_+) + (m_-))$
- $\text{cost10}/\text{cost01} = 1$ ，（当然 ROC 曲线通过坐标轴变换，也是可以表示 $\text{cost10}/\text{cost01} \neq 1$ 的情况，不过书里没有提到，就不展开说了）

基于 ROC 曲线的这两个前提，可知对于某一条 ROC 曲线来说，由于其概率代价是确定的，即 $p * \text{cost10}$ 是一个定值，那么一条 ROC 曲线映射到 cost curve 的 x 轴坐标，也就仅仅对应 cost curve 的 x 轴上的某一个 $x_0 = p * \text{cost10} / (\text{cost10} + (1-p) * \text{cost01})$ ！！！而 ROC 曲线映射到 cost curve 的 y 轴坐标，即为 $y = \text{FNR} * x_0 + \text{FPR} * (1-x_0)$ 。



赵文半

9 人读过

阅读

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引
应用 · 工作 · 申请开通知乎机构号
侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
儿童色情信息举报专区
电信与服务业务经营许可证
网络文化经营许可证
联系我们 © 2019 知乎



下载知乎客户端

与世界分享知识、经验和见解

相关问题

《神经网络和机器学习》，PRML，周志华的《机器学习》和李航《统计机器学习》请各位推荐学习顺序？ 14 个回答

周志华的《机器学习》这本书怎么来学习？ 20 个回答

机器学习方向，南大周志华组 lamda 实验室 vs 申请国外 cs？ 29 个回答

如何评价周志华在微博对 AlphaGo 和机器学习技术的评论？ 36 个回答

南大周志华老师的机器学习和深度学习「花书」这两本书的区别以及学习顺序可以给介绍一下吗？ 32 个回答

相关推荐



翘臀不粗腿：成为办公室的共 30 节课

▶ 试听



我们一起开始机器学习吧

王晋东不在家

★★★★★ 1967 人参与



这年头，谁还没点儿心病

那么问题来了，既然一条ROC曲线只能映射到cost curve的某个 x_0 ，那么cost curve又代表什么？要解决这个问题，首先要明确一点：cost10和cost01是人为设定的值，它与测试集样本的正反例分布无关，因此，无论cost10和cost01如何取值，只要p确定，FNR和FTR的组合就确定了。那么，cost curve的x轴坐标，即为不同cost10, cost01组合的 $p * \text{cost10} / (p * \text{cost10} + (1-p) * \text{cost01})$ 值；y轴坐标，即为 $y = \text{FNR} * x_0 + \text{FPR} * (1-x_0)$ 的最小值。

2. 如何理解cost curve中的归一化

理清楚了cost curve和ROC曲线的关系，再来说说cost curve中的归一化。

1) 归一化前提：

- 测试集样本的正反例分布是相同的，即p是相同。如果p不相同，则基于不同p归一化之后的值无法进行对比。

举个例子，假设某个NBA球员A,投2分球的概率为0.8,2分球的命中率为0.6;投3分球的概率为0.2,3分球的命中率为0.5,则其归一化之前每一次投篮的得分为：

$$2 * 0.8 * 0.6 + 3 * 0.2 * 0.5 = 1.26$$

归一化之后每一次投篮的得分为：

$$(2 * 0.8 * 0.6 + 3 * 0.2 * 0.5) / (2 * 0.8 + 3 * 0.2) = 0.663$$

假设某个NBA球员B,投2分球的概率为0.1,2分球的命中率为0.5;投3分球的概率为0.9,3分球的命中率为0.6,则其归一化之前每一次投篮的得分为：

$$2 * 0.1 * 0.5 + 3 * 0.9 * 0.6 = 1.72$$

归一化之后每一次投篮的得分为：

$$(2 * 0.1 * 0.5 + 3 * 0.9 * 0.6) / (2 * 0.1 + 3 * 0.9) = 0.593$$

显然，球员A的实际得分能力低于球员B，但是归一化之后的结果确实球员A单次投篮的得分大于球员B;所以，要保证归一化之后，E[cost]的值依然能进行对比，前提就是要保证p相等。

2) 归一化的作用：

- 将p相同，cost01,cost10不同的的E[cost]的值映射到[0,1]进行对比。

还是上面提到的球员得分的例子，假设NBA修改规则，将2分球算作4分，3分球算作6分，那么再次计算其归一化之前每一次投篮的得分为：

2.52

其归一化之后每一次投篮的得分为：

0.663;

由于本次规则修改，只是把得分数乘以2,所以球员的相对得分能力不会和规则修改之前有所不同，显然，归一化之后的数据才是对比球员得分能力变化的指标。

假设NBA修改规则，将2分球算作4分，3分球算作8分，那么再次计算其归一化之前每一次投篮的得分为：

2.72

其归一化之后每一次投篮的得分为：

0.567

由于本次规则修改，加强了3分球得分的收益，所以三分球出手占比和命中率高的球员，其相对得分能力会变得更高。显然，归一化之后的数据正确地反映了该球员得分能力的变化。



赵文平
9人读过

阅读

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引
应用 · 工作 · 申请开通知乎机构号
侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
儿童色情信息举报专区
电信与服务业务经营许可证
网络文化经营许可证
联系我们 © 2019 知乎



下载知乎客户端
与世界分享知识、经验和见解

相关问题

[《神经网络和机器学习》，PRML，周志华的《机器学习》和李航《统计机器学习》请各位推荐学习顺序？](#) 14个回答

[周志华的《机器学习》这本书怎么来学习？](#) 20个回答

[机器学习方向，南大周志华组lamda实验室vs申请国外cs？](#) 29个回答

[如何评价周志华在微博对AlphaGo和机器学习技术的评论？](#) 36个回答

[南大周志华老师的机器学习和深度学习「花书」这两本书的区别以及学习顺序可以给介绍一下吗？](#) 32个回答

相关推荐



翘臀不粗腿：成为办公室的
共 30 节课

▶ 试听



我们一起开始机器学习吧
王晋东不在家

★★★★★ 1967 人参与



这年头，谁还没点儿心病

赵文半
9人读过

阅读

同理， $E[\text{cost}]$ 需经过归一化之后，才能进行比较。

最后，答主解释一下cost curve下部面积的意义，即某一 p 分布的测试集下，学习器基于不同 cost_{01} 和 cost_{10} 组合的总体泛化性能。

发布于 2019-01-18

▲ 赞同 3 ▾ 3 条评论 分享 收藏 感谢

收起 ^

知乎用户

1的话楼上解释的很好，就是直线的参数方程，归一化也应该是直线，注意横坐标不再是单纯的先验概率

至于后面的，我是这样理解的，对于每一个先验概率，我都可以训练出一个分类器使得期望代价最小，所以期望代价曲线取所有线段的最小值。

不知道对不对？@xf3227

另外，我觉得之所以要归一化，就是为了方便求积分

编辑于 2018-08-05

▲ 赞同 3 ▾ 3 条评论 分享 收藏 感谢

匿名用户

8人赞同了该回答

今天看书看到这个没看明白，看了半天没看出所以然，网上也看了一堆，包括这个问题下的回答，都特别高深。后来偶然看到一份读书笔记强调了下其实就是切线的概念，然后恍然大悟。回过头看看一些回答真是有些愤怒。

这里贴出自己的理解，让大家少走弯路。少走弯路，少浪费时间。

首先我很同意关于分母是用来归一化的观点。

首先讲讲为什么分母是这个 $p * \text{cost}_{01} + (1 - p) * \text{cost}_{10}$ 。这个可以理解为我们把全部的样本全部判错的代价，即正的全判负，负的全判正的期望损失。没有分类器可以做到比这种情况更糟的了吧？所以这个保证了我们接下来要说的P和Cost一定不会大于1。所以用这种分母进行归一化。

然后讲讲题主的问题，同时也是我早些时候的困惑

2) 书上说期望总体代价，是线段下的面积，为什么？

3) 所有线段的下界，围城的面积，是所有条件下学习器的期望总体代价，为什么？

2) 其实很简单。

纵轴Y是代价（归一化），横轴X是正例概率代价。而“该条件下”，意思就是分类器f, FNR, FPR, 样本空间给定（这也会使得 $\text{cost}_{10}/\text{cost}_{01}$ 给定），但是P（也就是说p）是变化的（我们可能取到的样本全为正，或全是负，或两者之间变动。）

期望总体代价定义就是 $\int_P \text{Cost} dP$ ，其中P就是 $P(+)\text{cost}$ ，Cost就是 cost_{norm} ，后面都将这么表示。我们将这个积分就暂且设为A吧。为啥是P不是p呢？我是这么理解的，为了将代价引入而且保证这是个概率（所以进行了归一化）。

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引
应用 · 工作 · 申请开通知乎机构号
侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
儿童色情信息举报专区
电信与服务业务经营许可证
网络文化经营许可证
联系我们 © 2019 知乎



下载知乎客户端

与世界分享知识、经验和见解

相关问题

《神经网络和机器学习》，PRML，周志华的《机器学习》和李航《统计机器学习》请各位推荐学习顺序？ 14 个回答
周志华的《机器学习》这本书怎么来学习？ 20 个回答

机器学习方向，南大周志华组lamda实验室vs申请国外cs？ 29 个回答

如何评价周志华在微博对AlphaGo和机器学习技术的评论？ 36 个回答

南大周志华老师的机器学习和深度学习「花书」这两本书的区别以及学习顺序可以给介绍一下吗？ 32 个回答

相关推荐



翘臀不粗腿：成为办公室的
共 30 节课

▶ 试听



我们一起开始机器学习吧
王晋东不在家

★★★★★ 1967 人参与



这年头，谁还没点儿心病

这玩意儿就是在作者给的坐标系中要去求Cost曲线覆盖的面积。而 $\text{Cost} = \text{FNR} * P + \text{FPR} * (1 - P), 0 \leq P \leq 1$ ，这个熟悉吧，就是从(0,FPR)到(1,FNR)的线段。那它覆盖的面积就是A，就是Cost对于P从0到1的积分。所以现在2解决了吧。

那3) 呢？

我们来看看，啥是所有条件。学习器是给定的，但是我们的样本是变化的，设定的阈值也是变化的。啥意思呢，样本变化就导致1.设定的阈值（这个会导致FNR,FPR等等这些玩意儿变化），2.样本（比如阈值一定，样本有变化，那么我们做出来的AUC曲线就会有变化，即FPR,TPR会受到影响会变动）。

再具体点啥意思？

从图上看，

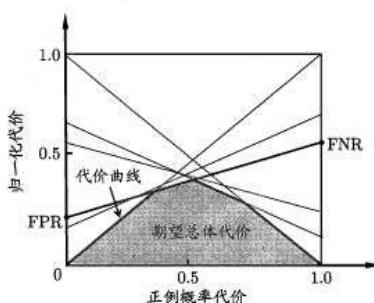


图 2.5 代价曲线与期望总体代价 知乎用户

上面2种变化可能导致线段移动。但是具体会咋移动呢，这个我们也不知道，想起来也挺费劲。可能FPR不变，FNR变；也可能FNR不变FPR变，或者都变。

这种情况下，变量就不仅仅是P了，因为我们要考虑所有情况，那么期望总体代价这个时候可以看作

$$A = \int \text{Cost}(FNR, FPR, p) dFNR dFPR dp = \iint \text{Cost}(\text{阈值}, \text{样本}) d\text{阈值} d\text{样本}$$

，因为要知道 $FNR/FPR/p$ 可是样本和阈值的因变量。

(注意，样本空间和分类器是给定的，所以 $\text{cost}_{01}, \text{cost}_{10}, f$ 都是给定的)

但这个东西好像画不出来，太多种情况了，有点复杂，咋办呢。

好消息是，我们知道给定FNR,FPR后A对于P的求导，就是那条线段。因为A也能表示成

$$\int_P \text{Cost} dP$$

这样呀！而且我们本身就是要求在P的度量下Cost的期望啊。

那么办法来了，那我们就划出A在这个坐标系先关于P的所有切线段（变化阈值以及样本获得所有可能的FPR与FNR），那么它们下界围成的面积就是这个A。

完。



赵文半

9人读过

阅读

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引
应用 · 工作 · 申请开通知乎机构号
侵权举报 · 网上有害信息举报专区
违法和不良信息举报：010-82716601
儿童色情信息举报专区
电信与服务业务经营许可证
网络文化经营许可证
联系我们 © 2019 知乎



下载知乎客户端

与世界分享知识、经验和见解

相关问题

[《神经网络和机器学习》，PRML，周志华的《机器学习》和李航《统计机器学习》请各位推荐学习顺序？](#) 14个回答

[周志华的《机器学习》这本书怎么来学习？](#) 20个回答

[机器学习方向，南大周志华组lambda实验室vs申请国外cs？](#) 29个回答

[如何评价周志华在微博对AlphaGo和机器学习技术的评论？](#) 36个回答

[南大周志华老师的机器学习和深度学习「花书」这两本书的区别以及学习顺序可以给介绍一下吗？](#) 32个回答

相关推荐



翘臀不粗腿：成为办公室的
共 30 节课

▶ 试听



我们一起开始机器学习吧
王晋东不在家

★★★★★ 1967 人参与



这年头，谁还没点儿心病

赵婉平
9人读过

阅读

刚刚接触这东西然后大概想出了点头绪，整体不一定完善但是方向我认为基本没大问题。应该是不难理解的。

主要是要知道这个条件是啥，给定条件的直接影响就是FPR,FNR不变而p (P) 在变动，所有条件就是FPR,FNR,p都会变，然而分类器和样本空间(决定了cost系数)都不会变。

编辑于 2018-11-18

▲ 赞同 8 ▼ ● 2 条评论 分享 ★ 收藏 ♥ 感谢

收起 ^

chenyh

在2.6节阅读材料里提出 是Drummond and Holt,2006 发明了代价曲线，论文名称是Cost curves: An improved method for visualizing classifier performance。在文章中提到了三种正例概率 p+(train):训练集的正例概率, p+(test):测试集的正例概率, p+(deploy):实际使用的整理概率。

书本里代价曲线里的p是第三个，实际使用的数据正例比例，并且事先不知道。我个人认为这个很重要，一开始就是不懂p是什么所以不能理解。

一条ROC曲线里的每一个点实际对应了一个分类器，它可以把FPR比例的真实反例预测为正例，把(1-TPR)比例的真实正例预测成反例。这些分类器在实际使用条件不明确的情况下是无法比较好坏的，比如ROC曲线的(0,0)和(1,1)两个点（两个分类器），前者把所有例子都分为反例，后者把所有例子分为正例，在实际使用中(deploy)，如果p为0（都是反例），那采用第一个分类器效果最好，如果p为1，那用第二个分类器最好。

所以一个分类器在真实情况下遇到不同正例比例时代价也会不同，所以一个ROC的点对应代价平面的一条线段。那期望代价不就是关于p的积分吗。

为什么取下界？那是因为在某一个p的情况下，由于实际情况固定了，我们要选一个代价最小的分类器对吧？所以要取每个p的下界。

还有就是为什么是线段？注意到 $P(+)\text{cost}$ 是关于p的连续递增函数，所以代价平面里的横轴每个点都与p一一对应。纵轴的归一化代价使用的是min-max规范化，把代价式子分子分母同时乘以M(样本总数)就知道是代价的归一化。我认为这个横轴就是人为构造的，能让代价关于P是一个线段（数学不好，抱歉）。

编辑于 2019-03-18

▲ 赞同 ▼ ● 添加评论 分享 ★ 收藏 ♥ 感谢

写回答



机器学习

关注者

10

被浏览

1,698

周志华机器学习2.3.3中的排序损失的定义怎样理解?

[图片] 为什么它的损失刚好是ROC曲线之上的面积 [图片] [图片]...显示全部 ▾

关注问题

写回答

邀请回答

添加评论

分享

...

默认排序 ▾

2个回答



JohnnyMa

各种变化的奇妙旅行者

20人赞同了该回答

用特例证明一下:

假设有5对正反例:

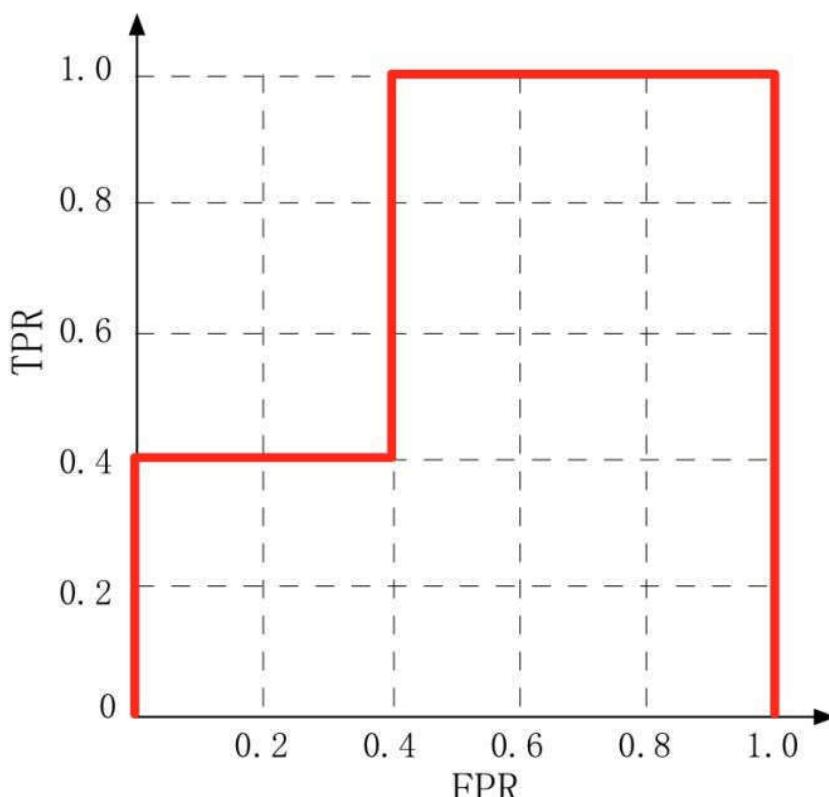
正例预测结果概率为 (0.9, 0.8, 0.5, 0.4, 0.3) ;

反例预测结果概率为 (0.7, 0.6, 0.2, 0.1, 0.01) ;

那么总排序为 (红色为正例概率, 黑色为反例概率) :

(0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.01)

故 ROC 曲线可绘制如下:



可以观察到, 每个虚线所框出的方格面积为 $\frac{1}{m^+ m^-}$, 所以反例概率比正例概率大的情况共有

6

▲ 赞同 20

▼

● 13条评论

分享

★ 收藏

♥ 感谢

收起 ^



下载知乎客户端

与世界分享知识、经验和见解

相关问题

[周志华的《机器学习》这本书怎么来学习?](#) 20个回答

[机器学习方向, 南大周志华组lamda实验室vs申请国外cs?](#) 29个回答

[南大周志华老师的机器学习和深度学习「花书」这两本书的区别以及学习顺序可以给我介绍一下吗?](#) 32个回答

[《神经网络和机器学习》, PRML, 周志华的《机器学习》和李航《统计机器学习》请各位推荐学习顺序?](#) 14个回答

[机器学习中, 有没有给定的阈值返回聚类结果的算法?](#) 29个回答

相关推荐



[翘臀不粗腿: 成为办公室的](#)
共 30 节课 [▶ 试听](#)



[数据挖掘经典评分模型开发](#)
 95 人参与



[专业投机原理 \(珍藏版\)](#)
维克托·斯波朗迪
6 人读过 [阅读](#)

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引
应用 · 工作 · 申请开通知乎机构号
侵权举报 · 网上有害信息举报专区
违法和不良信息举报: 010-82716601
儿童色情信息举报专区
电信与服务业务经营许可证
网络文化经营许可证
联系我们 © 2019 知乎





再考虑另一种情形：



依然有5对正反例：

正例预测结果概率为 (0.9, 0.8, 0.5, 0.4, 0.3)；

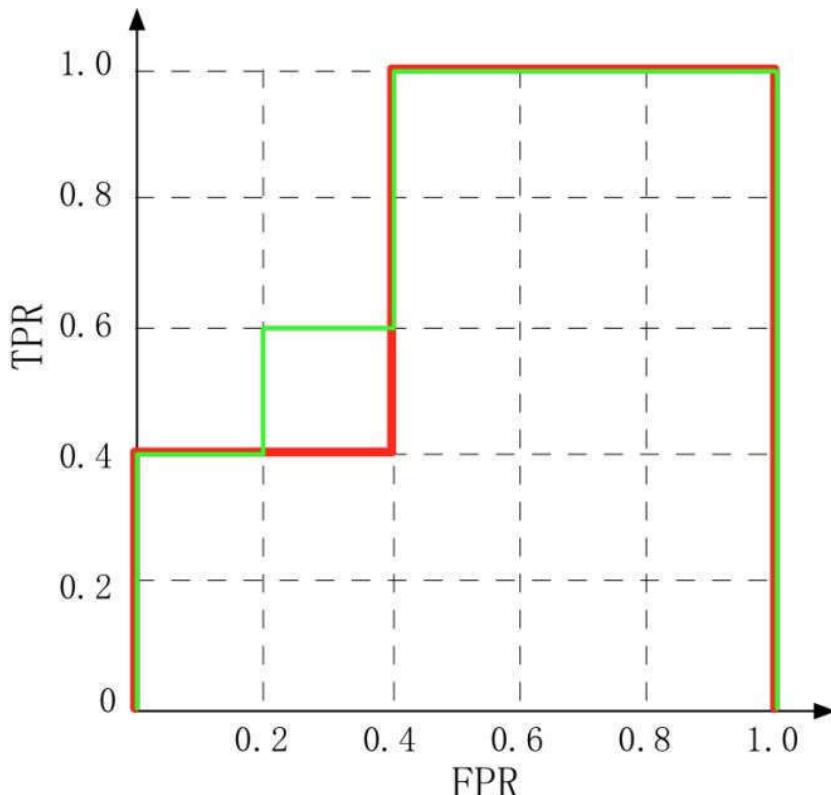
反例预测结果概率为 (0.7, 0.5, 0.2, 0.1, 0.01)；

那么总排序可以有两种情况为（红色为正例概率，黑色为反例概率）：

(0.9, 0.8, 0.7, 0.5, 0.5, 0.4, 0.3, 0.2, 0.1, 0.01)

(0.9, 0.8, 0.7, 0.5, 0.5, 0.4, 0.3, 0.2, 0.1, 0.01)

故 ROC 曲线有相应的两种情况（红色和绿色分别对应情况一与二）：



这时，反例概率比正例概率大的情况共有 $1+2*2=5$ 种，而反例概率与正例概率相同的情况可认为各占一半，所以得出总的排序损失为 $\frac{1}{m^+ + m^-} \times (5 + \frac{1}{2})$ 。

由此推论得证原公式：

$$l_{rank} = \frac{1}{m^+ + m^-} \sum_{x^+ \in D^+} \sum_{x^- \in D^-} (\mathbb{I}(f(x^+) < f(x^-)) + \frac{1}{2} \mathbb{I}(f(x^+) = f(x^-)))$$

编辑于 2018-04-09



zjwang

1 人赞同了该回答

▲ 赞同 20

▼

● 13 条评论

↗ 分享

★ 收藏

♥ 感谢

收起 ^

2019/4/6

周志华机器学习2.3.3中的排序损失的定义怎样理解？ - 知乎

从零点开始， f 会给所有样本一个排序：从 $\max f(x)$ 开始，若是正例，往上走一步，若是反例，往右走一步。（具体是不是这样我也记不太清了）

基本思想是

惩罚施加给每个判错的反例，也就是每一右步竖条上面的面积（这就是指示函数 $f(x_+) < f(x_-)$ 的用处）当然面积是要归一化的 ($1/m + m^-$)

编辑于 2018-11-27

▲ 赞同 1 ▾ 添加评论 分享 收藏 感谢

写回答

▲ 赞同 20 ▾ 13 条评论 分享 收藏 感谢

收起 ^