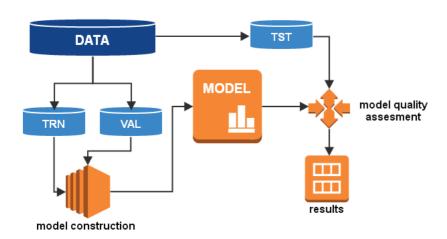


- 测试数据 —— 将被用来评估模型的最终质量;
- 2. 模型构建(使用训练数据和验证数据)
- 3. 模型质量评估(使用测试数据来测试模型)
- 4. 模型应用和后续监控(定期检查模型是否运行正常,性能是否下降)



可以使用哪些指标来确定分类模型的质量呢?

这里有两种指标可以来评估一个分类模型的质量:

• 定量指标 —— 用具体数值来表示分类质量。

- 图形指标 —— 分类质量用一张图来表示。图形化表示方法简化了质量的评估,并且也增强了可视化效果。这些指标包括:
 - 混淆矩阵 (confusion matrix)
 - ROC 曲线
 - LIFT 图表

用于评估分类模型的一些基本概念

二分类和多分类

二分类:

一个类别被定义为正样本,一个类别被定义为负样本。

多分类

一个类别被定义为正类,其他类别的组合都是被定义为负类。

正样本应该是在建模过程中被识别出来的对象:例如在信用评分模型中,正样本包括拖欠债务的顾客,负样本就是剩下所有类别的顾客

TP, TN, FP, FN

• TP —— 模型预测样本是正样本,这个样本的真实结果也是正样本;

比如:模型预测该客户是一个流失客户,然后实际情况是这个客户确实是一个流失掉的客户。

• TN —— 模型预测样本是负样本,这个样本的真实结果也是一个负样本;

比如:模型预测该客户不是一个流失客户,然后实际情况是这个客户确实不是一个流失掉的客户。

• FP —— 模型预测样本是正样本,这个样本的真实结果却是负样本;

比如:模型预测该客户是一个流失客户,然后实际情况是这个客户不是一个流失掉的客户。

• FN —— 模型预测样本是负样本,这个样本的真实结果是正样本;

比如:模型预测该客户不是一个流失客户,然后实际情况是这个客户确是一个流失掉的客户。

predicted→ real↓	Class_pos	Class_neg
Class_pos	TP	FN
Class_neg	FP	TN

对于一个完美分类器来说,我们希望得到如下参数:

FP = 0

FN = 0

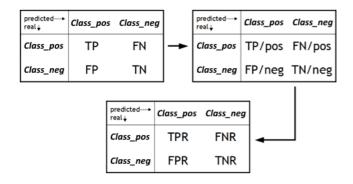
TP = 所有正样本的个数

TN = 所有负样本的个数

一些派生的质量指标

上面我们已经介绍了一些基本的质量指标,现在我们来介绍一些派生出来的质量指标。

- TPR(True Positive Rate)—— 反映了分类器检测正样本的能力。 TPR = TP / (TP + FN) = TP / P , P 表示全部的正样本
- TNR (True Negative Rate) —— 反映了分类器检测负样本的能力。
 TNR = TN / (TN + FP) = TN / N , N 表示全部的负样本
- FPR (False Positive Rate) —— 反映了分类器检测假正性的能力。
 FPR = FP / (FP + TN)
- FNR (False Negative Rate) —— 反映了分类器检测假负性的能力。 FNR = FN / (FN + TP)



- SE (敏感度) —— 反映了分类器检测正样本的能力。
 SE = TP / (TP + FN)
- SP (特异性) —— 反映了分类器检验负样本的能力
 SP = TN / (TN + FP)

所以我们可以得到如下等式:

SE = TPR

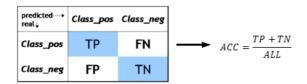
SP = TNR

1 - SE = FNR

1 - SP = FPR

• ACC(总准确率)——反映了分类器的总体预测准确性,即进行正确预测的概率,也就是等于正确决策的个数与总决策个数的比例。

ACC = (TP + TN) / (TP + TN + FP + FN)



例子:如何去选择合适的评估指标

从上面我们可以看到,分类器有很多的评估质量的指标。那么如何选取当前模型的指标呢?这就变成了一个至关重要的问题。

例子

在客户流失分析模型中,我们的数据有两个类别:客户流失和客户不流失。假设我们有100个数据,那么:

- 90 个观察到的数据被标记为 "0" (正类)的类别,这个标记属于客户决定停止使用我们的服务或者产品;
- 10 个观察到的数据被标记为 "1" (负类)的类别,这个标记属于客户决定继续使用我们的服务或者产品;

然后,我们设计的分类器获得了如下结果:

TP = 85 (85个被正确分类为正样本)

TN = 1 (1个被正确分类为负样本)

FP = 9 (9个原来是负样本的被分类为正样本)

FN = 5 (5个原来是正样本的被分类为负样本)

上述指标总和是:

TP + TN + FP + FN = 100

我们现在计算分类的总体准确率:

ACC = (85 + 1) / 100 = 0.86

86%的准确率已经是一个不坏的结果了,但是让我们看一下另一个指标:TNR,这个指标反映了分类器检测负样本的能力,也就是在我们这个例子中们软件的客户:

TNR = 1 / 10 = 0.1

也就是说,只有一个不会流失的顾客被识别出来了,剩下的不流失的顾客都被错误分类了。如果从这个角度分析,那么这个模型是非常糟糕的一个模型 所以,只有一个 ACC 是不能很准确来衡量一个模型的好坏的。

在接下来,我们将采用图形指标来分析模型的好坏。我们将从混淆矩阵开始,这是表示分类结果的最简单方法。

什么是混淆矩阵?

混淆矩阵是一个 N*N 矩阵,其中每一行代表了数据的真实归属类别,每一行的数据总数表示该类别的数据实例的数目。每一列代表了模型预测得到的的总数表示预测为该类别的数据的数目。如下左上角的图,第一行第一列中的 85表示有 85个实际归属为第一类的数据被正确预测为了第一类。同理列中的 9表示有 9个实际是第二类的数据被错误预测为了第一类。

predicted→	Class_pos	Class_neg	predicted→ real↓	Class_1	Class_2	Class_3
-	05		Class_1	80	19	1
Class_pos	85	5	Class_2	23	55	22
Class_neg	9	1	Class_3	13	26	61

	Class ₁	Class ₂	 Class _m
Class ₁	n ₁₁	n ₁₂	 n _{1m}
Class ₂	n ₂₁	n ₂₂	 n _{2m}
		•••	
Class _m	n _{m 1}	n _{m2}	 n_{mm}

混淆矩阵的形式

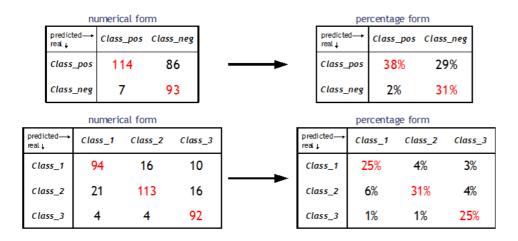
各种形式的混淆矩阵让我们可以更加容易的观察分类模型性能的这种特征,主要有如下两种形式:

• 数字形式 —— 包含分配给特定类别的数字表示

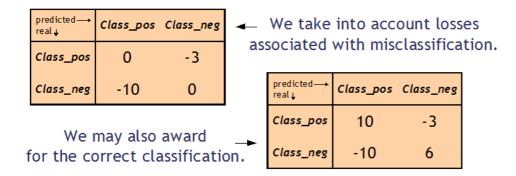
predicted→ real ↓	Class_pos	Class_neg
Class_pos	114	86
Class_neg	7	93

predicted → real ↓	Class_1	Class_2	Class_3
Class_1	94	16	10
Class_2	21	113	16
Class_3	4	4	92

• 百分数形式 —— 包含分配给特定类别的百分比 , 计算为分配给该类别的数量与总数量之间的比值。



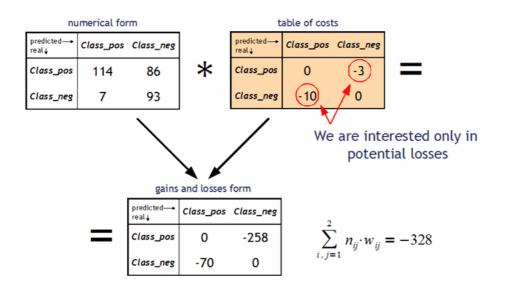
• 收益和损失形式 —— 包含由于正确和错误的分类而产生的收益和损失信息。



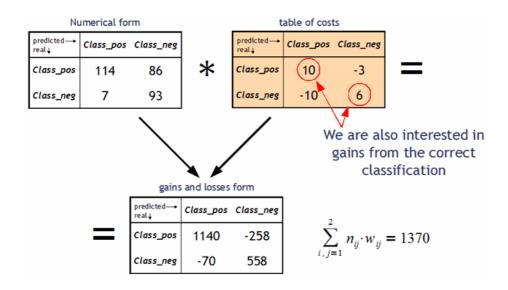
收益和损失形式中的混淆矩阵包含分类决策造成的成本总和。

接下来举几个例子。

例子1



例子2



截止点和混淆矩阵

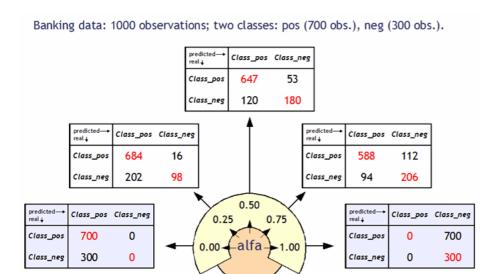
截止点是一个确定的阈值,可以用来确定一个观测属性是否属于一个特定的类别。

如果 P(class(x) = 1) > alfa , 那么这个则分配给第一类。其中 , alfa 就是一个截止点。P(class(x) = 1) 就是属于第一类的概率。

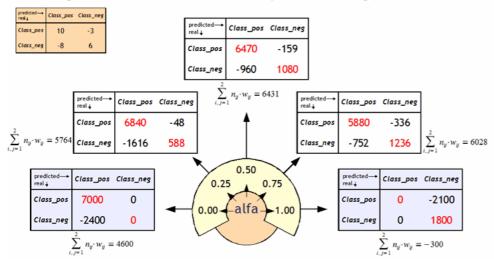
例如:

如果给定一个概率 60%, 那么贷款申请人不善于偿还贷款的概率大于这个值(这个值是我们模型计算出来的), 那么就将这个申请人转入坏账类别, 好账类别。

对于相同的问题, 我们可以考虑不同的截止点, 这也将导致不同的混淆矩阵。通过分析这些矩阵我们就可以选择一个最佳的截止点。







混淆矩阵总结

- 收集分类结果的一种简单易懂的方法
- 使得分类模型的评估更加容易
- 混淆矩阵的不同形式可以帮助更好的观察分类器的性能

什么是 ROC 曲线?

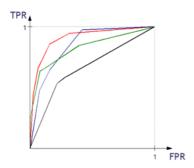
ROC 曲线是分类质量可视化的另外一种方法,显示了TPR 和 FPR 之间的依赖关系。

predicted→ real↓	Class_pos	Class_neg
Class_pos	TP	FN
Class_neg	FP	TN

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

$$FPR (1-specificity) = \frac{FP}{TN + FP}$$

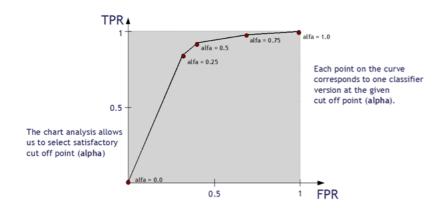
曲线越凸,分类器的性能越好。在下面的例子中,区域1中的"绿色"分类器更好,区域2中的"红色"分类器更好。

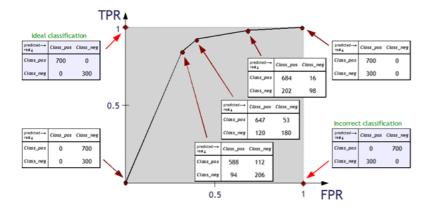


如何构建一条 ROC 曲线

- 1. 我们需要计算决策函数的值;
- 2. 我们测试不同的 alpha 阈值的分类器。回想一下,alpha 是估计概率的一个阈值,高于这个概率的观察值被分配到一个类别(正样本),低于这个阈 被分配至(负样本);
- 3. 对于具有 alpha 阈值的每一个分类,我们获得 (TPR , FPR) 对,其对应于 ROC 曲线上的一个点。
- 4. 对于每个具有 alpha 阈值的分类, 我们也有相应的混淆矩阵;

例子



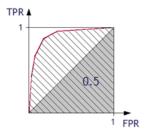


根据 ROC 曲线来评估分类器

分类质量可以使用 ROC 曲线下面的面积大小来计算衡量,这个曲线下的面积就是 AUC 系数。

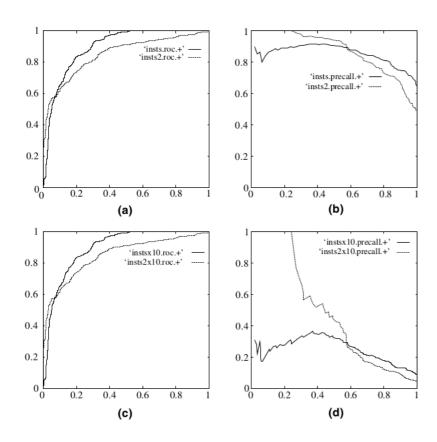
AUC 系数越高越好。AUC = 1 意味着这是一个完美的分类器,我们把所有的东西都分类准确了。对于纯随机数的分类,我们的 AUC = 0.5。如果 AU 意味着这个分类器的性能比随机数还要差。

这里再说一个概念:基尼系数(Gini Coefficient),GC = 2 * AUC - 1。基尼系数越高,代表模型的效果越好。如果 GC = 1,那么这就是一个完美的模 C = 0,那么代表这只是一个随机数模型。



为什么使用ROC曲线

既然已经这么多评价标准,为什么还要使用ROC和AUC呢?因为ROC曲线有个很好的特性:当测试集中的正负样本的分布变化的时候,ROC曲线能够实际的数据集中经常会出现类不平衡(class imbalance)现象,即负样本比正样本多很多(或者相反),而且测试数据中的正负样本的分布也可能随下图是ROC曲线和Precision-Recall曲线的对比:



最后讲讲最常用的精确率和召回率

精确率与召回率多用于二分类问题。精确率(Precision)指的是模型判为正的所有样本中有多少是真正的正样本;召回率(Recall)指的是所有正样本有多少被模型即召回。设模型输出的正样本集合为A,真正的正样本集合为B,则有:

$$\operatorname{Precision}(A,B) = \frac{|A \cap B|}{|A|}, \operatorname{Recall}(A,B) = \frac{|A \cap B|}{|B|}.$$

有时候我们需要在精确率与召回率间进行权衡,一种选择是画出精确率-召回率曲线(Precision-Recall Curve),曲线下的面积被称为AP分数(Average precisio 外一种选择是计算 F_{eta} 分数:

$$F_{\beta} = (1+\beta^2) \cdot \frac{\text{precision-recall}}{\beta^2 \cdot \text{precision+recall}}$$

当 $\beta=1$ 称为 F_1 分数,是分类与信息检索中最常用的指标之一。

完整 PPT 可以关注微信公众号: coderpai, 后台回复 20171223 获得。

作者:chen_h

微信号 & QQ:862251340

简书地址:https://www.jianshu.com/p/c281d52c54d8

CoderPai 是一个专注于算法实战的平台,从基础的算法到人工智能算法都有设计。如果你对算法实战感兴趣,请快快关注我们吧。加入AI实战微信群,ACM算法微信群,ACM算法QQ群。长按或者扫描如下二维码,关注"CoderPai"微信号(coderpai)





chen_h

赞赏是最好的支持



'长按识别'赞赏码 看看谁在赞赏

区块链开发成长路线,10万程序员点赞!

区块链以太坑DApp开发是怎么炼成的?区块链的日益火爆和备受追捧,使得区块链开发人才稀缺,同时更加伴随着高薪,想要转型却不了解,特整理出学习大纲分享

想对作者说点什么?

我来说一句

机器学习模型的评价指标和方法

▲ ◎ 2.6万

http://blog.csdn.net/pipisorry/article/details/52574156 衡量分类器的好坏 对于分类器,或者说分类算法,评价...

如何评估模型好坏

介绍"所有模型都是坏的,但有些模型是有用的"。我们建立模型之后,接下来就要去评估模型,确定这个模型是...



如何评价模型的好坏?

介绍 "所有模型都是坏的,但有些模型是有用的"。我们建立模型之后,接下来就要去评估模型,确定这个模...

回归模型的几个评价指标

对于回归模型效果的判断指标经过了几个过程,从SSE到R-square再到Ajusted R-square, 是一个完善的过程:SSE...

模型评估总结

模型评估是模型开发过程的不可或缺的一部分。它有助于发现表达数据的最佳模型和所选模型将来工作的性能如...

机器学习学习中评价模型好坏的标准

准确率(Accuracy):这个公式计算比较简单,就是预测正确的样本数量除以总的样本数量 但是很多时候,这个...

中国海参85%都是假的!10年女参农曝出30年行业丑闻

升平科技 · 燨燚

分类器性能评价总结

◎ 123

参考: http://blog.csdn.net/sinat_26917383/article/details/51114244 — 、混淆矩阵 如图所示,是分类的混淆矩...

分类器评价指标

转自: http://www.cnblogs.com/549294286/p/3621740.html 精度和召回率是广泛用于信息检索和统计学分类...

分类模型的评估方法

在检验集上计算出的准确率或错误率可以用来比较不同分类器在相同领域上的性能,但需要检验记录的类标号必...

评价模型

评价模型 HeartGo 关注2017.01.19 12:10* 字数 4802 阅读 2941评论 0喜欢 6数据挖掘之评价模型层次分析法(A...

相关热词 js如何如何转义 如何提问如何交流 火星如何如何通讯 如何java项目 如何remix

博主推荐



JeemyJon

Horky

298篇文章



前澳超

103篇文章

算法模型---算法调优---数据挖掘模型效果评估方法汇总

≥ 0 4000

1、基于损失函数的标准1.1、混淆矩阵混淆矩阵用在分类器中,是对每一类样本的统计,包括正确分类和错误...

1.模型评估的方法与性能度量

模型评估的方法与性能度量标签(空格分隔):王小草机器学习笔记周志华《机器学习》笔记整理1.评估模型的...

机器学习模型性能评估方法笔记

● ◎ 1.5万

模型性能评估综述对于模型性能的评估,我们通常分为一下三步: 1. 对数据集进行划分,分为训练集和测试集两...

青羊大道惊爆价!德国扫地机器人 348元今日特价

分类模型的评估方法简介

原创地址:http://www.toutiao.com/i6399510627546563073/,也是本博主的另一个博客地址,感兴趣的可以...

预处理之类别特征编码

前言:在数据预处理过程中,通常需要对类别特征进行编码,以便机器学习模型能够使用数据。本文介绍sklearn ...

模式识别:分类器的性能评价

● 6507

最近开始了模式识别的学习,对模式和模式类的概念有一个基本的了解,并使用MATLAB实现一些模式类的生成...

怎样判断一个诊断(风险预测)模型的好坏?

© 2951

有一类研究看似简单,其实水挺深的。比如你觉得前列腺特异性抗原(PSA)用来筛查前列腺癌不准确,你收集...

机器学习中模型评估与选择中的几个小问题

≥ 0 1.1

Training set、Validation set 与 Testing set有关于训练数据的过程中, validation与testing有何区别, validation的...



百度广告

什么是好产品?如何评价产品?谈产品评价体系模型

△ 2534

什么是好产品?如何评价产品? 作为管理者通常要面临如何评价某一个产品的好坏,以及面向产品线或多产品要...

构建风控评分卡模型介绍(WOE/KS/ROC)

什么是评分卡(信贷场景中)以分数的形式来衡量风险几率的一种手段对未来一段时间内违约/逾期/失联概率的...

精确率、召回率、F1 值、ROC、AUC 各自的优缺点是什么?

作者:京东白条链接:https://www.zhihu.com/question/30643044/answer/222274170来源:知乎著作权归...

机器学习中非平衡数据集的常用处理方法

● 5101

不平衡数据集: 在分类等问题中,正负样本,或者各个类别的样本数目不一致。:在人脸检测中,比如训练库...

为了使分类结果能更好的反应数据的特征,评估分类准确率的方法,主要有保持、随机子抽样、交叉验证和自助...

穷玩车,富玩表,北上广的土豪们却在玩这个,你想不到!

众慧商留· 蟻燚

笔记 | 风控分类模型种类 (决策、排序)比较与模型评估体系 (ROC/g... • ● ③ 3.7万

本笔记源于CDA-DSC课程,由常国珍老师主讲。该训练营第一期为风控主题,培训内容十分紧凑,非常好,推...

无监督聚类算法该如何评价

学过机器学习的小伙伴应该都很清楚:几乎所有的机器学习理论与实战教材里面都有非常详细的理论化的有监督...

简单的评价有监督学习(分类和回归)和无监督学习的技巧

假如只是想要简单的评价一下模型的好坏,并以此作为对比的话,那么使用简答的评价标准就好,不用太复杂。...

衡量模型泛化能力的评价标准

性能度量:衡量模型泛化能力的评价标准。模型的"好坏"是相对的,他不仅取决于算法和数据,还取决于任务需...

数据挖掘(三)分类模型的描述与性能评估,以决策树为例

分类模型的描述主要是混淆矩阵,精确率,召回率等等,分类模型的性能评估则主要讲述了交叉验证这种方式,...



一点点奶茶加盟

百度广告

机器学习模型评价指标总结

评价指标介绍分类模型准确率 Accuracy=N (correct)/N(total)准确率评价没有对不同类别进行区分。...

算法模型好坏、评价标准、算法系统设计

● ◆ 4920

算法模型好坏的评价通用标准: 1、解的精确性与最优性。基于正确性基础上。 2、计算复杂度,时间成本。 3...

深度 | 机器学习中的模型评价、模型选择及算法选择

⊕ ② 2237

作者:Sebastian Raschka翻译:reason_W编辑:周翔简介正确使用模型评估、模型选择和算法选择技术无论是...

常用的一些算法模型评价指标

◎ 1000

1. 基本概念 FN TP TN FP TP -- True Positive (真正, TP)被模型...

干货 | 机器学习中的模型评价、模型选择与算法选择

本论文回顾了用于解决模型评估、模型选择和算法选择三项任务的不同技术,并参考理论和实证研究讨论了每一...

"人喝茶三年,茶养人一辈子"已被科学证实!

三亿茶叶·燨燚

分类算法评价标准

● 8334

一、引言 分类算法有很多,不同分分类算法又用很多不同的变种。不同的分类算法有不同的特定,在不同的数据...

分类模型的评估方法-正确率(Accuracy)

△ ⊚ 2135

我们知道,机器学习的一大任务是"分类"。我们构建了一个分类模型,通过训练集训练好后,那么这个分类模型...

机器学习算法模型评价指标ROC AUC

【导读】在机器学习研究或项目中使用分类精度、均方误差这些方法衡量模型的性能。当然,在进行实验的时候...

如何评估模型的预测性能?

© 2325

you can refer to 如何评估模型的预测性能? 一、模型评估简介 http://www2.mccombs.utexas.edu/faculty/Mayta...

分类器的性能评估

◎ 3766

本文转载于http://funhacks.net/2015/08/12/classifier-evaluation/感谢原作者的分享 1. 背景 当我们使用一个分...



一点点加盟

🧱 百度广告

机器学习实战:模型评估和优化

● 1.2万

原文: Real-World Machine Learning: Model Evaluation and Optimization 作者: Henrik Brink, Joseph W. Richa...

分类评价指标和回归评价指标及python实现

● 1108

目录 目录 1、概念 1.1、前提 2、评价指标(性能度量) 2.1、分类评价指标 2.1.1 值指标-Accuracy、Precision、...

机器学习分类器模型评价指标

分类器评价指标主要有: 1, Accuracy 2, Precision 3, Recall 4, F1 score 5, ROC 曲线 6, AUC 7, PR 曲线混...

分类算法中常用的评价指标

● ● 982

转自:分类算法中常用的评价指标来自网络,备份勿忘!对于分类器(分类算法),评价指标主要有:1)Prec...

目标检测模型中的性能评估——MAP(Mean Average Precision)

▲ ◎ 1.2万

目标检测模型中性能评估的几个重要参数有精确度,精确度和召回率。本文中我们将讨论一个常用的度量指标:...

中国海参85%都是假的!10年女参农曝出30年行业丑闻

升平科技·爔燚

Roc曲线评价标准

1、一个二分问题,即将实例分成正类(positive)或负类(negative)。对一个二分问题来说,会出现四种情况...

分类之性能评估指标

本文主要介绍几种常用的分类评估指标,同时介绍如何绘制ROC曲线以及AUC值的便捷的计算方法。最后再附上...

二分类模型评价指标-总结

knitr::opts_chunk\$set(echo = TRUE) 主要介绍二分类评价的一些指标。 混淆矩阵 预测为正类 预测为负类 实...

性能测试模型

313

关于性能测试模型的探讨如下: 随着单位时间流量的不断增长,被测系统的压力不断增大,服务器资源会不...

2.模型性能的比较检验

标签(空格分隔): 王小草机器学习笔记通过模型评估的方法与性能的度量我们知道了学习器的性能指标,但是...



一点点加盟费

百度广告

衡量分类器的性能指标

评价一个分类器的metric的确有不少。 error rate: 这个不讲了, 当然是错误率越低越好。 不过值得提到的就是...

分类算法评价(集合)

◎ 1791

一、引言 分类算法有很多,不同分类算法又用很多不同的变种。不同的分类算法有不同的特定,在不同的数...

没有更多推荐了,返回首页



关注

向TA提问

评论 原创 粉丝 喜欢 333 88 28 14

等级: 博客 5

访问: 10万+

积分: 4292 排名: 1万+



-点点加盟











最新文章

利用 Python 计算资产 beta 值和市场 beta

第四篇机器学习投资组合--模型测试

第三篇机器学习投资组合——特征选择

第二篇机器学习投资组合——数据清洗

第一篇机器学习投资组合——简介

个人分类

Tensorflow 48篇 ACM 3篇 人工智能 197篇

深度学习 11篇

展开

归档

 2018年9月
 9篇

 2018年8月
 2篇

 2018年7月
 1篇

 2018年6月
 1篇

 2018年5月
 236篇

展开

热门文章

TensorFlow学习系列(二):形状和动态维

度

阅读量:8604

TensorFlow学习系列(五):如何使用队列

和多线程优化输入管道

阅读量:6410 GAN 论文大汇总 阅读量:5506

金融套利策略:理解统计套利的工作原理

阅读量:5378

TensorFlow学习系列(三):保存/恢复和

混合多个模型 阅读量:5085

最新评论

多任务学习与深度学习

larisa_cc: 您好,请问,/data/img1.jpg, 0, 1 /data/img2.jpg, 1, 2 ...

u/ III62.Jp6, 1, 2 ...

TensorFlow学习系列(三)...

guicai1647855685: 微调的时候。加载图就行吗? 不要加载参数?

使用数据驱动进行配对交易:简单交易… brucewong0516:博主,文章写得不错,申请转载 哈哈

Flashtext:大规模数据清洗...

weixin_42668309:想问一下博主,flashtext怎么 处理中文关键字重复而无法替换后一个中文关键字

Flashtext:大规模数据清洗...

weixin_42668309:想问一下博主,flashtext怎么处理中文关键字重复而无法替换后一个中文关键字

的问题