



# 1.1 机器学习中的模型评估



**HUAWEI TECHNOLOGIES CO., LTD.**

www.huawei.com

# 目录

## Contents

1

分类评估指标的计算

2

回归评估指标的计算

# 分类模型基本评估指标

分类模型评估指标可以定量的对模型的效果进行分析，对模型“准确率”进行准确的评估。

**定义：**假定一个二分类问题，标签是<0，1>两种。如果是一个多分类问题，则可以站在某一个类别标签的角度看待其它所有的类别标签都可以归为“其它类”的范畴，将多分类转换为二分类。各种评估指标如表所示。

评估指标	计算公式
准确率 ( accuracy )	$TP + TN / P + N$
错误率 ( error rate )	$FP + FN / P + N$
召回率 ( recall )	$TP / P$
真负例率 ( specificity )	$TN / N$
精度 ( precision )	$TP / TP + FP$
F分数	$2 \times \text{precision} \times \text{recall} / \text{precision} + \text{recall}$
$F_{\beta}$ 分数	$(1 + \beta^2) \times \text{precision} \times \text{recall} / \beta^2 \times \text{precision} + \text{recall}$

**说明：**假定站在标签0的角度。

P：标签为0的样本个数

N：标签为1的样本个数

TP：标签为0且模型判定为0的样本个数

TN：标签为1且模型判定为1的样本个数

FP：标签为1且模型判定为0的样本个数

FN：标签为0且模型判定为1的样本个数

$\beta$ ：非负实数，为了赋予precision和recall不同的权重，一般常用的 $\beta$ 值是2和0.5。

# 分类基本评估指标的使用

分类模型评估指标的使用跟具体的业务场景相关，而且因为可以站在不同类别标签的角度去计算指标，这样就产生（ $7 \times$  类别个数）个基本评估指标，需要在具体的业务场景中进行分析。

**业务场景举例1**：银行判定贷款申请者有无风险。此业务场景下，银行既希望能够扩大贷款业务、又希望降低坏账风险，所以对于“有风险”和“无风险”两种类别的判定评估都需要考虑，各项指标（错误率除外）都要尽可能的高。

**业务场景举例2**：某制造产线判定某生产设备有无故障。此业务场景下，首先要考虑的是故障的判定要准，所以以故障类别计算的召回率要接近100%，在此前提下，specificity要尽可能的高，越高越好。

**评估指标的使用原则：**

- 1、确立标签重要性。要确定哪一个类别是要非常关注的类别。
- 2、重要类别的召回率、精度需要制定一个高的标准。比如必须达到xx%。
- 3、非重要类别的指标尽可能的高。

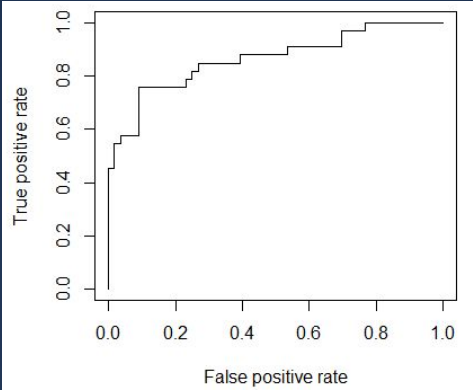
# 分类模型其它评估指标

1、**混淆矩阵**：当面临一个多分类问题，且每个类别的权重几乎等同时，利用混淆矩阵进行模型评估。

单位为样本个数		预测类别		
实际类别		类别1	类别2	类别3
	类别1	500	10	5
	类别2	2	300	1
	类别3	3	4	200

注：矩阵对角线上的值越大越好，其它位置的值越小越好

2、**ROC曲线**：当一些分类器（比如基于神经网络）给出的判定值并不是类别标签，而是一些数值，那么就需要给这些数值一个阈值去产生类别，阈值不同，则类别判定结果不同，这样就产生了基于阈值的recall和（1-specificity），每一对这样的两个值看做平面上的一个点，多个阈值产生的点相连就产生了ROC曲线。ROC曲线越凸，表明模型效果越好。



3、**AUC值**：AUC值为ROC曲线所覆盖的区域面积，AUC越大，分类器分类效果越好。  
**AUC = 1**：是完美分类器，采用这个预测模型时，不管设定什么阈值都能得出完美预测。绝大多数预测的场合，不存在完美分类器。  
**0.5 < AUC < 1**：优于随机猜测。这个分类器（模型）妥善设定阈值的话，能有预测价值。  
**AUC = 0.5**：跟随机猜测一样（例：丢铜板），模型没有预测价值。  
**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。

# 目录

Contents

1

分类评估指标的计算

2

回归评估指标的计算

# 回归评估指标的计算

## 典型回归方法

线性回归

KNN回归

分类回归树（基于平方误差）

.....

## 回归问题的评估指标

平均绝对误差（mean absolute error, MAE）

$$\frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|$$

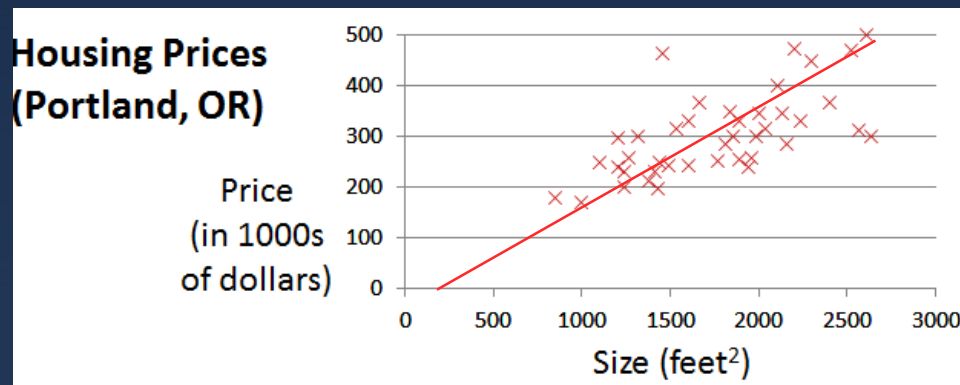
均方根误差（root mean squared error, RMSE）

$$\sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}$$

均方误差（Mean Squared Error, MSE）

$$\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

Size in feet <sup>2</sup> (x)	Price (\$) in 1000's (y)
2104	460
1416	232
1534	315
852	178
...	...



# 回归评估指标的使用原则

回归评估指标计算的都是真实值和预测值之间的误差：

1、使用误差的方式从总体上来判断模型的好坏，原则是误差越小越好，那么这个“小”的程度需要提前规定。比如“平均绝对误差不能超过真实值平均值的10%”，这就是一个程度规定，程度规定要根据具体场景具体制定。

2、可以使用“接受度”的概念将依据误差的评估进行转换，比如设定一个接受度的值为80%，即当预测值与真实值之间的比值处于0.8-1.2之间则视为“可接受”，然后设定一个可接受的样本数，比如“测试集中必须有90%以上的样本达到可接受状态”，这样就相当于用分类评估的思想去处理回归评估。

3、可以根据实际场景的不同设定各种规则，以满足模型验收需求为前提。





# Thank You.

**Copyright©2016 Huawei Technologies Co., Ltd. All Rights Reserved.**

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

华为云机器学习服务MLS  
[www.huaweicloud.com/product/mls.html](http://www.huaweicloud.com/product/mls.html)