

评分卡模型的评价标准

目录

模型的区分度

模型的准确度

模型的稳定性

从概率到分数

模型的区分度

□ 模型的区分度

评分卡模型的结果需要能对好、坏人群给出一定的区分度。

衡量区分度的常用方法：

1, 好、坏人群的分數(或违约概率)的 分布的差异

- KS

2, 好、坏人群的分數(或违约概率)的距离

- Divergence

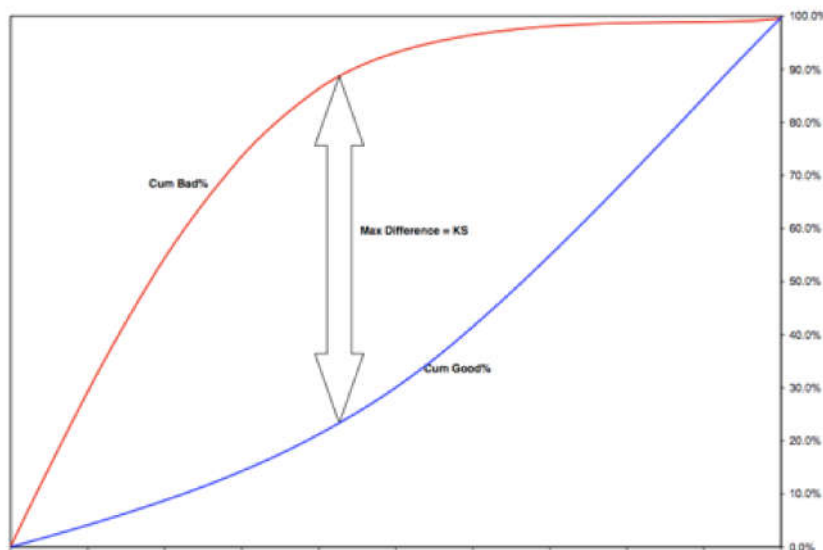
3, 好、坏人群浓度的差异

- Gini

模型的区分度：KS

□ KS(Kolmogorov-Smirnov)

衡量分数区分能力的指标



把样本按分数由低到高排序，X轴是总样本累积比例，Y轴是累积好、坏样本分别占总的好、坏样本的比例，两条曲线在Y轴方向上的相差最大值即KS。KS越大说明模型的区分能力越好。

Bad_k 和 $good_k$ 分别为分数累积到第k个分位点的坏样本个数和好样本个数，KS计算公式是：

$$KS = \max \left\{ \frac{Bad_k}{Bad_{total}} - \frac{Good_k}{Good_{total}} \right\}$$

模型的区分度：KS

□ KS(续)

➤ KS的判断阈值

>0.3 :好

$0.2\sim0.3$:可用

$0\sim0.2$:较差

<0 :评分与好坏程度相悖，模型错误

➤ KS的缺陷

只能表示区分度最好的分数的区分度，不能衡量其他分数

模型的区分度: Divergence

□ Divergence

反映好、坏样本分数(或概率)的“距离”的指标。

Divergence越大, 两类样本的“距离”越大、差异越大。

$$Divergence = \frac{(\mu_{good} - \mu_{bad})^2}{\frac{1}{2} \times (var_{good} + var_{bad})}$$

其中, μ_{good} , μ_{bad} 指好坏的分数(或概率)的均值, var_{good} , var_{bad} 指好坏的分数(或概率)的方差

(可参考t分布)

模型的区分度：Gini

□ Gini

表示分数(或概率)段中的坏样本浓度的差异

$$Gini = \sum \frac{n_i}{N} (1 - p_i^2 - (1 - p_i)^2) = 2 \sum \frac{n_i}{N} (1 - p_i) p_i$$

其中,

- 1, 需要对分数(或概率)进行分组
- 2, n_i 是每组的样本量, p_i 是每组的坏样本率, N 是样本总量
- 3, Gini越小, 表示划分的纯度越高
- 4, Gini同时也受划分方式影响!

目录

模型的区分度

模型的准确度

模型的稳定性

从概率到分数

模型的准确度

□ 模型的准确度

在评分卡模型中，不能用预测的正确率来衡量准确度。

例：

1000个样本，其中有10个坏样本。某模型预测结果中全部样本都是好样本，则正确率 = $(1000-10)/100=99\%$!

评价：

- 该模型的正确率很高
- 该模型没有用

模型的准确度

□ 混淆矩阵

		真实	
		坏	好
预测	坏	True Positive (TP)	False Positive (FP)
	好	False Negative (FN)	True Negative (TN)

模型的准确度

□ 混淆矩阵(续)

真正类率(True Positive Rate)TPR:

$TP/(TP+FN)$, 代表分类器预测的**正类中实际正实例**占有所有正实例的比例。Sensitivity

负正类率(False Positive Rate)FPR:

$FP/(FP+TN)$, 代表分类器预测的**正类中实际负实例**占有所有负实例的比例。1-Specificity

真负类率(True Negative Rate)TNR:

$TN/(FP+TN)$, 代表分类器预测的**负类中实际负实例**占有所有负实例的比例, $TNR=1-FPR$ 。Specificity

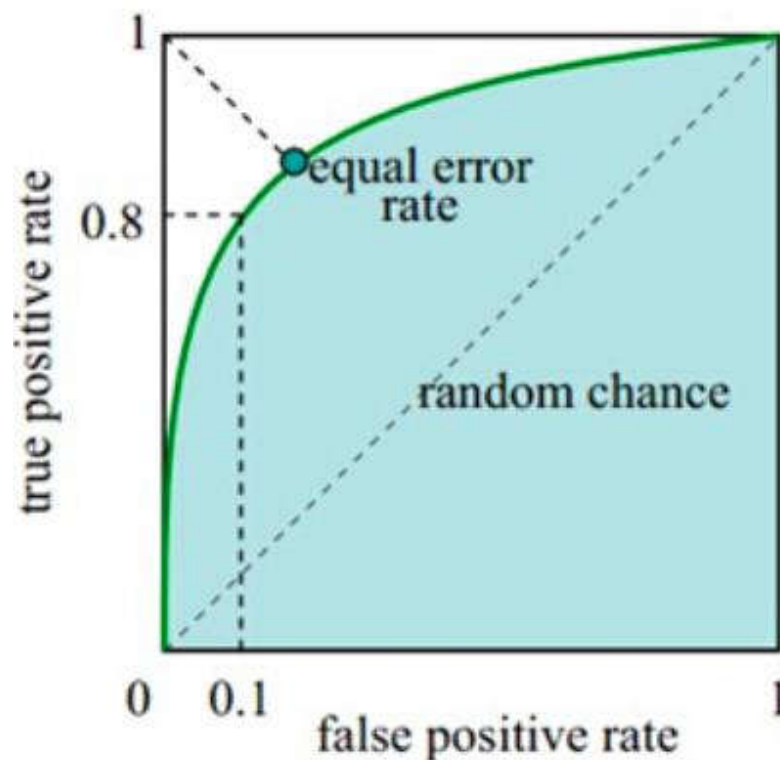
模型的准确度

□ ROC 曲线

逻辑回归分类器给出针对每个样本为坏的概率，那么通过设定一个阈值如0.4，概率大于等于0.4的为坏，小于0.4的为好。对应的就可以算出一组(FPR,TPR),在平面中得到对应坐标点。随着阈值的逐渐减小，越来越多的实例被划分为坏类，但是这些坏正类中同样也掺杂着真正的好类，即TPR和FPR会同时增大。阈值最大时，对应坐标点为(0,0),阈值最小时，对应坐标点(1,1)。

模型的准确度

□ ROC曲线(续)



- 图中实线为ROC曲线，线上每个点对应一个阈值。
- 横轴FPR：越大，预测坏类中实际好类越多。
- 纵轴TPR：越大，预测好类中实际好类越多。
- 理想目标：TPR=1，FPR=0，即图中(0,1)点，故ROC曲线越靠拢(0,1)点，越偏离45度对角线越好，Sensitivity、Specificity越大效果越好。

模型的准确度

□ AUC(Area Under Curve)

ROC曲线下的面积，常介于0.5和1之间(极端情况下低于0.5)，可以直观的评价分类器的好坏，值越大越好。

□ AUC的概率意义

AUC值是一个概率值，当你随机挑选一个坏样本以及好样本，当前的分类算法根据计算得到的概率值将这个坏样本排在好样本前面的概率就是AUC值，AUC值越大，当前分类算法越有可能将坏样本排在好样本前面，从而能够更好地分类。

□ AUC的数值计算

梯形法

模型的准确度

□ AUC(续)

AUC的常用阈值

- >0.7 : 有很强的区分度
- 0.6~0.7 : 有一定的区分度
- 0.5~0.6 : 有较弱的区分度
- 低于0.5 : 区分度弱于随机猜测

□ AUC的优化

AUC可以作为分类模型优化的方向，从而得出参数估计。

目录

模型的区分度

模型的准确度

模型的稳定性

从概率到分数

模型的稳定性

□ 模型的稳定性

评分卡模型需要保持一定的稳定性

- 测试集(out of time sample, OOT) vs 训练集的稳定性
- 在使用阶段的稳定性

模型的稳定性

□ PSI(Population Stability Index)

衡量分数稳定性的指标，通常要求低于25%

$$PSI = \sum_i (R_i - B_i) * \ln(R_i / B_i)$$

其中，

1，将两组样本分为若干份

2， R_i ， B_i 是两组样本在第*i*份上的占比

注：

1，PSI还受到分组的影响

2，PSI越小，说明分数分布变化越小

目录

模型的区分度

模型的准确度

模型的稳定性

从概率到分数

从概率到分数

□ 评分卡分数的计算

评分卡模型的最终产出是分数，且与违约概率呈负相关。

$$score = Base\ Point + \frac{PDO}{\ln(2)}(-y)$$

其中， $y = \text{logit}(p) = \log(\frac{p}{1-p})$

Base Point: 基准分，无实际意义

PDO: Point-to-Double Odds, 好坏比每升高一倍，分数升高 PDO 个单位

从概率到分数

□ 评分卡分数的计算(续)

PDO的证明:

$$y = \text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \log\left(\frac{p_{bad}}{p_{good}}\right)$$

当好坏比升高一倍时,

$$\begin{aligned} -y' &= -\text{logit}(p') = \log\left(\frac{1-p'}{p'}\right) = \log\left(2 \times \frac{p_{good}}{p_{bad}}\right) = \log(2) + \log\left(\frac{p_{good}}{p_{bad}}\right) = \log(2) - y \\ score' &= base\ point + \frac{PDO}{\log(2)} \times (-y') = base\ point + \frac{PDO}{\log(2)} \times (\log(2) - y) \\ &= base\ point + \frac{PDO}{\log(2)} (-y) + PDO = score + PDO \end{aligned}$$

注: 也可以用其他的好坏比率, 比如PTO(point to triple odds), 表示好坏比升高两倍, 分数上升PTO个单位

疑问

□ 小象问答官网

■ <http://wenda.chinahadoop.cn>

课程视频资源扫码：





联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：小象学院
- 新浪微博：小象AI学院

