

# 评分卡

2022-01-14

## 1. 原理

- 1.1. WOE
- 1.2. IV
- 1.3. Logistics
- 1.4. A和B估计
- 1.5. 建模过程

## 2. 评分卡

## 3. 参考文档

## 1. 原理

### 1.1. WOE

假设违约客户为正样本(bad)，其标签为1，正常客户为负样本(good)，其标签为0。则WOE(Weight of Evidence)的定义为：

$$WOE_i = \ln \left( \frac{P_{y1}}{P_{y0}} \right) = \ln \left( \frac{B_i/B_{total}}{G_i/G_{total}} \right)$$

其中， $i$ 表示某特征的第 $i$ 个分组

$B_i$ 表示该分组中bad样本数量， $G_i$ 表示good样本数量  
 $B_{total}$ 表示bad样本总数量， $G_{total}$ 表示good样本总数量

以Age特征为例：

Age	bad	good	WOE
0-10	50	200	$= \ln \frac{50/100}{200/1000} = 0.92$
10-18	20	200	$= \ln \frac{20/100}{200/1000} = 0.0$
18-35	5	200	$= \ln \frac{5/100}{200/1000} = -1.39$
35-50	15	200	$= \ln \frac{15/100}{200/1000} = -1.39$
50以上	10	200	$= \ln \frac{10/100}{200/1000} = -0.69$
汇总	100	1000	

### 1.2. IV

IV相当于WOE的加权平均，公式如下：

$$IV_i = \left( \frac{B_i}{B_{total}} - \frac{G_i}{G_{total}} \right) * \ln \left( \frac{B_i/B_{total}}{G_i/G_{total}} \right)$$
$$IV = \sum_{k=0}^n IV_i$$

其中， $n$ 为分组个数

以Age为例，IV为：

$$(0.5 - 0.2) * 0.92 + (0.2 - 0.2) * 0 + (0.05 - 0.2) * (-1.39) + (0.15 - 0.2) * (-1.39) + (0.1 - 0.2) * (-0.69) = 0.623$$

### 1.3. Logistics

假设违约客户为正样本，其标签为1，概率为 $p$ ，正常客户为负样本，其标签为0，概率为 $1 - p$ 。

令正样本概率为：

$$p = \frac{1}{1 + e^{-(\theta^T X + b)}} \quad (11)$$

定义正负样本概率的比值为 $odds$ ：

$$odds = \frac{p}{1 - p} \quad (12)$$

因此，正样本（违约）的概率越大， $odds$ 越大。

对个人信用评分卡，我们希望违约的概率越大，其分数越低。那么，可使用下式计算Score：

$$\begin{aligned} Score &= A - B \cdot \ln(odds) \\ &= A - B \cdot (\theta^T X + b) \end{aligned} \quad (13)$$

其中， $A$ 和 $B$ 为常数

一般情况下，我们将连续值进行离散化，然后转为WOE作为特征值，使用Logistics进行建模。得到其系数后，便可进行分数的计算。

### 1.4. A和B估计

对评分卡的计算公式：

$$Score = A - B \cdot \ln(odds) \quad (14)$$

对于参数A和B，可基于以下两个假设进行计算：

1. 某个特定的违约概率下的预期分值
2. 指定的违约概率翻倍的分数（PDO）

假设，当对数几率为 $\frac{1}{60}$ 时，设定的特定分数为600，PDO=200，那么对数几率为 $\frac{1}{30}$ 时的分数是620。带入公式得：

$$\begin{aligned} 600 &= A - B * \log\left(\frac{1}{60}\right) \\ 620 &= A - B * \log\left(\frac{1}{30}\right) \end{aligned} \quad (15)$$

解得，A=522，B=29

### 1.5. 建模过程

1. 特征分箱

对特征进行分箱处理

2. 特征选择

选择 $IV \geq 0.1$ 的特征

3. Logistics模型建立

将特征值根据分箱结果转为对应的WOE，然后将WOE作为训练数据进行建模

4. 预测数据-Score计算

$$Score = A - B \cdot (\theta^T X + b) \quad (16)$$

其中， $A$ 和 $B$ 为常数， $\theta$ 和 $b$ 为模型参数， $X$ 为特征对应的WOE

## 2. 评分卡

评分卡：

		left_interval	right_interval	woe	score
feature					
30-59	0	0.00	0.00	-0.48	14.0
	1	1.00	1.00	0.91	-27.0
	2	2.00	3.00	1.71	-51.0
	3	4.00	5.00	2.35	-70.0
	4	6.00	13.00	2.65	-79.0
60-89	0	0.00	0.00	-0.25	8.0
	1	1.00	1.00	1.85	-62.0
	2	2.00	3.00	2.72	-91.0
	3	4.00	11.00	3.11	-104.0
>=90	0	0.00	0.00	-0.35	18.0
	1	1.00	1.00	1.99	-102.0
	2	2.00	3.00	2.82	-144.0
	3	4.00	5.00	3.25	-166.0
	4	6.00	17.00	3.10	-159.0
age	0	21.00	35.00	-0.54	-8.0
	1	36.00	42.00	-0.30	-4.0
	2	43.00	47.00	-0.25	-4.0
	3	48.00	52.00	-0.17	-3.0
	4	53.00	57.00	0.00	0.0
	5	58.00	63.00	0.38	6.0
	6	64.00	70.00	0.86	13.0
	7	71.00	99.00	1.08	16.0
percentage	0	0.00	0.03	1.20	23.0
	1	0.03	0.14	1.16	22.0
	2	0.14	0.51	0.26	5.0
	3	0.51	1.00	-1.05	-20.0

计算示例：

Age	WOE	COE	B	Score=WOE * COE * B
(21, 35]	-0.54	-0.51	29	4 = -0.25*(-0.51)*29
(35, 42]	-0.30			
(42, 47]	-0.25			
(47, 52]	-0.17			
(52, 57]	0.00			
(57, 63]	0.38			
(63, 70]	0.86			
(70, 99]	1.08			

Age = 45      评分

### 3. 参考文档

1. [信用卡评分模型（A卡） - 知乎\(zhihu.com\)](#)
2. [AI智能风控\(二\) ——风控评分卡全流程建模看这篇就够了 - 知乎\(zhihu.com\)](#)
3. 《SAS开发经典案例解析》（杨驰然）