

消费金融场景下的用户购买预测

比赛页面: <https://www.datafountain.cn/competitions/287/details/data-evaluation>

本次目的在于模型的构建, 故数据进行了简化。

数据地址: 链接:https://pan.baidu.com/s/1n9erbLLXEoGBd8aD5P_AhA 密码:mth8

数据背景

个人属性与信用卡消费数据: 包含 80000 名信用卡客户的个人属性与信用卡消费数据, 其中包含枚举型特征和数值型特征, 均已转为数值并进行了脱敏和标准化处理。数据样例如下:

USRID	V1	V2	v3	...	V30
000001	-1.2212	0.4523	1.3251	...	-1.2212
000002	-1.0987	0.0165	-1.0684	...	0.0925

任务

请将数据集分为训练集合与测试集合进行建模。构建以下四个分类器:

- (1) 逻辑回归 LR 模型
- (2) SVM 模型
- (3) 决策树模型
- (4) 神经网络模型

此问题是一个二分类问题, 构建模型后对测试集合求出以下指标:

- (1) precision (2) recall (3) F1-score (4) accuracy

并给出结果的各指标的意义。

思考题

- (1) 分析构建的模型 recall, precision 的指标是否够好。
- (2) 思考有哪些构造提高各指标的方法, 若有思路请尝试实现并与之前的效果进行对比。

PS: 可以在要求的四个分类器基础上, 增加对模型的优化 (如逻辑回归模型中损失函数加入正则项防止模型过拟合)