数据分析 打金工作室 第一轮 解题过程说明

1. 特征：

**uin粒度的特征：**

- 登陆次数

- 登出次数

- 同时在线角色个数

- 最早上线小时点

**ip特征：**

- uin对应ip个数

- uin-roleid对应ip个数的统计特征

- 对ip进行count编码/target编码 https://blog.csdn.net/ssswill/article/details/90271293

**时间特征：**

- 在线总时长

- 最早上线时间点

- 最晚上线时间点

- 各个时间段登陆次数

- 每个uin-roleid层次上在线时长统计特征std\mean\sum\max等

**属性特征：**

- 角色个数，onehot编码后求sum

- 角色职业的纯度

- 朋友总数、朋友数量变化，统计特征std\mean\sum\max等

- 等级变化、统计特征std\mean\sum\max等

- 能力变化、统计特征std\mean\sum\max等

**Create与chat特征：**

- 角色创建时间

- 聊天数目

**Money与item特征：**

- hour粒度上的交易统计特征

- 6小时间隔的统计特征

1. 模型
2. 模型选用lightgbm
3. 参数如下：

params = {

"objective": "binary",

'num\_leaves': 63,

"learning\_rate": 0.05,

"lambda\_l1": 0.1,

"lambda\_l2": 0.1,

'feature\_fraction': 0.75,

'bagging\_fraction': 0.9,

'bagging\_freq': 3,

'seed': 2050,

"iterations": 6666,

"num\_threads ": 16,

'metric': {'binary\_logloss', 'auc'},

'is\_unbalance': True

}

1. 后处理

在用5-fold交叉验证获取5-flod数据后，根据指标对threshhold进行微调，确定为0.875