



数据分析及实践

Analysis and Practice of the Data

实验课

刘 淇

Email: qiliuql@ustc.edu.cn

课程主页:

<http://staff.ustc.edu.cn/~qiliuql/AD2021.html>



实验三

□ 实验要求

□ 给定一个数据集和预测任务，需要分析数据以及抽取特征

□ 数据分析

- 单个特征的分布
- 异常样本
- 特征间的相关性
- 特征与预测目标的相关性
- ...

□ 特征抽取

- 特征的变换，如：str -> int, 取log
- 尝试组合特征
- 特征 aggregation
- ...



实验三

数据集说明

LoL比赛数据集，每一行数据代表一场5V5比赛，

queueId	mapId	seasonId	team1_win	team1_firstBlood	team1_firstTowerKills	team1_firstInhibitorKills	team1_firstBaronKills	team1_firstDragonKills	team1_firstRiftHeraldKills	player1_championId	player1_kills	player1_deaths	player1_assists	player1_gold	player1_role	player1_lane
420	11	13	0	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	107	2	4	0	5385	DUO_SUP	NONE
420	11	13	1	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	76	10	5	6	11865	NONE	JUNGLE
420	11	13	0	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	51	7	6	2	12985	DUO_CAR	BOTTOM
420	11	13	0	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	28	5	3	5	8953	NONE	JUNGLE
420	11	13	1	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	79	4	2	10	8347	NONE	JUNGLE
420	11	13	1	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	555	4	8	8	9650	DUO_SUP	BOTTOM
420	11	13	1	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	145	19	3	9	16990	DUO_CAR	BOTTOM
420	11	13	1	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	127	3	6	16	10413	SOLO	MIDDLE
420	11	13	0	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	98	3	5	1	7337	SOLO	TOP
420	11	13	0	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	119	10	7	6	15504	DUO_CAR	BOTTOM
420	11	13	0	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	0	0	0	0	10710	SOLO	MIDDLE

team1_win代表team1是否取得胜利

team1_firstBlood代表team1是否取得一血

其他类似特征同理

player1_championId代表玩家1选择的英雄ID

player1_kills代表玩家1的击杀数

其他类似特征同理

player1~player5属于team1， player6~player10属于team2

预测任务

预测比赛胜负，即team1_win



实验三

注意事项

- 该实验需要掌握：Jupyter、numpy、pandas库、matplotlib库
- 每位同学将代码和图表保存在 Jupyter notebook 中
- 实验报告中记录数据分析结论和提取的特征，注意不要将代码放在实验报告中
- 数据集会发布于QQ群中



实验三

□ 提交要求

- 将 **jupyter** 文件和实验报告打包成一个压缩文件，发送给助教：apdata2021@163.com
- 邮件标题：姓名_学号_exp3
压缩文件命名格式：姓名_学号_exp3.zip (rar)
- 截止日期：4月25日

□ 评分标准：

- 格式是否规范
- 数据分析、特征提取是否详尽
- 提交是否及时

实验三-参考资料

□ 参考资料：

- kaggle、天池等数据科学网站的初学者教程，如：
<https://www.kaggle.com/startupsci/titanic-data-science-solutions>
- 《利用Python进行数据分析-第2版》

