**2017 “贝贝网·种子杯”初赛解题报告**

# 使用语言及运行环境

**1）python 3.6运行环境：Windows 10（64位）**

**2) matlab R2017a 运行环境Windows10 (64位)**

# 2.依赖的脚本

生成球队总数据.py 输入：球员数据（teamdata.csv） 输出：球队数据（Team\_Per\_Game\_State.csv）

生成日程.py 输入：未来比赛日程（）输出：格式化后的比赛日程（）

生成训练集.py 输入：训练集数据（） 输出：格式化后训练集数据（）

prediction.py 输入：所有球队数据 （）输出： 预测结果

测试集

训练集

主队置信度.py 输入：预测结果 （） 输出：置信度结果（）

logisticTrainClassifier.m 输入：预测数据+相应球队数据构成的矩阵

输出：第一场所获胜的对应的置信度和结果

# 3.数据特征提取思路

### （1）.球队数据提取

| **数据名** | **含义** |
| --- | --- |
| Rk -- Rank | 排名 |
| G -- Games | 参与的比赛场数（都为82场） |
| MP -- Minutes Played | 平均每场比赛进行的时间 |
| FG--Field Goals | 投球命中次数 |
| FGA--Field Goal Attempts | 投射次数 |
| FG%--Field Goal Percentage | 投球命中次数 |
| 3P--3-Point Field Goals | 三分球命中次数 |
| 3PA--3-Point Field Goal Attempts | 三分球投射次数 |
| 3P%--3-Point Field Goal Percentage | 三分球命中率 |
| 2P--2-Point Field Goals | 二分球命中次数 |
| 2PA--2-point Field Goal Attempts | 二分球投射次数 |
| 2P%--2-Point Field Goal Percentage | 二分球命中率 |
| FT--Free Throws | 罚球命中次数 |
| FTA--Free Throw Attempts | 罚球投射次数 |
| FT%--Free Throw Percentage | 罚球命中率 |
| ORB--Offensive Rebounds | 进攻篮板球 |
| DRB--Defensive Rebounds | 防守篮板球 |
| TRB--Total Rebounds | 篮板球总数 |
| AST--Assists | 助攻 |
| STL--Steals | 抢断 |
| BLK -- Blocks | 封盖 |
| TOV -- Turnovers | 失误 |
| PF -- Personal Fouls | 个犯 |
| PTS -- Points | 得分 |

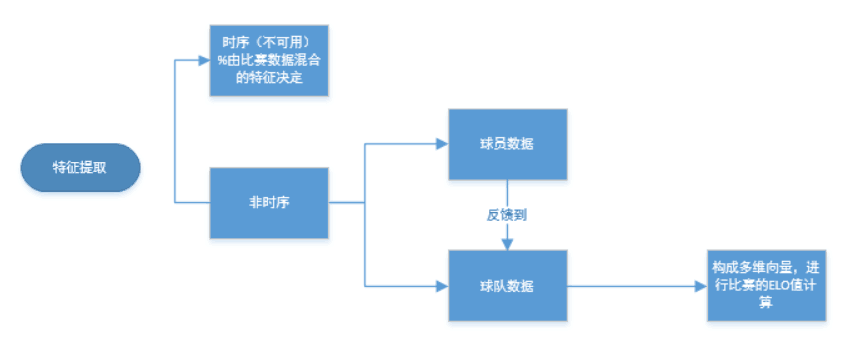
### （2）比赛数据提取

| **数据名** | **含义** |
| --- | --- |
| 缺省 | 主队 |
| 缺省 | 客队 |
| 缺省 | 主队之前胜场 |
| 缺省 | 客队之前胜场 |

### 提取思路:

**首先考虑时序（连胜）和非时序，在询问主办方得知无法获取时序后，我们考虑非时序方法。**

**其中包括球队和球员两个部分（见上表），球员数据反馈到球队数据上，作为球队数据的一个维度，最终每个球队数据处理为一个多维向量。**



# 4.预测模型选取

将上述数据项构成77维向量，在进行数据模型的试训练过程中，发现在去除部分重复维度降维降到50时，对模型的影响不大，故考虑训练速度和非有效项的干扰，我们后面采用了降维后的50向量

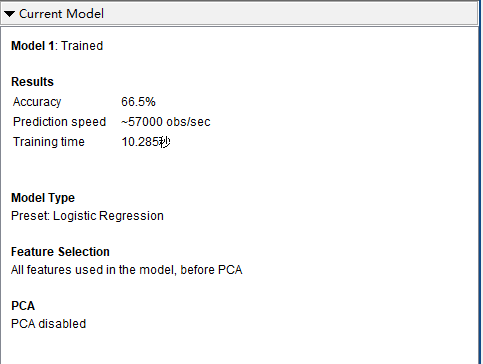
在matlab中进行classsifiing learner 其中尝试进行disabled PCA 降维。由图得，在此我们使用logistic Regression

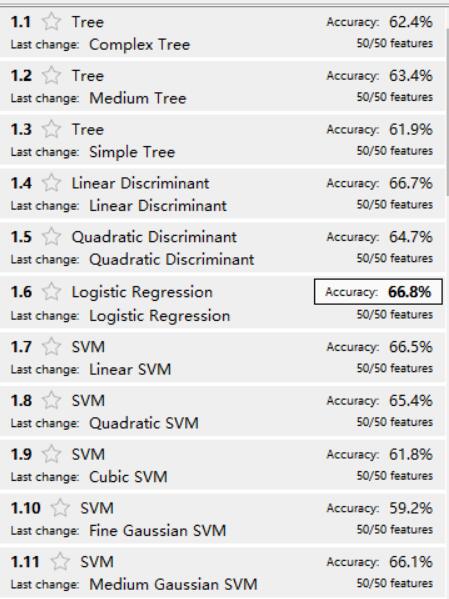


C:\Users\89383\Documents\Tencent Files\893833413\Image\Group\Z3J$SQ(9G0{52C`)JHYXYAU.jpg

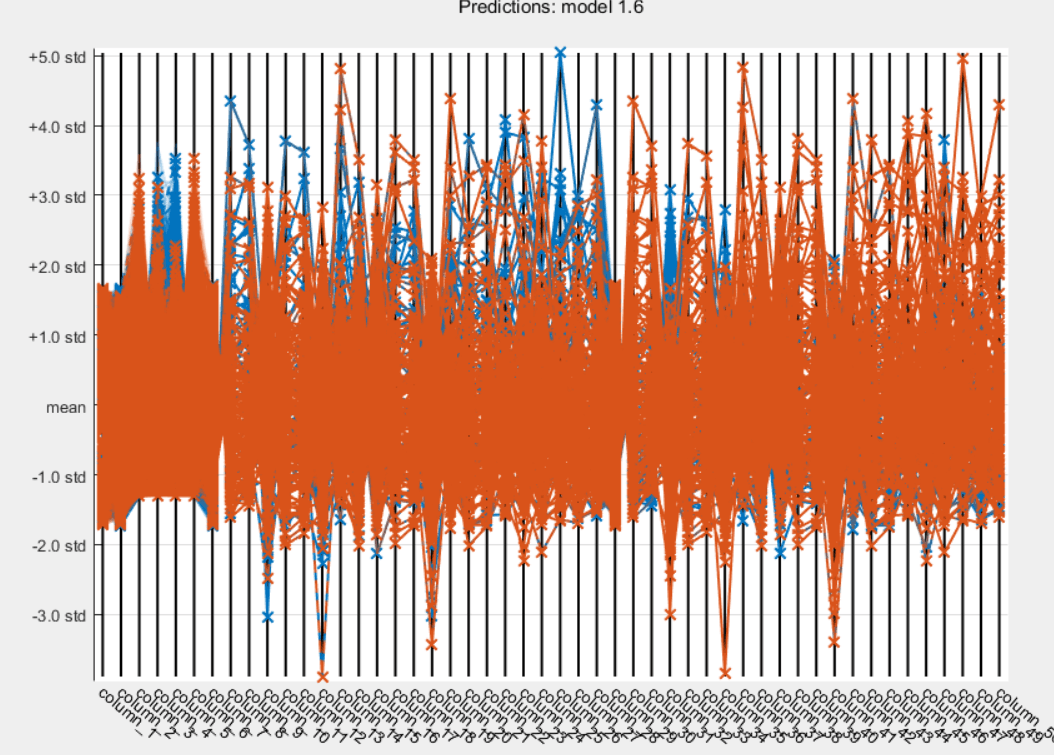
我们不使用PCA降维

logistic Regression：









# 5.模型参数的选择与优化思路

1.历史战绩修正：

对于构成50维向量的相应各维度权重进行修正，其中对历史战绩的相关参数进行 净胜场差值除以平均常数（作为主场对于客场的历史优势），，进行比分的修正，同时在最后进行预测的时候，对于比分修正，判定正负和置信度。

1. 主场优势修正
2. 球队参数修正