

# 问题一：建立各国奖牌数的预测模型

---

数据feature：前9届奖牌数量 参数人数 参数项目数 是否东道主（个人项目与团体项目？） *岭回归模型基本解决*

1. 预测金牌榜（给出预测区间？） 判断国家进步退步。eg线性回归 数值预测
2. 不确定性与模型性能（对比：ARIMA相比较） *任务：孙1.跑代码出结果。（模型数据 相关图 预测结果：榜单和具体值（明天做成美观的图）） 陈恺2.找背景知识和怎么写论文(模型建立和求解，模型检验)，用于明天。*  
思路：搜关键词：看一下有没有美赛好的论文用同样模型。去找论文，基于这个模型的应用。（他们是怎么介绍这个模型的背景，一两句体现模型优势的言论重点记下来可以作为引用）。完整的语句：文章名字作者

# 问题二：预测尚未获得奖的国家首次获奖的概率

---

*XGBoost用于解决小概率、样本较小的二分类问题，且将目标函数设为binary: logistic则输出为概率值，完美符合要求*

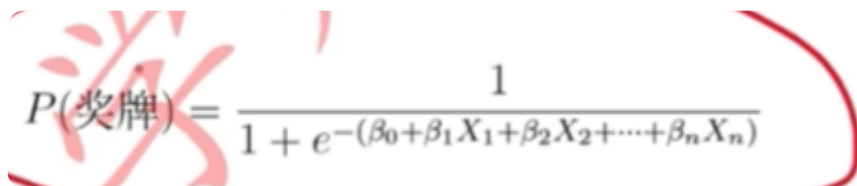
思路：用XGBoost模型来训练，feature：人数 项目数 已经参加的届数，输出：是否获奖。贝叶斯优化来寻找最佳超参数组合：调节：学习步长 弱学习器数量 树的最大深度等7个超参数）并采用交叉验证。

将准确率从80%提升到90% *优化后的模型准确率：0.9040*

最佳超参数组合：{'colsample\_bytree': 0.9050404201359931, 'gamma': 0.19563055435549098, 'learning\_rate': 0.13405901590902525, 'max\_depth': 2, 'min\_child\_weight': 1, 'n\_estimators': 2, 'subsample': 0.8

陈翔\*后续任务： 1.完善数据：扩充样本数量 陈恺2.学习论文机理XGboost和贝叶斯优化，看看怎么进一步优化调参数以及论文撰写。 3.2028年国家（参数人数 参数项目 已经参加的届数）（获得结果，同上） 拿奖牌概率为xxx

二分类：逻辑回归：概率 $0.5 >$ 可以 小于 $0.5 <$ 不能 \*


$$P(\text{奖牌}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

回归系数用最大似然估计 特征选择：运动员数量 参赛项目数量 历史成绩

概率模型：获得首枚奖牌概率，极值理论 生存分析 贝叶斯层次模型

## 问题三：项目设置（数量和类型）与奖牌数关系：

---

探究新增项目对东道主奖牌数的影响；各国有哪些擅长的运动项目？ \*\* 东道主选择项目 哪些运动对各国最重要

1：您的模型还应考虑特定奥运会的项目（数量和类型）。探索项目与国家获得奖牌数量之间的关系。 \*任务：

项目数量和各个类型sport作为feature： 关系就是探讨x对y的作用

1. 整理出来每届比赛每个国家参赛的Sport总数量和sport名单0 1 哪些项目对不同国家最重要？ 本国选择的项目对 结果有何影响？
2. \*整理数据：中国：足球 篮球 排球.....拿金银铜（分数？）以及参赛人数
3. 评价模型用于体现重要性
4. 统计新增项目的表现：可以直接分为东道主和其他国家。例如  
2024breaking霹雳舞：东道主拿牌情况与其他国家拿牌情况。后续处理思路：先统计所有的情况：暴力：东道主新增一个项目，平均多拿x 金 x 银。比例：法国其他的项目拿了10块金牌，其他一共有20项。法国作为东道主在新增项目上的拿牌率为xx，优于。

## 问题四：将教练因素考虑其中

---

原问题：研究数据，寻找可能由 "伟大教练"效 应引起的变化的证据。您估计这种效应对奖牌数的影响有多大？ 选择三个国家， 确定它们应考虑投资 "优秀"教练的体育项目，并估计其影响。

???

郎平和体操教练为例。用小概率模型，数据集：为体操和排球项目数据，数据只有：金银铜牌，参加了几届。郎平执教中美时的成绩为真实。??? 关注得到奖牌的教练在跳槽后执教的成绩（需要数据）

## 问题五：揭示其他奥运奖牌数规律

---

您的模型还揭示了哪些有关奥运奖牌数的独到见解？解释这些见解如何为各国奥委会提供信息。

idea：团体项目与个人项目。理解：个人项目对于小国家更易拿奖，看哪些项目获奖国家更新比较快的，建议小国重点投资（尤其是该国家也在历届比赛中派出选手参赛，具有一定的经验）陈翔：对数据集进行挖掘 找出第一次获奖的Sport都是哪些，然后看一下还未获奖的小国哪些是一直有在派人参赛的，给予推荐。