# 建模思路

2024.5.17

1. 总体思路

目的：根据不同考生分数、偏好等因素，为考生提供最优专业与学校选择建议。

衡量标准：各个选择（学校+专业）选择指数、效用值、效用概率、录取概率，使尽可能大。

目标函数：

#todo:完善录取概率计算

#为考生i高考分数排名，为该选择的招生人数，具体函数形式需要讨论

假设：

* 各个选择的各项客观指标不会变化随考生变化而变化，而各项主观指标则会随考生变化而变化
* 各个考生对各项指标的权重分配和不会随选择变化而变化
* 选择指数分为2个部分，效用概率和录取概率，效用概率满足MNL模型，录取概率满足？

1. 参数指标、设置与量化

说明：，大部分指标均需要根据其它参数进行折算

#todo每个指标的具体折算方法

* 1. 客观指标：

学校相关：

：学校排名指数（参考QS、泰晤士、USnews等）

：学校综合实力指数（参考学校等级（如985、211、双一流等）、学校科研经费、师资力量）

：学校地理位置（参考城市等级、交通便利程度、与家乡距离等）

#可以考虑把与家乡距离单独拎出来

：学校环境与设施情况（参考图书馆、宿舍、体育馆、实验室等）

：学校校友资源（参考知名校友数量、？等）

……

专业相关：

：专业就业压力（参考招聘市场需求量、毕业生数量、平均薪水等）

：专业学习难度与压力（参考主修课程平均成绩、学长评价等）

：专业发展潜力（参考平均薪水、国家政策支持度、专业跳槽率、毕业生去向分布等）

：专业社会影响力（参考？）

……

* 1. 主观指标：

：专业的兴趣与爱好

：家庭经济压力（参考学费、生活成本、消费成本、收入情况等）

……

1. 模型
   1. 调查数据
      1. 面向往届考生

数据用于训练模型、测试模型

原始数据：考生的最终录取去向、考生对各项指标的偏好情况和、考生的主观指标。其中根据问卷设置，

* + 1. 面向应届考生

数据用于模型应用、测试模型

原始数据：考生的最终录取去向、考生对各项指标的偏好情况和、考生的主观指标。其中根据问卷设置，

* 1. 既有数据
     1. 客观指标数据

待收集

* + 1. 各学校专业的招生人数和最低录取分数线

为该选择的最低分数排名，为该选择的招生总人数

* + 1. 考生排名和一分一段表
  1. 数据处理
     1. 效用模型
     2. 录取模型

#todo:完善录取模型函数

* + 1. 选择模型

#todo:完善选择模型函数

#todo:完善参数设定

涉及参数、、?

总方程满足：

通过神经网络、机器学习的方式（考虑RNN、LTSM等神经网络），根据训练集数据得出相关参数、、?

* + 1. 模型验证

根据(3)所得的参数集，代入验证集数据进行验证，根据验证结果再次迭代(3)(4)步，直至预测准确率足够高