

# Attention is All My Need

## 《人工智能与 Python 程序设计》期末建模报告

郭立为

中国人民大学经济学院

2025 年 6 月 4 日



中國人民大學

RENMIN UNIVERSITY OF CHINA

# 建模“两步法”

- 房屋价格的决定因素有很多，本次提供的数据文件中也包含异常丰富的特征信息
- 然而，这些信息并不是同等重要的（OLS 算法的局限性）
- 课程中学习了大量的建模方法，它们可以协助我判断哪些信息对于房屋价格有重要的影响
- 然而，机器是没有感情的，人类的经验和直觉往往是有效的

## 一个思考

地理位置在房屋价格的决定中有举足轻重的地位，其他任何因素都只是在此基础上对价格进行一定的增减。

# 建模"两步法"

## 建模的基本思路

- 分两步拟合模型
- **第一步**: 计算房屋所在版块房价的平均值 (实际是一个线性模型)
- **第二步**: 利用其他特征, 对房屋在平均值上下浮动的百分比进行估计 (更进一步的模型)

# 信息提取

## 利用 Word2Vec 处理文本信息

- 部分特征是描述性的而非定量的，相应特征也将会为购买者所直接阅读，因此适合使用文本处理方法
- 具体而言，采用 Word2Vec 工具，我们得以发现最重要的文本信息，从而将冗长的文本划分为特征变量

## 利用 PCA 处理高维信息

- 我们尽可能将连续变量都作为虚拟变量处理，由此可能造成信息维数过高的问题
- 利用主成分分析方法，可以提取出比较重要的信息，从而降低维数

# 第一步模型

- 预测变量：地理位置信息（城市、区域、板块）
- 模型类别：OLS、Lasso、Ridge、ElasticNet

## 模型表现

## 第二步模型

- 预测变量：除地理位置信息外的其他信息及他们与城市-区域虚拟变量的交乘项
- 模型类别：随机森林模型、神经网络模型

### 模型表现