

湖南大学实训报告

课程名称： 计算与人工智能概论 实验类型： 团队实训

实验项目名称： 基于各种回归模型预测 boston 房价

学生姓名： 王彬宇 班级： 数学 2402 班 学号： 202410040201

同组姓名： 王熙琳，周隆生 班级： 数学 2402 班 学号：

 202410040204,202410040202

一、实训目的

1. 巩固计算与人工智能概念课程所学基础知识
2. 拓展训练学生计算和 AI 思维能力
3. 加强团队分工与合作，培养学生团队协作能力

二、实训内容和团队分工

基于可数学分析与解释的各种回归模型对 **boston** 房价预测，项目由三人共同完成，无明确分工。

三、实训环境

我们采用标准数据分析库 **numpy,matplotlib,pandas** 以及机器学习库 **sklearn** 的各种子库，经尝试最终决定以 **boston** 房价作为标准数据集使用。

四、实训方法和步骤

我们采用学习、模仿、创新、总结的研究方法：

1. 学习资料，借助课上所学以及 **ai** 大模型理解各种回归模型及优化算法；

2. 从 `sklearn` 库导入 `boston` 数据集作为预测对象，并进行初步预测；
3. 分析比对不同回归模型性能，并不断优化参数，使均方误差减小；此外，对于与任务不适配的模型我们并没有摒弃，如 `logistic regression` 我们对 `target` 划区间分类预测，模型表现良好。
4. 分析不同模型的此任务中的性能优劣，进一步得出其使用范围，总结经验教训。

五、实训结果和分析

各种回归模型在预测 `boston` 房价任务上均有不错的表现，具体有以下几点：

- 一. 对比不同 `degree` 的多项式回归模型发现，`degree` 值较小，如 `degree = 1` 时，多项式回归退化为线性回归，模型形式简单，只能拟合线性关系。`degree` 值增大，模型能拟合更复杂的非线性关系，能更好地贴合训练数据。但过高的次数可能导致过拟合，模型在训练集上表现很好，在测试集或新数据上泛化能力差。在此任务中，`Degree = 2` 时均方误差最小。
- 二. 分位数回归像“给房价分组，每组单独算关系”，能看到低价、中间、高价房各自的规律，预测时更精准（尤其对极端价位），还能分析不同因素对各价位房价的影响。我们发现当 `q = 0.5` 时，即中位数回归均方误差最小，说明波士顿房价整体平稳，中价房较多，预测较为准确。
- 三. 随机森林回归、`xgboost` 等对复杂数据（如波士顿房价的多特征关系）拟合效果更好。
- 四. ~~~~
.....

六、讨论与心得

在讨论中，我们意识到没有绝对最优的模型。模型选择需综合考量数据特点、预测目标及计算资源等因素。同时，交叉验证、特征工程等环节对提升模型性能至关重要，我们在特征工程环节遇到了一些问题难以解决才转而使用 `boston` 房价标准数据集。通过此次小组讨论，我们不仅深入理解了各回归模型原理与应用，也体会到团队协作在机

器学习研究中集思广益、攻克难题的力量，为今后更复杂的建模任务积累了宝贵经验。