



中國人民大學

RENMIN UNIVERSITY OF CHINA

第12组模型展示（补充）

實事求是

日期：



2023200118 Midterm Exam Presentation

【处理思路】

- 1. 首先，运行完整的数据分析、数据清洗和数据预处理流程，实现了对复杂数据的处理。
- 2. 然后，进行了全面的特征工程，包括对数变换、多项式特征、交互项等。
- 3. 最后，从初始的基础线性模型出发，使用内置的交叉验证方法有效优化了正则化参数，综合评估了参数调优后的多种线性模型，选择最佳的线性回归模型。

【创新点】①将房价和房租两个任务进行一体化建模和预测；②采用了内置CV方法代替GridSearchCV；③引入了更多有意义的比率特征来增强模型性能。

Metrics	In sample	out of sample	Cross-validation	Kaggle Score
OLS	0.42	0.42	0.42	24.07
Ridge	0.42	0.42	0.42	24.07
LASSO	0.42	0.42	0.42	24.07
Elastic Net	0.44	0.44	0.44	18.10
Best Linear Model	0.41	0.41	0.41	26.95

(1) 去除异常值后的房价预测总数: 103,871
(2) 去除异常值后的房价预测总数: 98,899