

实验二：数据预处理的基本方法

2.1 实验目的

理解数据分析的基本过程，使学生掌握基本的数据预处理方法。

2.2 实验内容及要求

【内容描述】利用给出的房屋价格数据集，进行以下任务：

- (1) 缺失值的检测与缺失值处理；
- (2) 异常值检测；
- (3) 特征间的相关性分析；
- (4) 对 price 属性进行标准化；
- (5) 根据 price 属性进行离散化；
- (6) 找出与 price（房价）相关性最高的三个特征，并给出合理的解释。

【实验要求】建议使用 Jupyter Notebook 完成该任务，给出完成任务的过程。

2.3 实验环境

采用 python 语言并利用 Jupyter Notebook 编程实现。

2.4 评分标准

满分 10 分		
程序 8 分	缺失值的检测是否全面	1 分
	缺失值的处理方式是否更加符合实际问题（例如选取 KNN、树等处理方式填充缺失值等等）	1 分
	异常值检测是否全面	1 分
	处理异常值的方式是否符合实际问题（例如选取平均值等）	2 分

	能否提取出关键特征删去冗余特征	2
	能否对数据进行合适的标准化	1
文档 2 分	是否对程序代码有清晰的解释，并且描述自己的分析结论	1 分
	具有更加深度和发散的思维，思考数据内部之间存在的隐藏的逻辑关系	1 分