

数据预处理(Data Preprocessing)与

特征工程(Features Engineering)思考题

1. 数据预处理(Data Preprocessing)的目的/目标是什么？评价数据集质量的主要指标有哪些？
2. 数据清洗(Data Cleaning)的内容是什么？处理缺失数据的方法有哪些？歧义值与异常值处理方法有哪些？
3. 数据集成(Data Integration) 的含义是什么？数据集成的常用方法及含义是什么？
4. 什么是数据的特征(Feature)？特征工程(Features Engineering)的动机与目标是什么？
5. 特征提取(Features Selection)的方法有哪些？它们的原理和实施步骤是什么？
6. 特征生成(Features Generating)的方法有哪些？它们的原理和实施步骤是什么？
7. 数据约简(Data Reduction)的目的是什么？数据约简的方法有哪些？它们的作用分别是什么？
8. 数据变换(Data Transformation)的目标是什么？数据标准化的方法及其特征是什么？偏态分布的处理方法及其特征是什么？连续变量的离散化方法是什么？常用特征编码方法及适用范围？