第4章 数据预处理

教案

**课程名称：**Python数据分析与挖掘实战

**课程类别：**必修

**适用专业：**大数据技术类相关专业

**总学时：**112学时（其中理论49学时，实验63学时）

**总学分：**7.0学分

**本章学时**：6学时

# 材料清单

* 1. 《Python数据分析与挖掘实战》教材。
  2. 配套PPT。
  3. 引导性提问。
  4. 探究性问题。
  5. 拓展性问题。

# 教学目标与基本要求

### 教学目标

介绍数据分析的数据预处理过程，即数据清洗、数据集成、数据变换和数据归约。数据清洗介绍对缺失值和异常值的处理。数据集成介绍将多个数据源中的数据合并存放到一个数据存储的过程。数据变换介绍了如何从不同的应用角度对已有属性进行函数变换；数据归约从属性（纵向）归约和数值（横向）归约两个方面介绍了如何对数据进行归约，使挖掘的性能和效率得到很大的提高。

### 基本要求

1. 掌握Python中数据清洗的方法。
2. 掌握Python中数据集成的方法。
3. 掌握Python中数据变换的方法。
4. 掌握Python中数据归约的方法。
5. 了解Python主要数据预处理函数。

# 问题

### 引导性提问

引导性提问需要教师根据教材内容和学生实际水平，提出问题，启发引导学生去解决问题，提问，从而达到理解、掌握知识，发展各种能力和提高思想觉悟的目的。

* 1. 数据质量不高的情况下如何提高数据质量？
  2. 数据预处理包含哪些内容？

### 探究性问题

探究性问题需要教师深入钻研教材的基础上精心设计，提问的角度或者在引导性提问的基础上，从重点、难点问题切入，进行插入式提问。或者是对引导式提问中尚未涉及但在课文中又是重要的问题加以设问。

1. 数据预处理各个步骤是否有先后？
2. 数据变换的目的是什么？

### 拓展性问题

拓展性问题需要教师深刻理解教材的意义，学生的学习动态后，根据学生学习层次，提出切实可行的关乎实际的可操作问题。亦可以提供拓展资料供学生研习探讨，完成拓展性问题。

1. 数据清洗除了缺失值处理和异常值处理外，还能有什么操作？
2. 能否将数据预处理当中的方法写成自定义函数?

# 主要知识点、重点与难点

### 主要知识点

1. 数据清洗的方法。
2. 数据集成的方法。
3. 数据变换的方法。
4. 数据归约的方法。
5. 主要的数据预处理函数。

### 重点

1. 数据清洗的方法。
2. 数据集成的方法。
3. 数据变换的方法。
4. 数据归约的方法。

### 难点

1. 数据变换的方法。
2. 数据归约的方法。

# 教学过程设计

### 理论教学过程

1. 缺失值、异常值处理方法。
2. 常见的数据集成的方法。
3. 数据规范化和离散化的方法。
4. 属性构造的方法。
5. 小波变换的方法。
6. 属性归约和数值归约的方法。
7. Python主要数据预处理函数。

### 实验教学过程

1. 使用Python进行数据清洗。
2. 使用Python进行数据集成。
3. 使用Python进行数据变换。
4. 使用Python进行数据归约。

# 教材

### 教材

张良均，谭立云，刘名军，江建明．《Python数据分析与挖掘实战（第2版）》[M]．北京：机械工业出版社．2019．