第5章 挖掘建模

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**：24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

根据数据挖掘技术的基本任务，即分类与预测、聚类、关联规则、时序模式、离群点检测五个方面，介绍了对应的数据挖掘建模方法及实现过程。分类与预测主要介绍决策树和人工神经网络两个分类模型、回归分析预测模型及其实现过程；聚类主要介绍K-Means聚类算法，建立分类方法按照接近程度对观测对象给出合理的分类并解释类与类之间的区别；关联规则主要介绍Apriori算法，以在一个数据集中找出各项之间的关系；时序模式从序列的平稳性和非平稳型出发，对平稳时间序列主要介绍ARMA模型，对差分平稳序列建立ARIMA模型，应用这两个模型对相应的时间序列进行研究，找寻变化发展的规律，预测将来的走势；离群点检测主要介绍基于模型和离群点的检测方法，是发现与大部分其他对象显著不同的对象。

### 基本要求

1. 掌握Python中分类与预测的方法。
2. 掌握Python中聚类分析的方法。
3. 掌握Python中关联规则分析的方法。
4. 掌握Python中时序模式的分析方法。
5. 掌握Python中离群点的检测方法。

# 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. 常见的模型算法使用场景有哪些？
  2. 关联规则算法有哪些？

### 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. 回归模型有几种，它们各自适用于怎样的条件？
2. 聚类和分类的区别是什么？

### 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. 不同的算法，可解释性不同，能否挑选一种可解释性强的算法对算法结果进行解释？
2. 聚类算法的评价指标还有哪些？

# 主要知识点、重点与难点

### 主要知识点

1. 常用的分类与预测算法。
2. 分类与预测算法基本原理及实现过程。
3. 常用的聚类分析算法。
4. 聚类分析算法简介及基本原理。
5. 聚类分析算法实现过程及评价。
6. 常用关联规则算法。
7. Apriori算法的原理与实现过程。
8. 常见的时间序列模型。
9. 时间序列的预处理。
10. 平稳时间序列分析。
11. 非平稳时间序列分析。
12. 常用的离群点检测方法。
13. 基于模型的离群点检测方法。
14. 基于聚类的离群点检测方法。

### 重点

1. 分类与预测算法基本原理及实现过程。
2. 聚类分析算法简介及基本原理。
3. 聚类分析算法实现过程及评价。
4. Apriori算法的原理与实现过程。
5. 时间序列的预处理。
6. 平稳时间序列分析。
7. 非平稳时间序列分析。
8. 基于模型的离群点检测方法。
9. 基于聚类的离群点检测方法。

### 难点

1. 分类与预测算法基本原理及实现过程。
2. 聚类分析算法简介及基本原理。
3. Apriori算法的原理与实现过程。
4. 平稳时间序列分析。
5. 非平稳时间序列分析。
6. 基于模型的离群点检测方法。
7. 基于聚类的离群点检测方法。

# 教学过程设计

### 理论教学过程

1. 常用的分类与预测算法。
2. 回归分析。
3. 决策树的原理。
4. 人工神经网络的原理。
5. 分类与预测算法的评价方法。
6. Python分类预测模型特点。
7. 常用聚类分析算法。
8. K-Means聚类算法的原理。
9. 聚类分析算法的评价方法。
10. Python主要聚类分析算法。
11. Apriori算法的原理。
12. 时间序列预处理的方法。
13. 平稳时间序列分析。
14. 非平稳时间序列分析。
15. Python主要时序模式算法。
16. 离群点检测方法。
17. 基于模型的离群点检测方法。
18. 基于聚类的离群点检测方法。

### 实验教学过程

1. 使用逻辑回归算法进行分类预测。
2. 使用决策树算法进行分类预测。
3. 使用人工神经网络算法进行分类预测。
4. 使用K-Means进行聚类分析。
5. 使用TSNE进行数据降维并展示聚类结果。
6. 使用Apriori算法进行关联分析。
7. 使用ARIMA模型进行时间序列分析。
8. 使用Python离散点检测。

# 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．