



中山大學南方學院

NANFANG COLLEGE OF SUN YAT-SEN UNIVERSITY

[课程名称] 大数据分析与应用
本科课程教学大纲 (文学、外语、经济、管理、
艺术类)

| 课程信息 | | | |
|--------|--|--------|----------------|
| 开课单位： | 商学院 | 开课学年学期 | 2018-2019 第二学期 |
| 授课年级： | 2016 级 | 授课对象专业 | 经济学 |
| 课程学分 | 3 | 课程学时 | 60 |
| 课程性质 | <input type="checkbox"/> 专业必修 <input type="checkbox"/> 专业选修 <input type="checkbox"/> 公共必修 <input type="checkbox"/> 公共选修 <input type="checkbox"/> 成长必修 <input checked="" type="checkbox"/> 专业限选 <input type="checkbox"/> 公共限选 | | |
| 先修课程要求 | 计算机基础, 统计学 | | |
| 教师信息 | | | |
| 授课教师 | 陈放 | 联系电话 | 1591878369 |
| 答疑地点 | 8 教-110 | 答疑时间 | |
| 电子邮箱： | math_chen@126.com | | |

主 审:

1 课程描述

1. 大数据时代已经到来，在商业、经济及其他领域中基于数据和分析去发现问题并做出科学、客观的决策越来越重要。数据分析技术将帮助企业用户在合理时间内获取、管理、处理以及整理海量数据，为商业决策提供积极的帮助。数据分析作为一门前沿技术，广泛应用于物联网、云计算、移动互联网等战略新兴产业。有实践经验的数据分析人才已经成为了各企业争夺的热门。为了推动商学院大数据，人工智能行业的发展，满足日益增长的数据分析人才需求，特开设大数据分析与应用课程，大数据不能空谈，因此本课程所采用的软件为当前使用非常广泛和免费的 python 软件为工具，通过理论与实践结合的方式讲授该课程。
2. 通过本课程的学习，使学生学会使用 Python 进行科学计算、可视化绘图、数据处理，分析与建模，并详细拆解学习聚类、回归、分类三个企业案例，将理论与实践相结合，为将来从事商业领域数据分析挖掘研究、工作，奠定基础。
3. 考核方式：期末试卷成绩 *40%+ 平时成绩 *60%
 - (a) 出勤率：10%
 - (b) 课堂提问：10%
 - (c) 上机练习：30%
 - (d) 作业：10%
 - (e) 期末考试：40%

2 课程目标

第 1 章 Python 数据分析概述

- 主要知识点**
1. 数据分析的概念、流程与应用场景。
 2. Python 常用的数据分析库。
 3. Windows/Linux 系统下 Anaconda 安装。
 4. Jupyter Notebook 的常用功能。

教学要求 1. 了解数据分析的概念。

2. 了解数据分析的流程。

3. 了解数据分析在实际中的应用。

4. 了解 Python 在数据分析的优势。

5. 了解 Python 常用的数据分析库。

6. 掌握 Windows / Linux 系统下 Anaconda 的安装。

教学重点与难点 重点: (1) 数据分析的概念、流程与应用场景。(2) Jupyter Notebook 的常用功能. 难点: 数据分析的概念与流程.

教学方法 课堂展示, 教学条件使用多媒体教学, 本章教学方法要注重理论和实际操作。

问题 1. 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. 数据分析能够做什么?

2. 现实生活中存在哪些数据分析技术?

3. 该如何进行数据分析?

4. Python 语言有哪些优势?

5. Pycharm 这个 PythonIDE 有那些优点和缺点?

2. 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. 数据分析的完整流程是怎样的?

2. 数据分析的能够应用在那些场景?

3. Jupyter Notebook 的哪些优缺点?

3. 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. 数据分析是不是万能的?

2. Python 语言用于做数据分析有哪些优势?

3. 分析本班人员的基本信息可以从哪些角度入手?

- 讲授教学环节**
1. 数据分析的概念。
 2. 数据分析的流程。
 3. 数据分析的应用场景。
 4. 数据分析的常用工具。
 5. Python 数据分析的优势。
 6. Python 数据分析的常用类库。
 7. Python 的 Anaconda 发行版。
 8. 在 Windows 操作系统上安装 Anaconda。
 9. 在 Linux 系统上安装 Anaconda。
 10. Jupyter Notebook 的基础功能。
 11. Jupyter Notebook 的高级功能。
- 实践教学部分**
1. 在 Windows/Linux 系统上安装 Anaconda。
 2. Jupyter Notebook 的常用功能。

第 2 章 NumPy 数值计算基础

- 主要知识点**
1. NumPy 创建多维数组与生成随机数的方法。
 2. 数组的索引与转换。
 3. NumPy 中数组矩阵的运算及通用函数的基本使用方法。
 4. NumPy 统计分析的方法。
- 教学要求**
1. 掌握 NumPy 创建多维数组与生成随机数的方法。
 2. 掌握数组的索引与转换。
 3. 了解 NumPy 中数组矩阵的运算及通用函数的基本使用方法。
 4. 掌握 NumPy 统计分析的方法。

教学重点与难点 重点：（1）ndarray 对象的创建与生成随机数的方法。（2）数组的索引与变换。（3）矩阵的运算及通用函数的基本使用方法。难点：（1）ndarray 对象的创建与生成随机数的方法。（2）数组的索引与变换。

教学方法 课堂展示，教学条件使用多媒体教学，本章教学方法要注重理论和实际操作。

问题 1. 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. NumPy 从这个库的名字理解，这个库的作用是什么？
2. 数组内部存放的数据是否可以为字符串？
3. 怎样取出数组内部的某个元素？
4. 数组和矩阵是否可以转换

2. 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. 数组和矩阵的异同点？
2. 数组和矩阵能否互相转换？
3. 能否自定义 ufunc？

3. 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. 使用 NumPy 能否进行本班人员的基本信息？
2. 如何将所有的分析方法做成一个 ufunc？

讲授教学环节

1. 创建数组对象。
2. 生成随机数。
3. 通过索引访问数组。
4. 变换数组的形态。
5. 创建 NumPy 矩阵。
6. ufunc 函数。
7. 读写文件。
8. 使用数组进行简单的统计分析。

- 实践教学部分**
1. 创建 NumPy 数组对象 ndarray。
 2. 查看 ndarray 的常用属性。
 3. 花式索引 ndarray。
 4. 变换 ndarray 的形态。
 5. 创建 NumPy 矩阵并使用。
 6. 使用常见 ufunc。
 7. 使用 NumPy 读写文件。

第 3 章 Matplotlib 数据可视化基础

- 主要知识点**
1. pyplot 常用绘图参数调节。
 2. 子图的绘制方法。
 3. 散点图和折线图的作用与绘制方法。
 4. 直方图和饼图绘制的作用与绘制方法。
 5. 箱线图和茎叶图的作用与绘制方法。

- 教学要求**
1. 掌握 pyplot 常用绘图参数调节。
 2. 掌握子图的绘制方法。
 3. 掌握散点图和折线图的作用与绘制方法。
 4. 掌握直方图和饼图绘制的作用与绘制方法。
 5. 掌握箱线图和茎叶图的作用与绘制方法。

教学重点与难点 重点：（1）pyplot 的基础语法。（2）散点图和折线图的作用与绘制方法。（3）直方图、饼图和箱线图的作用与绘制方法。难点：（1）子图的绘制方法。（2）散点图和折线图的作用与绘制方法。（3）直方图、饼图和箱线图的作用与绘制方法。

教学方法 课堂展示，教学条件使用多媒体教学，本章教学方法要注重理论和实际操作。

问题 1. 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. 图形和文字哪一种更容易让大脑记住？
2. 常见的统计学图形有哪些？
3. 不同的图形所表示的意义是否相同？

2. 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. 散点图，折线图的主要功能是什么，有什么异同点？
2. 饼图，直方图的主要功能是什么，有什么异同点？
3. 箱型图的主要功能是什么？

3. 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. 能否绘制一个雷达图，该怎么做？
2. 如何将本班同学的数据做可视化？

讲授教学环节 1. pyplot 的基础语法。

2. 设置 pyplot 的动态 rc 参数。

3. 绘制散点图。

4. 绘制折线图。

5. 绘制直方图。

6. 绘制饼图。

7. 绘制箱线图。

- 实践教学部分**
1. 设置 pyplot 的动态 rc 参数。
 2. 绘制散点图；绘制折线图。
 3. 绘制直方图。
 4. 绘制饼图。
 5. 绘制箱线图。

第 4 章 pandas 统计分析基础

- 主要知识点**
1. 常见的数据读取方式。
 2. DataFrame 常用属性与方法。
 3. 基础时间数据处理方法。
 4. 分组聚合的原理与方法。
 5. 透视表与交叉表的制作。

- 教学要求**
1. 掌握常见的数据读取方式。
 2. 掌握 DataFrame 常用属性与方法。
 3. 掌握基础时间数据处理方法。
 4. 掌握分组聚合的原理与方法。
 5. 掌握透视表与交叉表的制作。

教学重点与难点 重点：（1）读写数据。（2）DataFrame 数据的常用属性与方法。（3）基础时间数据处理方法。（4）分组聚合的原理与方法。（5）透视表与交叉表的制作。
难点：（1）读写数据。（2）分组聚合的原理与方法。（3）透视表与交叉表的制作。

问题 1. 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. 数据分析中常见的结构化数据读取方式有哪些？
2. DataFrame 和数组有什么相似之处？
3. Excel 透视表如何制作？

2. 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. 读取数据库数据为什么需要别的库？
2. 时间数据中存在哪些信息？
3. 以百万级的数据为例，Excel 制作透视表和 Python 制作透视表哪个速度更快？

3. 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. 能否读取 word 数据，该如何做？
2. 为什么索引的时候有 loc 和 iloc，设计者的意何在？

教学方法 课堂展示，教学条件使用多媒体教学，本章教学方法要注重理论和实际操作。

讲授教学环节

1. 读写数据库数据。
2. 读写文本文件。
3. 读写 Excel 文件。
4. 查看 DataFrame 的常用属性。
5. 查改增删 DataFrame 数据。
6. 描述分析 DataFrame 数据。
7. 转换字符串时间为标准时间。
8. 提取时间序列数据信息。
9. 加减时间数据。
10. 使用 groupby 方法拆分数据。
11. 使用 agg 方法聚合数据。
12. 使用 apply 方法聚合数据。
13. 使用 transform 方法聚合数据。
14. 使用 *pivot_table* 函数创建透视表。
15. 使用 *crosstab* 函数创建交叉表实验教学过程。

实践教学部分

1. 读写数据库数据。
2. 读写文本文件。
3. 读写 Excel 文件。
4. 查看 DataFrame 的常用属性。
5. 查改增删 DataFrame 数据。
6. 描述分析 DataFrame 数据。
7. 转换字符串时间为标准时间。
8. 提取时间序列数据信息。
9. 加减时间数据。
10. 使用 groupby 方法拆分数据。
11. 使用 agg, apply, transform 方法聚合数据。
12. 制作透视表。
13. 制作交叉表。

第 5 章 使用 pandas 进行数据预处理

主要知识点

1. 数据清洗的基本方法。
2. 数据合并的原理与方法。
3. 数据标准化的概念与方法。
4. 常用的数据变换方法。

教学要求

1. 掌握数据清洗的基本方法。
2. 掌握数据合并的原理与方法。
3. 掌握数据标准化的概念与方法。
4. 掌握常用的数据变换方法。

教学重点与难点 重点：（1）数据合并。（2）数据清洗。（3）数据标准化。（4）数据转换。难点：（1）数据清洗。（2）数据转换。

问题 1. 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?
3. 数据预处理各个步骤是否有先后?

2. 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. 重复值是否一定要做清洗?
2. 时间数据中存在哪些信息?
3. 数据变换的目的是什么?

3. 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. 除了哑变量处理外, 还有那些方法可以处理离散型特征?
2. 能否将这些方法写成自定义函数?

教学方法 课堂展示，教学条件使用多媒体教学，本章教学方法要注重理论和实际操作。

- 讲授教学环节**
1. 堆叠合并数据。
 2. 主键合并数据。
 3. 重叠合并数据。
 4. 检测与处理重复值。
 5. 检测与处理缺失值。
 6. 检测与处理异常值。
 7. 离差标准化数据。
 8. 标准差标准化数据。
 9. 小数定标标准化数据。
 10. 哑变量处理类别型数据。
 11. 离散化连续型数据。

- 实践教学部分**
1. 堆叠、主键、重叠合并数据。
 2. 检测与处理重复值，缺失值，异常值。
 3. 离差标准化、标准差标准化。
 4. 小数定标标准化数据。
 5. 哑变量处理类别型数据。
 6. 离散化连续型数据。

第 6 章 使用 sklearn 构建模型

- 主要知识点**
1. sklearn 转换器，评估器的使用。
 2. sklearn 数据标准化与数据划分。
 3. sklearn 中聚类，分类，回归模型的构建。
 4. sklearn 中聚类，分类，回归模型的评价。

- 教学要求**
1. 掌握 sklearn 转换器，评估器的使用。
 2. 掌握 sklearn 数据标准化与数据划分。
 3. 掌握 sklearn 中聚类，分类，回归模型的构建。
 4. 掌握 sklearn 中聚类，分类，回归模型的评价。

教学重点与难点 重点: (1) sklearn 转换器的使用方法。(2) sklearn 估计器的使用方法。(3) 聚类模型的构建与评价。(4) 分类模型的构建与评价。(5) 回归模型的构建与评价。难点: (1) sklearn 转换器的使用。(2) sklearn 估计器的使用。(3) 分类模型构建。(4) 回归模型构建。

问题 1. 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. 不同的场景之间有什么区别?
3. scikit-learn 名字的由来?

2. 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. scikit-learn 的标准化和第 5 章的有何不同?
2. 聚类和分类的区别是什么?
3. 回归和分类的区别又是什么?

3. 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. 除了 PCA 降维以外,可以使用 RandomForest 模型进行特征的筛选,该如何做?

教学方法 课堂展示,教学条件使用多媒体教学,本章教学方法要注重理论和实际操作。

- 讲授教学环节**
1. 加载 datasets 模块中的数据集。
 2. 将数据集划分为训练集和测试集。
 3. 使用 sklearn 转换器进行数据预处理与降维。
 4. 使用 sklearn 估计器构建聚类模型。
 5. 评价聚类模型。
 6. 使用 sklearn 估计器构建分类模型。
 7. 评价分类模型。
 8. 使用 sklearn 估计器构建回归模型。
 9. 评价回归模型。

- 实践教学部分**
1. 加载 datasets 模块自带数据集。
 2. 划分数据集。
 3. 使用 sklearn 转换器进行数据预处理与降维。
 4. 构建与评价聚类模型。
 5. 构建与评价分类模型。
 6. 构建与评价回归模型。

第 7 章 航空公司客户价值分析

- 主要知识点**
1. RFM 模型的基本原理。
 2. K-Means 算法的基本原理与使用方法。
 3. 比较不同类别客户的客户价值，制定相应的营销策略。

- 教学要求**
- (a) 了解 RFM 模型的基本原理。
 - (b) 掌握 K-Means 算法的基本原理与使用方法。
 - (c) 比较不同类别客户的客户价值，制定相应的营销策略。

教学重点与难点 重点：（1）航空客户价值分析的步骤和流程。（2）RFM 模型的基本原理。（3）K-Means 算法的基本原理与使用方法。（4）比较不同类别客户的客户价值。难点：（1）RFM 模型的基本原理。（2）KMeans 算法的基本原理与使用方法。

问题 1. 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. 什么样的客户是航空公司的高价值客户？
2. 航空公司能够提供什么样的数据？
3. 为什么航空公司要做客户价值分析？

2. 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. RFM 模型对应的特征在本案例中是什么？
2. 为什么 RFM 模型不能够很好区分用户？
3. 为何 KMeans 可以用作客户分类？

3. 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. 除了 RFM 模型外还有什么用于客户分类的模型？
2. 除了 KMeans 还有什么算法能够用于客户分类？

教学方法 课堂展示，教学条件使用多媒体教学，本章教学方法要注重理论和实际操作。

讲授教学环节

1. 分析航空公司现状。
2. 认识客户价值分析。
3. 熟悉航空客户价值分析的步骤与流程。
4. 处理缺失值与异常值。
5. 构建爱你航空客户价值分析关键特征。
6. 标准化 LRFMC 5 个特征。
7. 了解 K-Means 聚类算法。
8. 分析聚类结果。
9. 模型应用。

- 实践教学部分**
1. 处理数据缺失值与异常值。
 2. 构建航空客户价值分析的关键特征。
 3. 标准化 LRFMC 5 个特征。
 4. 构建 K-Means 聚类模型。
 5. 评价 K-Means 聚类模型。

第 8 章 财政收入预测分析

- 主要知识点**
1. 财政收入预测的背景知识，分析步骤和流程。
 2. 相关性分析方法与应用。
 3. 用 Lasso 模型特征选取方法。
 4. 灰色预测和支持向量回归算法的原理与应用

- 教学要求**
1. 了解财政收入预测的背景知识，分析步骤和流程。
 2. 掌握相关性分析方法与应用。
 3. 掌握用 Lasso 模型特征选取方法。
 4. 掌握灰色预测和支持向量回归算法的原理与应用。

教学重点与难点 重点：（1）财政收入预测的步骤和流程。（2）相关性分析方法与应用。（3）使用 Lasso 模型选取特征。（4）灰色预测算法的原理与使用。（5）支持向量回归算法的原理与使用。难点：（1）使用 Lasso 模型选取特征。（2）灰色预测算法的原理与使用。（3）支持向量回归算法的原理与使用。

问题 1. 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. 市财政收入的构成是什么？
2. 影响财政收入的相关因素有哪些？
3. 市财政收入预测的意义在哪里？

2. 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. 相关性分析的使用场景有哪些？
2. Lasso 回归使用场景有哪些？
3. 为何要提取关键特征？

3. 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. 除了 SVR 还有很多回归算法，能否使用其他回归算法解决该需求？
2. 国家数据网有很多类似数据，能否预测某个省的财政收入呢？

教学方法 课堂展示，教学条件使用多媒体教学，本章教学方法要注重理论和实际操作。

- 讲授教学环节**
1. 分析财政收入预测背景。
 2. 了解财政收入预测的方法。
 3. 熟悉财政收入预测的步骤与流程。
 4. 了解相关性分析。
 5. 分析计算结果。
 6. 了解 Lasso 回归方法。
 7. 分析 Lasso 回归结果。
 8. 了解灰色预测算法。
 9. 了解 SVR 算法。
 10. 分析预测结果。

- 实践教学部分**
1. 分析财政收入数据特征的相关性。
 2. 使用 Lasso 回归选取财政收入预测的关键特征。
 3. 使用灰色预测和 SVR 构建财政收入预测模型。
 4. 评价 SVR 模型。

第 9 章 家用热水器用户行为分析与事件识别

- 主要知识点**
1. 掌握用水事件划分原理及与方法。
 2. 掌握阈值寻优的原理和方法。
 3. 熟悉用水行为特征构建的原理与方法。
 4. 了解神经网络算法的基本原理。
 5. 掌握使用 sklearn 神经网络算法构建。
 6. 掌握分类算法的评价方法。

- 教学要求**
1. 了解财政收入预测的背景知识，分析步骤和流程。
 2. 掌握相关性分析方法与应用。
 3. 掌握用 Lasso 模型特征选取方法。
 4. 掌握灰色预测和支持向量回归算法的原理与应用。

教学重点与难点 重点：（1）家用热水器用户行为分析的步骤与流程。（2）用水事件划分原理及与方法。（3）阈值寻优的原理和方法。（4）用水行为特征构建的原理与方法。（5）BP 神经网络算法的原理与使用方法。难点：（1）阈值寻优的原理和方法。（2）BP 神经网络算法的原理与使用方法。

问题 1. 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. 什么样的电器可以称得上智能的？
2. 智能电器具备怎样的特征？
3. 智能热水器有需要怎样的条件，需要哪些数据？

2. 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. 如何划分用水事件
2. 如何确定用水时长？
3. 不同的用水时长可能代表了什么意义？

3. 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. 除了热水器还有什么日常生活电器可以智能化？
2. 做到这些电器智能化需要哪些数据，分析的核心因素是什么？

教学方法 课堂展示，教学条件使用多媒体教学，本章教学方法要注重理论和实际操作。

讲授教学环节

1. 分析家用热水器行业现状。
2. 了解热水器采集数据的基本情况。
3. 熟悉家用热水器用户行为分析的步骤与流程。
4. 删除冗余特征。
5. 划分用水事件。
6. 确定单次用水事件时长阈值。
7. 构建用水时长与频率特征。
8. 了解灰色预测算法。
9. 构建用水量与波动特征。
10. 筛选候选洗浴事件。
11. 了解 BP 神经网络算法原理。
12. 构建模型。
13. 评估模型。

实践教学部分

1. 删除冗余特征。
2. 划分用水事件。
3. 确定单次用水事件的时长阈值。
4. 构建用水行为特征。
5. 筛选候选洗浴事件。
6. 构建 BP 神经网络模型。
7. 评价 BP 神经网络模型。

3 课程资料:

1. 教科书

黄红梅, 张良均. Python 数据分析与应用 [M]. 北京: 人民邮电出版社. 2018.

2. 参考书目

(a) 张良均. Python 数据分析与挖掘实战 [M]. 北京: 机械工业出版社. 2015.

(b) 张良均. Python 与数据挖掘 [M]. 北京: 机械工业出版社. 2016.

3. 电子文献链接

(a) <http://www.diveintopython3.net/>

(b) <https://www.cnblogs.com/dennis-liucd/p/7669161.html>

(c) <https://docs.python.org/3/>

4 教学进度计划表

| 周次 | 课程要点 | 教学活动 | 阅读资料 |
|----|----------------------------|---|------------|
| 1 | 第 1 章 Python 数据分析概述 | 数据分析的概念、流程、python 安装 | 教材及参考书相关内容 |
| 2 | 第 2 章 NumPy 科学计算基础 | 生成随机数, 创建矩阵, 使用通用函数计算, 以及利用 NumPy 进行数据分析的常用函数 | 教材及参考书相关内容 |
| 3 | 第 3 章 Matplotlib 可视化基础 (1) | pyplot 绘图的基本语法, 常用参数 | 教材及参考书相关内容 |
| 4 | 第 3 章 Matplotlib 可视化基础 (2) | 散点图、折线图、直方图、饼图、箱线图 | 教材及参考书相关内容 |
| 5 | 第 4 章 pandas 统计分析基础 (1) | DataFrame 的常用属性, 方法与描述性统计相关内容 | 教材及参考书相关内容 |
| 6 | 第 4 章 pandas 统计分析基础 (2) | 时间数据的转换, 信息提取与算术运算。剖析分组聚合方法 groupby 的原理 | 教材及参考书相关内容 |
| 7 | 第 5 章使用 pandas 进行数据预处理 (1) | 数据清洗、数据合并、数据标准化和数据转换 | 教材及参考书相关内容 |
| 8 | 第 5 章使用 pandas 进行数据预处理 (2) | 数据合并、数据标准化、数据转换 | 教材及参考书相关内容 |
| 9 | 第 6 章使用 sklearn 构建模型 (1) | 构建并评价聚类模型 | 教材及参考书相关内容 |

| | | | |
|----|---------------------------|----------------------------------|----------------|
| 10 | 第 6 章使用 sklearn 构建模型 (2) | 构建并评价分类模型 | 教材及参考书 相关内容 |
| 11 | 第 6 章使用 sklearn 构建模型 (3) | 构建并评价回归模型 | 教材及参考书 相关内容 |
| 12 | 第 7 章航空公司客户价值分析 (1) | K-Means 聚类算法在客户价值分析中的应用 | 教材及参考书 相关内容 |
| 13 | 第 7 章航空公司客户价值分析 (2) | 使用 K-Means 算法构建航空客户价值分析 LRFMC 模型 | 教材及参考书 相关内容 |
| 14 | 第 8 章财政收入预测分析 (1) | 财政收入预测的背景与方法、数据的特征 | 教材及参考书 相关内容 |
| 15 | 第 8 章财政收入预测分析 (2) | lasso 回归、灰色预测和 SVR 构建财政收入预测模型 | 教材及参考书 相关内容 |
| 16 | 第 9 章家用热水器用户行为分析与事件识别 (1) | 家用热水器用户行为分析的背景与步骤、数据预处理 | 教材及参考书 相关内容 |
| 17 | 第 9 章家用热水器用户行为分析与事件识别 (2) | 数据特征构建及筛选、BP 模型 | 教材及参考书 相关内容 |
| 18 | 复习 | | |

5 教学活动与预期成果的评估

1. 教学活动:

- (a) 课堂讲授
- (b) 案例分析
- (c) 课堂问答

2. 对预期学习成果的考察

| 预期学习成果 | 教学活动 | 评估方式 |
|----------------------|---------|-----------|
| 掌握 python 软件的基本编程 | 课堂讲解和实训 | 平时 + 期末成绩 |
| 会用 python 常用的数据分析包使用 | 课堂讲解和实训 | 平时 + 期末成绩 |
| 掌握教材中三个数据分析案例 | 课堂讲解和实训 | 上机实操 |

6 评分体系与标准（评分项目不少于 3 项）

1. 评分体系

- (a) 出勤率：10%
- (b) 课堂提问：10%
- (c) 上机练习：30%
- (d) 作业：10%
- (e) 期末考试：40%

2. 评分标准

| |
|----------------------|
| 课堂参与度 (30%) |
| 1) 课前预习 |
| 2) 全勤 |
| 3) 积极回答问题，参与课堂讨论 |
| 计算机操作 (30%) |
| 1) 课后实训练习 |
| 2) 案例分析作业 |
| 3) 能独立完成、结论符合逻辑 |
| 小组活动 (40% 展示) |
| 1) 概念清晰 |

| |
|-----------------------|
| 2) 学会数据分析的基本方法及解决简单问题 |
| 3) 数据分析案例 |

期末考试（闭卷考试）成绩占 40%，平时的课程参与度和作业累计占 60%（参考以上评分标准）。