**《面向数据科学的Python编程》教学大纲**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **课程编号** | （略） | **课程名称：** | 面向数据科学的Python编程 | |
| **课程类别** | 专业必修 | **学分/学时** | 3/54 | |
| **任课教师** | 朝乐门 | **答疑时间和地点** | 第1-19周周二 10：00-12：00 信息楼3\*\*\* | |
| **先修课程** | C语言程序设计/Java语言程序设计、数据结构、数据库 | | | |
| **授课对象** | 信08-1，信08-2 | | | |
| **课程目标** | 1.理解Python语言与其他编程语言（如C/Java等）的区别与联系；  2.掌握面向数据科学/数据分析的Python编程的基本思路、方法和技能；  3掌握从事基于Python的数据加工、可视化、统计分析和机器学习项目的能力；  4.掌握继续学习基于Python的Spark和NoSQL编程的基本能力；  5.提升基于Python的数据分析/数据科学项目的动手操作能力。 | | | |
| **课程简介** | 本课程以数据分析/数据科学为目的与导向，讲解Python编程思想，帮助学生掌握基于Python的数据加工、数据可视化、统计分析、机器学习、Spark编程和NoSQL数据库编程能力，提升学生的编程语言基础和实际动手操作能力，为后续相关课程的学习打下较好的基础。 | | | |
| **学习要求** | 1.需要认真听课，并完成上机操作；  2.需要完成任课教师指定的编程任务；  3.需要以小组形式完成一项综合应用项目。 | | | |
| **教材及**  **参考文献** | **1.教材**  朝乐门.Python编程:从数据分析到数据科学[M].北京:电子工业出版社，2019.  **2.参考文献**.  Jake VanderPlas. Python Data Science Handbook [M] O Reily, 2017. | | | |
| **平时考核(占总成绩比例%)** | 60 | **期中考试**  **(占平时成绩比例%)** | | 0 |
| **期末考核(占总成绩比例%)** | 40 | **课堂作业**  **(占平时成绩比例%)** | | 90 |
| **期末考核内容** | 见教学内容 | **课堂表现**  **(占平时成绩比例%)** | | 10 |

|  |  |  |  |
| --- | --- | --- | --- |
| **教学周** | **章节名称** | **讲授内容及掌握程度** | **上机及课后习题** |
| 第1周 | **模块（一） 准备工作**  1.1 Python语言的特征  1.2 Python开发环境的搭建  1.3 Python编程的基本步骤及注意事项 | * 了解为什么要学习Python和学习Python的哪些知识 * 掌握Python开发环境的搭建方法 * 掌握Jupyter Notebook的操作方法 | 习题1-3 |
| 第2周 | **模块（二） Python基础**  2.1数据类型及变量定义方法  2.2 语句书写规范及常用语句  2.3 运算符 | * 了解Python数据类型、语句和运算符与C/Java语言的区别与联系 * 理解Python中的变量及其定义方法 * 掌握Python语句的编写规范 * 掌握赋值、注释、if、for、while、pass语句的编写方法 | 习题4-13  上机练习本章代码 |
| 第3周 | 2.4列表、元组及序列  2.5集合与字典 | * 理解列表和元组的区别与联系 * 理解序列类型的共性特征 * 掌握列表、元组、集合及字典的操作方法 | 习题14-19  上机练习本章代码 |
| 第4周 | 2.6 Python中的函数 | * 理解Python函数的类型与特点 * 掌握自定义函数的定义和调用方法 * 掌握lambda函数的定义和调用方法 | 习题21-25  上机练习本章代码 |
| 第5周 | 2.7 Python 中的包与模块  2.8 Python帮助文档  2.9 搜索路径与当前工作目录 | * 理解Python中的模块与包的区别和联系 * 掌握Python帮助文档的查看方法 * 掌握搜索路径和当前工作目录的操作方法 | 习题26  习题27  习题28  上机练习本章代码 |
| 第6周 | 2.10 Python面向对象编程 | * 了解Python面向对象编程的特点 * 掌握类的定义和调用方法 | 习题31  上机练习本章代码 |
| 第7周 | **模块（三）数据加工**  3.1 NumPy与数组 | * 了解Python数据加工的常用包和模块 * 理解NumPy的多维数组对Python列表和元组的改进 * 掌握基于NumPy的矩阵/数组计算方法 | 习题36  上机练习本章代码 |
| 第8周 | 3.2 Pandas与数据框 | * 了解数据框与数组的区别 * 掌握基于数据框的数据加工方法 | 上机练习本章代码 |
| 第9周 | 3.3 Matplotlib 与数据可视化 | * 了解Python数据可视化中常用的模块与包 * 掌握基于Matplotlib的数据可视化方法 | 习题38 |
| 第10周 | **模块（四）数据分析**  4.1 统计分析（上） | * 了解Python统计分析的常用包/模块 * 理解基于Python和统计学的数据分析基本步骤 | 上机练习本章代码 |
| 第11周 | 4.1 统计分析（下） | * 掌握特定统计模型（如线性回归）的数据分析能力和技巧 | 习题43  综合训练项目（启动） |
| 第12周 | 4.2 机器学习（上） | * 了解Python机器学习的常用包/模块 * 理解机器学习和统计学的区别与联系 * 理解基于Python和机器学习的数据分析基本步骤 | 上机练习本章代码 |
| 第13周 | 4.2 机器学习（下） | * 掌握特定机器学习算法（如KNN算法）的数据分析能力和技巧 | 习题44 |
| 第14周 | **模块（五）大数据处理**  5.1 Spark编程（上）  （1）Spark 技术的简介  （2）pyspark包及其导入方法  （3）SparkSession及其创建 | * 了解Spark的核心技术与主要特点； * 理解用Python调用Spark的基本方法 * 掌握SparkSession及其构建方法 | 学生按照教材P445的提示搭建Spark Python开发环境 |
| 第15周 | 5.1 Spark编程（中）  （1）Spark数据抽象类型  （2）Spark 数据框  （3）Spark SQL编程 | * 掌握Spark数据框操作方法 * 理解 SparkSQL编程方法 | 上机练习本章代码 |
| 第16周 | 5.1 Spark编程（下）  Spark机器学习 | * 了解Spark机器学习的基本本步骤 | 习题45 |
| 第17周 | 5.2 NoSQL数据库（上）  （1）Memcached简介  （2）Memcached的安装 | * 了解NoSQL及Memcached的基本特征 * 掌握Memcached的安装方法 | 上机练习本章代码 |
| 第18周 | 5.2 NoSQL数据库（下）  （3）用Python调用Memcached | * 了解NoSQL及Memcached的基本特征 * 掌握用Python读写Memcached的方法 | 习题47 |
| 第19周 | 课程总结与讨论 | 总结与讨论 | 综合训练项目（提交） |

提示：

1.建议您根据自己课程的培养目标、学生基础和学时安排，对本书进行适当裁剪，部分内容略讲或不讲。

2..课程名应根据自己的教学需要进行适当调整，如面向数据分析的Python编程，Python数据分析等。

3.如有其他问题，请联系本书作者朝乐门老师chaolemen@ruc.edu.cn