



黑龙江科技大学

2020版 **计算机与信息工程学院
本科课程教学大纲**

UNDERGRADUATE COURSE SYLLABUSES FOR
SCHOOL OF COMPUTER AND INFORMATION ENGINEERING



黑龙江科技大学教务处

二〇二〇年

目 录

数据科学与大数据技术专业

《概率统计》教学大纲	1
《离散数学》教学大纲	4
《专业导论》教学大纲	6
《C 程序设计》教学大纲	8
《C++程序设计》教学大纲	11
《优化与运筹学》教学大纲	13
《数据结构 (C++)》教学大纲	15
《数据结构 (C++)》(实验) 教学大纲	17
《数据库原理及应用》教学大纲	19
《数据库原理及应用》(实验) 教学大纲	23
《操作系统原理与 Linux 系统》教学大纲	25
《数学建模与 MATLAB》教学大纲	28
《数学建模与 MATLAB》(实验) 教学大纲	30
《Hadoop 技术基础》教学大纲	32
《Hadoop 技术基础》(实验) 教学大纲	35
《计算机网络》教学大纲	38
《计算机网络》(实验) 教学大纲	41
《Python 程序设计》教学大纲	44
《算法分析与设计》教学大纲	46

《分布式计算系统原理》教学大纲	48
《大数据分析与应用》教学大纲	51
《大数据分析与应用》（实验）教学大纲	54
《NoSQL 数据库原理》教学大纲	56
《NoSQL 数据库原理》（实验）教学大纲	59
《数据可视化技术》教学大纲	61
《机器学习基础》教学大纲	63
《机器学习基础》（实验）教学大纲	66
《Spark 大数据处理技术》教学大纲	69
《Spark 大数据处理技术》（实验）教学大纲	71
《专业外语》教学大纲	74
《R 语言统计分析》教学大纲	76
《复变函数》教学大纲	78
《数据分析》教学大纲	80
《数字图像处理》教学大纲	83
《人工智能》教学大纲	86
《深度学习基础》教学大纲	89
《软件工程技术》教学大纲	92
《数值分析》教学大纲	95
《最优化方法》教学大纲	98
《云计算与大数据》教学大纲	100
《认识实习》教学大纲	103

《专业实训（基础）》教学大纲	105
《Spark 大数据处理技术课程设计》教学大纲	107
《大数据分析与应用课程设计》教学大纲	109
《专业实训（高级）》教学大纲	111
《生产实习》教学大纲	114
《毕业实习》教学大纲	116
《毕业设计（论文）》教学大纲	118

《概率统计》教学大纲

课程代码：175003

课程类别：通用基础

适用专业：数据科学与大数据技术

总学时：56

总学分：3.5

执笔人：张兴华

审定人：史健婷

审批人：张兴华

一、课程性质及教学目标

《概率统计》是数据科学与大数据技术专业大学本科必修的通用基础课程，在人才培养中占有重要的地位。包含了概率统计的基本原理、计算方法、逻辑推导，具备在大数据分析中应用的优势。

课程的教学目标是使学生掌握随机现象的基本概念、基本理论，掌握概率论与数理统计的论证方法，较熟练地获得本课程所要求的基本计算方法和能力，培养学生运用概率统计方法分析和解决实际问题尤其是数据分析问题的能力，为后续课程及其它相关学科的学习建立良好的知识储备。

二、教学内容及要求

（一）随机事件与概率（12学时）

1. 样本空间的概念，理解随机事件的概念，熟练掌握事件之间的关系与运算。
2. 概率的定义（古典概率、几何概率、概率的统计定义和概率的公理化定义），掌握概率的性质并且会应用性质进行概率的计算。
3. 条件概率的概念，掌握概率的加法公式、乘法公式、全概率公式和贝叶斯（Bayes）公式并会用这些公式进行概率计算。
4. 事件独立性的概念，熟练掌握伯努利（Bernoulli）概型。

（二）随机变量及其分布（12学时）

1. 随机变量的概念、离散型随机变量及分布列的概念和性质，会利用分布列计算概率。
2. 连续型随机变量，掌握分布函数与密度函数的关系。
3. 二点分布、二项分布、泊松（Poisson）分布、几何分布以及它们之间的关系。
4. 连续型随机变量及其概率密度的概念，掌握均匀分布、正态分布、指数分布及其应用。
5. 泊松定理的结论和应用条件。
6. 数学期望与方差的概念，掌握它们的性质与计算。
7. 简单随机变量函数的概率分布。
8. 一维随机变量的数字特征。
9. 切比雪夫不等式。

（三）多维随机变量及其分布（12学时）

1. 联合分布函数概念。
2. 联合分布函数与边际函数的关系。
3. 二维随机变量的联合分布列、联合概率密度的概念和性质，并会计算有关事件的概率，了解二维随机变量的边际分布及条件分布。
4. 随机变量的独立性概念。
5. 随机变量函数的分布，会求两个独立随机变量的几类函数的分布。
6. 二维均匀分布、二维正态分布。
7. 多维随机变量的数字特征。

数据科学与大数据技术专业教学大纲

（四）大数定律与中心极限定理（8 学时）

1. 伯努利、切比雪夫大数定律以及他们相互关系。
2. 随机变量序列的两种收敛性。
3. 独立同分布的中心极限定理及其应用。

（五）统计量及其分布（4 学时）

1. 总体、个体、样本、简单随机样本、样本值、样本容量、统计量等数理统计的基本概念。
 2. 样本均值、样本方差和样本矩的计算，分布函数。
 3. 三个重要分布 χ^2 分布、t 分布、F 分布的定义及其简单性质，了解常用概率分布分位数的概念。
- 次序统计量及其分布。

（六）参数估计（8 学时）

1. 参数点估计的概念及两种点估计法的基本思想。
2. 点估计的两种方法：矩估计法（一阶，二阶）与极大似然估计法。
3. 了解估计量的评价标准（无偏性、有效性、一致性）。

本课程重点内容：古典概率、几何概率、条件概率、概率的加法公式、乘法公式、全概率公式和贝叶斯（Bayes）公式，离散型和连续型随机变量及其分布，数学期望与方差，多维随机变量联合分布和边际分布，二维均匀分布、二维正态分布，统计的基本概念和基本计算。

本课程难点内容：随机变量函数的分布，多维随机变量条件分布、随机变量独立性、多维随机变量数字特征，大数定律和中心极限定理及其应用，统计量及其分布，矩估计法与极大似然估计法，了解估计量的无偏性、有效性、一致性。

三、过程考核方式

课程一般采用平时表现与期末考试相结合的方式进行考核。平时表现占 30%，包括课堂表现 15%，作业完成情况 15%；期末考试 70%。也可以只采用笔试考核的方式。

四、教学方法与手段

教学方法以课堂的讲授为主，课前预习、教师课后辅导、答疑为辅助手段。鼓励学生充分利用现代教学技术，到网上收集，查找本学科的相关资料，巩固学习成果，开阔视野。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
随机事件与概率	12				12
随机变量及其分布	12				12
多维随机变量及其分布	12				12
大数定律与中心极限定理	8				8
统计量及其分布	4				4
参数估计	8				8
合计	56				56

六、参考书

- [1] 齐淑华, 刘强, 丁淑妍等. 概率论与数理统计. 北京: 清华大学出版社, 2019

数据科学与大数据技术专业教学大纲

- [2] 宗序平. 概率论与数理统计 (第 4 版). 北京: 机械工业出版社, 2019
- [3] 茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程 (第 3 版). 北京: 高等教育出版社, 2019
- [4] 吴赣昌. 概率论与数理统计. 北京: 中国人民大学出版社, 2017

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《离散数学》教学大纲

课程代码：175018 课程类别：通用基础 适用专业：数据科学与大数据技术

总学时：48 总学分：3

执笔人：付喜辉 审定人：史健婷 审批人：李海华

一、课程性质及教学目标

本课程为数据科学与大数据技术专业本科生必修的一门基础核心课程。

教学目标是要求学生掌握集合论、数理逻辑、图论和代数系统等离散数学的基本内容，为数据结构、数据库、操作系统、编译原理、人工智能、计算机网络等计算机专业各后续课程做好必要的知识准备，使学生得到良好的数学训练，提高学生抽象思维和逻辑推理能力，为从事计算机的应用提供坚实的理论基础。

二、教学内容及要求

（一）命题逻辑（16学时）

1. 了解命题与联结词。
2. 理解命题公式及分类。
3. 掌握等值式。
4. 掌握范式。
5. 了解联结词的完备集。
6. 理解推理理论。
7. 理解一阶逻辑基本概念。
8. 掌握一阶逻辑合式公式及解释，理解一阶逻辑等值式。

（二）集合（6学时）

1. 了解集合的基本概念。
2. 掌握集合的基本运算。
3. 理解集合元素的计数。

（三）二元关系和函数（14学时）

1. 理解有序对与笛卡尔积与二元关系。
2. 掌握关系的运算。
3. 掌握关系的性质。
4. 掌握关系的闭包。
5. 掌握等价关系和偏序关系。
6. 理解函数的定义与性质。
7. 了解函数的复合和反函数。

（四）图论（4学时）

1. 了解无向图及有向图及通路、回路和图的连通性。
2. 理解图的矩阵表示及无向树及生成树。

（五）代数结构（8学时）

1. 了解二元运算及其性质。
2. 理解代数系统。

数据科学与大数据技术专业教学大纲

3. 掌握代数系统的一般性质。

4. 掌握几个典型的代数系统。

本课程重点内容：命题逻辑、二元关系、图论、代数结构。

本课程难点内容：命题逻辑、代数结构。

三、过程考核方式

课程最终成绩按百分制评定。成绩包括平时成绩和期末考试成绩，平时成绩占 40%，期末考试成绩占 60%，平时成绩由课堂表现、作业、测验等组成，可根据情况适当修改考核方式。

四、教学方法与手段

利用多媒体教学，采取学练结合的方式进行教学。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
命题逻辑	16				16
集合	6				6
二元关系和函数	14				14
图论	4				4
代数结构	8				8
合计	48				48

六、参考书

[1] 张清华. 离散数学及其应用. 北京：清华大学出版社, 2016

[2] 刘铎. 离散数学及其应用（第 2 版）. 北京：清华大学出版社, 2018

[3] 刘忠艳. 离散数学（第 1 版）. 北京：清华大学出版社, 2016

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《专业导论》教学大纲

课程代码：175001

课程类别：专业基础

适用专业：数据科学与大数据技术

总学时：8

总学分：0.5

执笔人：史健婷

审定人：丁文妙

审批人：李福军

一、课程性质及教学目标

《专业导论》是数据科学与大数据技术专业的必修课程，主要讲授数据科学的基础知识、数据处理和分析的基本技术，以及大数据技术的典型应用案例和不同场景的作用等。目的是使学生能够了解数据科学的基本概念和 IT 领域的大数据技术，并能够应用计算思维和分析方法解决问题。

通过各教学环节，本课程应达到下列要求，具体包括：

能够建立对大数据专业科学知识体系的轮廓性认识，了解专业发展历程、基本概念、主要影响、应用领域、关键技术、计算模式和产业发展；

能够了解数据科学与大数据技术专业的人才培养体系和人才培养模式以及就业方向；

能够联系实际，了解数据科学与大数据技术的典型应用，对信息技术的未来发展趋势有一定了解。

二、教学内容及要求

（一）专业概述（2学时）

1. 数据科学与大数据技术的概念。
2. 数据科学与大数据技术专业的发展历史。
3. 数据科学与大数据技术的研究方向。

（二）专业发展（2学时）

1. 专业发展概况。
2. 专业发展趋势。
3. 专业目前的分布状况。

（三）新技术的应用（2学时）

1. 大数据技术。
2. 什么云计算技术。
3. 什么是非关系型数据库。
4. 企业和公司的新技术应用概况。

（四）专业人才培养（2学时）

1. 专业人才培养体系。
2. 专业人才培养模式。
3. 数据科学与大数据技术的就业前景。

本课程重点内容：专业概述、专业的发展趋势、新技术的应用、专业人才培养。

本课程难点内容：新技术的应用、专业人才培养。

三、过程考核方式

课程采取学习表现与结课报告相结合的方式进行，学习表现占 40%，包括出勤情况，作业完成情况，课堂表现等，结课报告占 60%。

四、教学方法与手段

数据科学与大数据技术专业教学大纲

采取课堂讲授与课堂讨论的方式进行教学。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
专业概述	2				2
专业发展趋势	2				2
新技术	2				2
专业人才培养	2				2
合计	8				8

六、参考书

- [1] 梅宏. 大数据导论. 北京: 高等教育出版社, 2018
- [2] 刘鹏, 张燕, 付雯. 大数据导论. 北京: 清华大学出版社, 2018
- [3] 袁芳. 计算机导论. 北京: 清华大学出版社, 2016

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《C 程序设计》教学大纲

课程代码：176002 课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：48 总学分：3

执笔人：*聂红梅* 审定人：*史健婷* 审批人：*李瑞军*

一、课程性质及教学目标

本课程为数据科学与大数据技术专业大学本科生必修的一门专业基础课程。

教学目标是通过学习，使同学熟练掌握 C 语言基本语法、结构化程序设计思想及表示，函数的定义与应用，C 语言的指针、结构体、文件操作，锻炼同学使用计算机解决实际问题的思维及数据建模、数据抽象的能力。通过教学的各个环节，采用理论教学与实践操作相结合的形式，逐步培养学生抽象思维和概括问题的能力、逻辑推理能力、量化思维能力、自学能力、较熟练的运算能力和综合运用所学知识分析和解决问题的能力。注重培养学生动手能力、设计能力和应用能力，为进一步学习计算机程序设计知识，深入理解、掌握计算机程序设计的先进技术打下良好的基础。

二、教学内容及要求

（一）程序设计和 C 语言（2 学时）

1. 计算机程序。
2. 计算机语言。
3. C 语言的发展及其特点。
4. C 程序运行的步骤及编译环境的配置。

（二）算法（4 学时）

1. 算法。
2. 算法的特性。
3. 表示算法的方法（流程图、N-S 图）。
4. 结构化程序设计方法。

（三）顺序程序设计（6 学时）

1. 常量、变量和数据类型。
2. 运算符和表达式。
3. C 语句。
4. 数据的输入输出。

（四）选择结构程序设计（6 学时）

1. if 语句。
2. 关系运算符和关系表达式。
3. 逻辑运算符和逻辑表达式。
4. 选择结构的嵌套。
5. switch 语句。

（五）循环结构程序设计（6 学时）

1. while 语句。
2. do ... while 语句。
3. for 循环语句。

4. 循环的嵌套

5. break、continue 语句。

(六) 数组 (6 学时)

1. 一维数组的定义和引用。

2. 二维数组的定义和引用。

3. 字符数组。

(七) 函数 (6 学时)

1. 定义函数。

2. 调用函数。

3. 函数的嵌套调用。

4. 函数的递归调用。

5. 局部变量和全局变量。

(八) 指针 (4 学时)

1. 指针。

2. 定义和引用指针变量。

3. 通过指针引用数组。

4. 通过指针引用字符串。

(九) 用户自己建立数据类型 (4 学时)

1. 定义和使用结构体变量。

2. 使用结构体数组。

3. 结构体指针。

4. 共用体类型。

5. 枚举类型。

6. 用 typedef 声明新类型名。

(十) 文件 (4 学时)

1. 文件概述。

2. 打开与关闭文件。

3. 顺序读写数据文件。

4. 随机读写数据文件。

本课程重点内容：基本数据类型定义及表示，函数的调用及函数参数传递，结构化程序设计的思想，指针、结构体的定义及引用等。

本课程难点内容：数据的表示，函数参数的传递，结构化程序设计的思想，指针引用等。

三、过程考核方式

课程采取平时表现与期末考试相结合的方式进行考核。平时表现占 40%，包括：课堂表现 10%、作业 15%、测试 15%；期末考试 60%。

四、教学方法

数据科学与大数据技术专业教学大纲

教学方法：“教中学，学中做”教学模式，即边讲边练，理论知识教师以 PPT 和板书形式开展，实践部分通过程序实例进行训练。师生交流、答疑、作业等问题，利用 QQ 群和微信群进行有效的沟通。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
程序设计和 C 语言	2				2
算法	4				4
顺序程序设计	6				6
选择结构程序设计	6				6
循环结构程序设计	6				6
数组	6				6
函数	6				6
指针	4				4
用户自己建立数据类型	4				4
文件	4				4
合计	48				48

六、参考书

- [1] 谭浩强. C 程序设计（第五版）. 北京：清华大学出版社, 2017
- [2] 谭浩强. C 程序设计（第五版）学习辅导. 北京：清华大学出版社, 2017
- [3] 何钦铭, 颜晖. C 语言程序设计（第 4 版）. 北京：高等教育出版社, 2020
- [4] 颜晖, 张泳. C 语言程序设计实验与习题指导（第 4 版）. 北京：高等教育出版社, 2020

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《C++程序设计》教学大纲

课程代码：176003

课程类别：专业基础

适用专业：数据科学与大数据技术

总学时：32

总学分：2

执笔人：聂红梅

审定人：史健婷

审批人：李瑞华

一、课程性质及教学目标

本课程为数据科学与大数据技术专业大学本科生必修的一门专业基础课程。教学目标是学生通过本课程的学习，要了解面向对象程序设计的基本原理，包括抽象、封装、继承等关键要素以及其关键语法特性——类；熟悉 C++ 的新语法特性，了解 C 与 C++ 的区别和兼容性；掌握用 C++ 编写面向对象程序的一般方法。

二、教学内容及要求

（一）C++ 的初步知识（4 学时）

1. C++ 的发展历史；
2. C++ 程序的构成与书写形式；
3. 上机的方法和步骤。

（二）类与对象的特性（4 学时）

1. 面向对象程序设计的基本特点；
2. 类和对象；
3. 类的封装性和信息隐蔽。

（三）怎样使用类和对象（6 学时）

1. 构造函数和析构函数；
2. 对象指针的用法；
3. 对象的动态建立和释放；
4. 类的静态成员；
5. 类的友元。

（四）对运算符进行重载（2 学时）

1. 为什么要对运算符重载；
2. 对运算符重载的方法；
3. 重载运算符的规则；
4. 运算符重载函数作为类成员函数和有元函数。

（五）类的继承（6 学时）

1. 类的继承与派生；
2. 派生类的访问属性；
3. 派生类的构造函数和析构函数；
4. 基类与派生类的转换。

（六）多态性（4 学时）

1. 多态性的概念；
2. 虚函数；
3. 纯虚函数和抽象类。

数据科学与大数据技术专业教学大纲

(七) 输入输出流 (6 学时)

1. C++的输入和输出;
2. 标准输出流;
3. 标准输入流;
4. 对数据文件的操作与文件流。

本课程重点内容: C++的初步知识、类和对象、对运算符进行重载、类的继承、C++的多态性、输入输出流等内容。

本课程难点内容: 类和对象、类的继承、多态性。

三、过程考核方式

课程采取平时表现与期末考试相结合的方式进行考核。平时表现占 40%，包括：日常评价（提问、出勤与课堂纪律）10%、作业 15%、测试 15%；期末考试 60%。

四、教学方法

教学方法：“教中学，学中做”教学模式，即边讲边练，理论知识教师以 PPT 和板书形式开展，实践部分通过程序实例进行训练。师生交流、答疑、作业等问题，利用 QQ 群和微信群进行有效的沟通。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践（上机）、实训	合计
C++的初步知识	4				4
类与对象的特性	4				4
怎样使用类和对象	6				6
对运算符进行重载	2				2
类的继承	6				6
多态性	4				4
输入输出流	6				6
合计	32				32

六、参考书

- [1] 谭浩强. C++面向对象程序设计（第 3 版）. 北京：清华大学出版社, 2020
- [2] 谭浩强. C++程序设计（第 3 版）. 北京：清华大学出版社, 2015
- [3] 张娜. C++高级程序设计教程. 北京：清华大学出版社, 2017

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《优化与运筹学》教学大纲

课程代码：175020 课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：32 总学分：2

执笔人：付喜辉 审定人：史健婷 审批人：李海华

一、课程性质及教学目标

本课程为大数据专业大学本科生必修的一门专业基础课程。

本课程的教学目标是使学生掌握优化与运筹学中线性规划、非线性规划、动态规划的基本概念和基本内容，培养学生大数据领域技术问题性质的识别与判断的能力。通过计算机上机，使学生能把实际问题构成数学模型、选择适当方法、运用计算机求出最优解或满意解，以提高学生分析和解决实际问题的能力，也为进一步学习后继课程打下基础。

二、教学内容及要求

（一）线性规划（14 学时）

1. 掌握线性规划的单纯型法。
2. 掌握线性规划的对偶单纯型法及灵敏度分析。
3. 了解运输问题的数学模型及表上作业法。
4. 理解目标规划的数学模型及图解法。
5. 理解指派问题。
6. 掌握对策论的基本概念、基本理论及矩阵对策论的解法。
7. 掌握决策分析的基本问题、及决策方法。

（二）非线性规划（10 学时）

1. 掌握无约束非线性规划的基本概念与性质。
2. 掌握一维搜索方法。
3. 掌握共轭梯度法。
4. 掌握约束非线性规划问题的最优性条件。
5. 了解罚函数法。

动态规划与图论（8 学时）

1. 了解动态规划的基本概念及动态规划的最优性原理。
2. 理解动态规划的求解方法。
3. 掌握图的基本概念及性质、树及最短路问题。
4. 理解网络最大流问题及最小费用最大流问题。

本课程重点内容：线性规划的单纯型法、对偶单纯型法及灵敏度分析、无约束非线性规划问题的性质，约束非线性规划的最优性条件、图的基本性质及最短路问题。

本课程难点内容：线性规划的单纯型法、对偶单纯型法及灵敏度分析、约束非线性规划的最优性条件。

三、过程考核方式

课程最终成绩按百分制评定。成绩包括平时成绩和期末考试成绩，平时成绩占 40%，期末考试成绩占 60%，平时成绩由课堂表现、作业、测验等组成，可根据情况适当修改考核方式。

四、教学方法

数据科学与大数据技术专业教学大纲

利用多媒体教学，采取学练结合的方式进行教学。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
线性规划	14				14
非线性规划	10				10
动态规划	8				8
合计	32				32

六、参考书

- [1] 运筹学教材编写组. 运筹学（第四版）. 北京：清华大学出版社, 2013
- [2] 胡运权. 运筹学教材（第五版）. 北京：清华大学出版社, 2018

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《数据结构 (C++)》教学大纲

课程代码: 176004 课程类别: 专业基础 适用专业: 数据科学与大数据技术
总学时: 64/8 总学分: 4
执笔人: 丁文抄 审定人: 史健婷 审批人: 姜海军

一、课程性质及教学目标

本课程是数据科学与大数据技术专业的一门专业基础课程。

教学目标是使学生掌握各种主要数据结构的特点、计算机内的表示方法, 以及处理数据的算法实现, 能够分析算法所花费的时间和空间代价; 理解各种数据对象的特点, 数据的组织方法和实现方法, 提高学生程序设计能力。

二、教学内容及要求

(一) 数据结构的基本概念 (4 学时)

1. 数据结构的基本概念。
2. 术语以及算法的时间和空间量度。

(二) 线性结构 (20 学时)

1. 线性表的两种存储结构。
2. 两种存储结构的描述方法。
3. 栈和队列两种存储结构的特点, 表示和基本操作。
4. 建立在顺序, 链式两种存储结构上的基本应用。
5. 串的定义和一些基本概念。
6. 串的实现和表示

(三) 非线性结构 (20 学时)

1. 树的一些术语, 树与二叉树之间的转换, 线索化方法。
2. 最优树的有关知识, 二叉树的逻辑结构、性质和存储结构, 二叉树的遍历算法。
3. 图的存储结构, 图的两种遍历算法。
4. 最小生成树的生成算法, 有向无环图及其应用, 最短路径问题的算法。

(四) 查找和排序 (12 学时)

1. 静态查找表的查找算法。
2. 动态查找表的构造和查找算法。
3. 哈希表的构造和查找算法。
4. 排序的基本概念。
5. 插入排序, 希尔排序, 选择排序, 冒泡排序, 快速排序和归并排序算法。

本课程重点内容: 数据结构的基本概念、术语以及算法的时间和空间量度; 单链表的基本知识以及建立在线性表顺序和链式两种存储结构上的算法实现; 栈和队列的顺序, 链式两种存储结构的表示和实现, 栈和队列的特点; 串的定义和一些基本概念; 数组在以行为主序的存储结构中的地址计算方法; 二叉树的结构特性及性质, 二叉树的各种遍历算法的递归形式; 图的两种遍历策略, 最小生成树的生成方法和最短路径问题的解法; 折半查找的实现, 哈希表构造方法, 哈希表的查找以及衡量查找效率的平均查找长度的计算。

数据科学与大数据技术专业教学大纲

本课程难点内容：算法的时间和空间量度；线性表顺序和链式两种存储结构上基本运算的算法实现和算法度量；栈和队列的应用；串的模式匹配算法；二叉树的各种遍历算法的递归算法的理解以及 Huffman 编码；最小生成树的生成方法和最短路径问题的算法；动态查找各种方法的基本思想和实现；各种排序算法的基本思想。

三、过程考核方式

本课程采取平时表现与期末考核相结合的方式进行，平时表现占 30%，包括课堂表现，作业和实验成绩，期末理论知识闭卷考试占 70%。

四、教学方法

利用多媒体教学，采取学练结合的方式进行教学。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
绪论	4				4
线性表	12		2		14
栈和队列	4				4
串、数组和 广义表	4				4
树和二叉树	10		2		12
图	10		2		12
查找表	6				6
排序	6		2		8
合计	56		8		64

六、参考书

- [1] 汪沁等. 数据结构与算法（第 2 版），北京：清华大学出版社, 2018
- [2] 彭军等. 数据结构与算法，北京：人民邮电出版社, 2013
- [3] 严蔚敏等. 数据结构（C 语言版），北京：清华大学出版社, 2012

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《数据结构 (C++)》(实验) 教学大纲

课程代码: 176004 课程类别: 专业基础 适用专业: 数据科学与大数据技术
课程属性: 课内实验 课程总学时: 64 总学分: 4 实验学时: 8 学分: 0.5
执笔人: 丁文妙 审定人: 史健婷 审批人: 姜泊宇

一、实验性质及教学目标

《数据结构 (C++)》(实验) 属于课内实验, 是《数据结构 (C++)》理论课程的有效辅助, 通过理论与实践密切结合, 加深对理论课程中基本数据结构和算法的理解, 使学生能够用熟悉的计算机语言验证所学数据结构算法, 最终运用所学的逻辑结构和物理存储方案解决实际问题的能力。

二、实验项目

序号	实验项目名称	学时	实验类型	实验要求
1	求任意两个一元多项式的和	2	验证性	必做
2	二叉树的建立与遍历	2	验证性	必做
3	图的最小生成树	2	验证性	必做
4	二叉排序树	2	验证性	必做

三、实验概述

实验 1: 求任意两个一元多项式的和

目的: 掌握线性表的链式存储结构, 熟悉线性表的一些基本运算和函数定义, 并用于一元多项式的和运算。

原理概述: 一元多项式可用单链表存储, 节点结构如下:

系数域	指数域	指针域
-----	-----	-----

一个多项式可用一个按指数有序的单链表来存储, 两个多项式的和为指数相同的系数相加, 一个多项式中存在的指数而另一个里没有的结点直接插入即可。

方法与手段: 学生将自定义结点类型, 编写多项式链表的建立, 多项式输入, 显示及求和函数, 而后调试运行。

应得到的实验结果和数据: 输出任意的 2 个多项式及其和。

实验 2: 二叉树的建立与遍历

目的: 掌握建立二叉树二叉链表算法的基本思想和实现方法; 理解遍历的基本含义, 能够在二叉树二叉链表存储结构之上, 输出中序遍历序列。

原理概述: 利用队列实现创建二叉树二叉链表存储结构, 同时在二叉树二叉链表存储结构之上, 利用栈结构来实现中序遍历非递归算法, 输出中序遍历序列。

方法与手段: 学生将教材中的算法调试成可运行的程序。

应得到的实验结果和数据: 完成二叉树二叉链表存储结构的构造, 得出正确的中序遍历序列。

实验 3: 图的求最小生成树

目的: 掌握图的定义和术语, 掌握图的存储结构, 利用 Prim 算法和 kruskal 算法求图的最小生成树。

原理概述: 图的最小生成树是由图的所有顶点构成的, 而且所有边的权值之和最小的生成树。

方法与手段: 学生将教材中的算法调试成可运行的程序。

数据科学与大数据技术专业教学大纲

应得到的实验结果和数据：根据输入的顶点、边和权值，输出最小生成树。

实验 4：二叉排序树

目的：掌握二叉排序树的构造，查找，插入，删除及遍历算法，能够利用二叉排序树中序遍历进行排序。

原理概述：二叉排序树左子树的所有节点的关键字都小于根节点的关键字，右子树的所有节点的关键字都大于根节点的关键字，这样只要与根节点的关键字比较，如果小于就到左子树继续查找，大于就到右子树查找，而左右子树也是二叉排序树，这样形成递归，二叉树的中序遍历即可完成对关键字的排序。

方法与手段：学生将教材中的算法调试成可运行的程序。

应得到的实验结果和数据：输入要排序的关键字，输出最终的排序结果。

四、主要仪器设备

PC 机，C++语言软件开发工具。

五、教学形式

利用多媒体手段总体讲解有关实验的内容要求和注意事项，并且演示关于实验的相应程序，帮助学生从总体上把握每次实验的要求和细节，此后学生根据老师的讲解编制和调试程序，在调试的过程中教师给予积极的帮助和引导，当学生调试成功时，采用一定量的实验数据来验证程序的正确性和合理性，实验结束之后，学生将实验的总体过程详细地记录到实验报告中。

六、考核方式与成绩评定方法

实验成绩通过实验报告成绩确定，总分 100 分，每个实验项目报告 25 分。

七、教材及主要参考资料

- [1] 孙丽云等. 数据结构实验指导（C 语言版、第 1 版）. 北京：华中科技大学出版社, 2017
- [2] 邹永林等. 数据结构与算法习题解析与实验指导（第 1 版）. 北京：清华大学出版社, 2015
- [3] 毛养红等. 数据结构实验指导教程（第 1 版）. 北京：清华大学出版社, 2014

八、说明

在执行大纲过程中应结合专业的特点和学生的实际情况，可适当调整授课内容与进度。

《数据库原理及应用》教学大纲

课程代码：174006 课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：56/8 总学分：3.5

执笔人：陈学刚 审定人：史健婷 审批人：李福军

一、课程性质及教学目标

《数据库原理及应用》是数据科学与大数据技术专业本科生的专业基础课。数据库技术是现代软件技术的重要支撑，是诸多研究方向如信息系统、决策支持系统等的基础、也是支持人工智能、办公自动化软件，计算机辅助软件工程等的有力工具。《数据库原理及应用》课程在专业培养课程体系处于核心地位。

通过本课程的学习，学生应能掌握数据库系统的基本概念和原理；深入理解关系数据模型、关系数据理论和关系数据库系统，掌握关系数据语言；掌握数据库安全保护知识及手段；熟悉典型数据库系统开发技术；掌握数据库设计方法，具有一定的数据库设计能力；初步具备使用数据库技术和方法解决实际应用问题的能力，为今后从事信息系统的管理、开发等工作打下坚实的基础。

重点掌握以下几个方面的基础理论和应用技能：

- (1) 关系代数，关系数据库的完整性
- (2) 关系规范化理论
- (3) 数据库设计
- (4) 数据库与基本表的创建和管理
- (5) 视图与索引
- (6) 数据查询
- (7) 存储过程与触发器
- (8) 事务与锁
- (9) 权限系统
- (10) 数据备份和恢复

二、教学内容及要求

(一) 数据库系统概述 (2 学时)

1. 了解数据库、数据库管理系统。
2. 了解数据库系统基本概念。
3. 了解数据管理技术的发展和数据模型。
4. 理解数据库的体系结构。

(二) 关系代数 (2 学时)

1. 掌握关系的概念。
2. 掌握基本的集合运算。
3. 掌握专门的关系运算。

(三) 关系数据库的完整性 (2 学时)

1. 掌握实体完整性，候选码与主码的概念。
2. 掌握参照完整性。
3. 掌握用户定义的完整性。

（四）关系规范化理论（2 学时）

1. 理解函数依赖概念。
2. 掌握第一范式、第二范式和第三范式。

（五）数据库设计（4 学时）

1. 掌握数据库需求分析。
2. 掌握概念结构设计。
3. 掌握逻辑结构设计。
4. 掌握物理结构设计。

（六）MySQL 数据库（2 学时）

1. 掌握数据库安装。
2. 了解数据库对象。
3. 掌握数据库管理工具基本使用方法。

（七）数据库与基本表的创建和管理（8 学时）

1. 掌握数据库及表的创建、修改和删除。
2. 掌握数据插入、修改和删除。

（八）数据更新与查询（8 学时）

1. 掌握数据查询语言（DDL），包括单表条件查询，单表有条件查询、聚集函数的使用，分组与排序，多表连接查询，嵌套查询，集合查询。
2. 掌握数据更新语言（DML），包括插入，修改，删除等。

（九）视图与索引（4 学时）

1. 掌握视图定义、使用。
2. 掌握索引定义、使用。

（十）存储过程与触发器（6 学时）

1. 了解存储过程类型。
2. 了解触发器类型。
3. 掌握存储过程与触发器的创建、修改和删除。
4. 掌握执行存储过程。

（十一）事务与锁（4 学时）

1. 理解事务的定义。
2. 掌握事务的 ACID 特性。
3. 掌握事务隔离级别。
4. 掌握 MySQL 的锁定机制，分析 InnoDB 行锁争用情况及死锁的处理。

（十二）权限系统（2 学时）

1. 掌握权限表。
2. 了解 MySQL 权限系统的工作原理。
3. 掌握账户管理。
4. 掌握权限管理。

（十三）数据备份和恢复（2 学时）

数据科学与大数据技术专业教学大纲

1. 掌握数据备份的方法。
2. 掌握数据恢复的方法。
3. 了解数据库迁移。
4. 掌握表的导入与导出。

本课程重点内容：实体完整性、参照完整性和用户定义的完整性、数据库基本表的创建、数据的更新与查询、存储过程与触发器。

本课程难点内容：复杂的多表查询，视图的更新，完整性约束的定义，存储过程和出发其的定义和使用。

三、过程考核方式

课程采取平时学习表现与期末考核相结合的方式进行，学习表现占 30%，包括出勤情况（10%）、作业完成情况（10%）及课堂表现（10%），期末考试占 70%。

四、教学方法与手段

课程采用多媒体方式进行教学，其中实践教学采用案例式、项目式等方式进行，在教学中使用 MySQL、Navicat 等软件进行相关理论的分析与实践。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
数据库系统概述	2				2
关系代数	2				2
关系数据库的完整性	2				2
关系规范化理论	2				2
数据库设计	4				4
MySQL数据库	2				2
数据库与表的管理	8		2		10
数据更新与查询	8		4		12
视图与索引	4				4
存储过程与触发器	6		2		8
事务与锁	4				4
权限系统	2				2
数据备份和恢复	2				2
合计	48		8		56

六、参考书

- [1] 王珊, 萨师焯. 数据库系统概论第 5 版. 北京: 高等教育出版社, 2014
- [2] 孟繁荣, 闫秋燕等. 数据库原理与应用 (MySQL 版). 北京: 清华大学出版社, 2019
- [3] 郑阿奇. MySQL 教程. 北京: 清华大学出版社, 2015

七、说明

数据科学与大数据技术专业教学大纲

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《数据库原理及应用》（实验）教学大纲

课程代码：174006 课程类别：专业基础 适用专业：数据科学与大数据技术 课程属性：课内实验
课程总学时：56 总学分：3.5 实验学时：8 学分：0.5
执笔人：陈学刚 审定人：史健婷 审批人：李海宇

一、实验性质及教学目标

实验性质：课内实验，非独立实验课

教学目标：通过实验使学生掌握数据库对象的定义和数据表的操作和管理，并深入理解数据库设计各个环节中的重点内容及所需掌握的专业技术。具体划分如下：

1. 掌握数据库和表结构的定义和修改，熟悉表设计器和菜单的使用；
2. 掌握数据表的增、删、改、查的操作；
3. 掌握视图的定义和查询，了解视图的更新限制及视图的作用；
4. 掌握存储过程和触发器的定义和使用。针对存储过程，需掌握参数的种类和应用场景，以灵活运用存储过程实现表应用过程中复杂功能的程序设计和实现；针对触发器，需掌握触发时机、触发体等语句的编写，了解在 MySQL 下触发器的语法限制。

二、实验项目

序号	实验项目名称	学时	实验类型	实验要求
1	数据库和数据表的定义和使用	2	验证性	必做
2	数据表的查询和更新	2	验证性	必做
3	视图的定义和使用	2	验证性	必做
4	存储过程和触发器的定义和使用	2	验证性	必做

三、实验概述

实验一：数据库和数据表的定义和使用

实验内容为表结构的建立和修改、约束的定义、索引的定义。使学生在给定数据和实验要求的前提下独立完成表结构的建立。

学时：2 学时

实验二：数据表的查询和更新

实验内容为单表查询、多表查询、带有集函数的查询。数据应用过程中，查询是主要功能，使学生熟练掌握基本查询语句结构，掌握复杂查询功能语句的设计。

学时：2 学时

实验三：视图的定义和使用

实验内容为视图的定义、视图的查询操作、视图的更新操作。实验过程中学生应掌握基于多表建立视图及更新视图的操作，同时使学生了解视图中常见的更新限制。

学时：2 学时

实验四：存储过程和触发器的定义和使用

实验内容为存储过程的定义和调用，触发器的定义和使用。实验应使学生掌握存储过程的定义、带参存储过程中参数的类型及使用场景和方法、触发器的定义，合理设定触发时机及触发体语句。

学时：2 学时

四、主要仪器设备

硬件为电脑，软件为 MySQL 时下稳定版本+Navicat。

五、教学形式

在教师指导下，给定实验要求和题目，由学生独立完成。

六、考核方式与成绩评定方法

非独立实验，作为数据库原理及应用理论课程过程性考核的一部分，其占期末考试成绩的 10%，即 10 分。

七、教材及主要参考资料

- [1] 王珊, 萨师焯. 数据库系统概论第 5 版. 北京: 高等教育出版社, 2014
- [2] 孟繁荣, 闫秋燕. 数据库原理与应用 (MySQL 版). 北京: 清华大学出版社, 2019
- [3] 郑阿奇. MySQL 教程. 北京: 清华大学出版社, 2015

八、说明

实验内容仅做参考，授课教师可根据学生对不同实验内容的掌握情况调整授课重点及内容。

《操作系统原理与 Linux 系统》教学大纲

课程代码：176005 课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：48 总学分：3

执笔人：常亮 审定人：史健婷 审批人：李海军

一、课程性质及教学目标

本课程是数据科学与大数据技术专业本科生必修的一门专业基础课程。

教学目标是重点讲述操作系统的运行过程和设计原理，使学生掌握操作系统的基本概念、设计方法、运行机制，从而建立起对操作系统的整体认识。培养学生对大型软件的分析与设计能力，为后续课程的学习提供基础理论知识，为学生将来从事系统开发和软件开发，奠定系统调度算法的基本设计能力。

二、教学内容及要求

（一）操作系统概述（4学时）

1. 操作系统的设计目标。
2. 操作系统的地位和作用。
3. 操作系统的主要功能。
4. 多道批处理系统、分时系统、实时系统。

（二）进程描述及控制（12学时）

1. 程序的顺序执行。
2. 并发执行及线程的概念。
3. 进程的概念、组成。
4. 进程的基本状态及状态间的转换。
5. 进程同步的概念及信号量机制。

（三）处理机管理（12学时）

1. 处理机调度，作业调度及常用算法。
2. 进程调度及常用算法。
3. 死锁的概念及产生死锁的必要条件。
4. 预防死锁、避免死锁和检测和解除死锁的方法。
5. 银行家算法。

（四）存储器管理（12学时）

1. 存储器的层次结构、程序装入和链接的方式。
2. 内存连续分配算法。
3. 对换及离散存储的概念，基本分页存储管理方式和基本分段存储管理方式。
4. 虚拟存储的概念。
5. 请求分页式存储管理方式。
6. 常用页面置换算法，抖动与工作集的概念。

（五）设备管理（4学时）

1. I/O系统的功能、模型和接口。
2. 串行和并行设备的特点、设备的分类。

数据科学与大数据技术专业教学大纲

3. 缓冲策略、直接存储器访问等，磁盘存储器调度算法。

(六) 文件管理 (4学时)

1. 文件、文件系统的相关概念。
2. 文件的逻辑结构、文件目录、文件共享。
3. 文件保护的概念及组织方式。

本课程重点内容：

1. 掌握操作系统的基本概念、特征、功能，了解操作系统的发展历史，理解操作系统的运行机制，培养应用、调试及优化操作系统的能力；

2. 掌握进程的概念，深刻理解处理机调度的准则、算法以及死锁问题，重点掌握进程同步机制和避免死锁的方法，并能够选择适当的软件开发工具实现处理机调度中的关键技术；

3. 了解存储器的分类、特点及主要的内存分配方法，掌握页式存储管理、段式存储管理及段页式存储管理方式，理解虚拟存储器的概念，重点掌握页面置换算法，并能够选择适当的软件开发工具实现存储器管理中的关键技术；

4. 了解设备管理及文件管理的基本概念，掌握输入输出系统的构成和文件目录的概念，重点掌握缓冲技术和磁盘调度算法。

本课程难点内容：系统的了解操作系统各功能模块的运行机制，理解操作系统应用与开发技术之间的密切关系和相关知识，培养学生对大型软件的分析 and 设计能力。

三、过程考核方式

本课程采取平时作业+期末考试+课堂表现，平时作业课堂表现40%，包括出勤情况，作业完成情况，其末理论知识闭卷考试占60%。

四、教学方法

利用多媒体教学，采取案例教学，PPT结合投屏演示，线上线下混合式，利用腾讯课堂和腾讯会议进行线上的演示和答疑。

五、各教学环节学时分配

授课内容 (知识点)	第一课堂	第二课堂	实验	实践实训	合计
操作系统概述	4				4
进程描述及控制	12				12
处理机管理	12				12
存储器管理	12				12
设备管理	4				4
文件管理	4				4
合计	48				48

六、参考书

- [1] 郑鹏等编. 计算机操作系统 (第三版). 武汉: 武汉大学出版社, 2020
- [2] 汤小丹. 计算机操作系统 (第三版). 西安: 西安电子科技大学出版社, 2007
- [3] 汤小丹, 梁红兵, 哲凤屏, 汤子瀛编. 现代操作系统. 北京: 电子工业出版社, 2009

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《数学建模与 MATLAB》教学大纲

课程代码：176006 课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：56/8 总学分：3.5

执笔人：丁文妙 审定人：史健婷 审批人：李海军

一、课程性质及教学目标

《数学建模与 MATLAB》是数据科学与大数据技术专业必修课程，注重锻炼学生的数学建模、数据分析能力等所需的基础知识和基本能力。

教学目标：MATLAB 是一种以数值计算和数据图示为主的计算机软件，并包含适应多个学科的专业软件包，以及完善程序开发功能。本课程要求学生掌握 MATLAB 的数据类型、矩阵输入和操作方法、语法结构、函数的使用以及二维、三维绘图功能等基础知识和数学模型，数据分析工具软件的使用，并能够熟练地将 MATLAB 应用于实际问题的数学建模中，解决相关问题的数学计算问题，培养解决工程问题的分析、建模和编程求解能力。

二、教学内容及要求

（一）MATLAB 系统简介（2 学时）

1. MATLAB 软件的发展历史。
2. MATLAB 的基本情况，以及学习的意义。
3. 启动和退出 MATLAB 的方法。
4. MATLAB 的各种功能介绍。
5. 命令窗口的使用，MATLAB 帮助。

（二）基本数据类型及基本运算（12 学时）

1. 数值类型，逻辑类型，字符和字符串，结构体类型，单元数据类型。
2. 矩阵创建、保存和提取方法。
3. 矩阵元素标识，矩阵函数，矩阵的运算。
4. 数组的运算和数组函数。

（三）MATLAB 编程（6 学时）

1. 变量类型、基本表达式、数据类型、算符与操作符
2. 关系运算与逻辑运算，三种程序结构。
3. 两种 M 文件和 M 文件的调试方法。

（四）符号及其运算（4 学时）

1. 符号变量，符号表达式。
2. 符号方程，符号函数的基本概念及运算。

（五）数据文件操作（4 学时）

1. 打开和关闭数据文件的方法。
2. 读、写格式化文本文件的方法。

（六）图形功能和 GUI 设计（8 学时）

1. 二维平面图形与坐标系。
2. 三维绘图，坐标轴的控制和图形标注。
3. 特殊图形绘图。

数据科学与大数据技术专业教学大纲

(七) MATLAB 在数学建模中的应用 (12 学时)

1. 数值计算工具。
2. 优化问题工具和数据分析工具的使用。

本课程重点内容：基本数据类型，矩阵的运算，三种程序结构，二维绘图。

本课程难点内容：数组的运算、函数 M 文件的创建，三维绘图，数学工具的使用。

三、过程考核方式

课程采取学习表现与期末考核相结合的方式进行，学习表现占 30%，包括课堂表现，作业完成情况和实验成绩，期末理论知识闭卷考试占 70%。

四、教学方法

利用多媒体教学，采取学练结合的方式进行教学。引导学生理解基本概念、理论和方法，掌握基本操作、基本技能和学习方法。鼓励学生自学和实践、主动获取知识，激发学生的学习主动性和兴趣。培养学生分析问题、解决问题的能力、自学能力、实践和创新能力。

五、各教学环节学时分配

授课内容 (知识点)	第一课堂	第二课堂	实验	合计
MATLAB 简介	2			2
数据表示及基本运算	12			12
MATLAB 语言编程	6		2	8
符号及其运算	4		2	6
数据文件操作	4			4
图形功能和 GUI 设计	8		4	12
MATLAB 在数学建模中的应用	12			12
合计	48		8	56

六、参考书

- [1] 苏庆堂等编. MATLAB 原理及应用案例教程. 北京：清华大学出版社, 2016
- [2] 薛定宇译. MATLAB 编程之父：编程实践. 北京：北京航空航天大学出版城, 2014
- [3] 赵骥等编. MATLAB 基础与实例教程. 北京：清华大学出版社, 2018

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《数学建模与 MATLAB》（实验）教学大纲

课程代码：176006 课程类别：专业基础 适用专业：数据科学与大数据技术
 课程属性：课内实验 课程总学时：56 总学分：3.5 实验学时：8 学分：0.5
 执笔人：丁文妙 审定人：史健婷 审批人：李海军

一、实验性质及教学目标

《数学建模与 MATLAB》（实验）属于课内实验，是《数学建模与 MATLAB》理论课程的有效辅助，通过理论与实践密切结合，加深对理论课程中基本知识的理解，使学生能够用熟悉的 MATLAB 语言进行数学计算和结果的绘图，进而达到对实际问题进行分析建模和求解的能力。

二、实验项目

序号	实验项目名称	学时	实验类型	实验要求
1	英文字符串中单词处理	2	验证性	必做
2	符号运算	2	验证性	必做
3	图形绘制	2	验证性	必做
4	gui 和 app	2	验证性	必做

三、实验概述

实验 1：英文字符串中单词处理

目的：理解字符串操作的相关概念，了解 MATLAB 字符串操作的相关函数；掌握 MATLAB 脚本和函数编程

原理概述：英文字符串中单词之间使用空格或标点符号分隔，利用 MATLAB 字符串操作的相关函数可以把单词分割出来。

方法与手段：1. 编写一个脚本，判断输入字符串中每个单词的首字母是否为大写，若不是，则将其改为大写，其他字母为小写。2. 编写一个函数，计算输入字符串中单词的个数，并提取每个单词，计算每个单词长度。

应得到的实验结果和数据：1. 输出首字母为大写的单词，2. 输出单词个数，每个单词及长度。

实验 2：符号运算

目的：掌握符号变量，符号表达式，符号方程，符号函数的创建，掌握符号代数方程，符号微分方程的求解。

原理概述：利用 MATLAB 符号工具可以创建符号方程，符号方程组，符号常微分方程，并提供对应的求解函数。

方法与手段：1. 编写一个脚本，求解方程组。2. 编写一个脚本，求常微分方程初值问题。

应得到的实验结果和数据：输出方程组的解和常微分方程初值问题的解。

实验 3：图形绘制

目的：掌握 MATLAB 基本绘图命令，掌握二维、三维 MATLAB 绘图。

原理概述：利用 MATLAB 绘图函数可以画出函数的图形。

方法与手段：1. 编写一个脚本，使用 subplot 分别画出 $\sin(x)$ ， $\cos(x)$ ， $\exp(x)$ ， x^2 并添加标题，标注 x 轴，y 轴，加网格。2. 编写一个脚本，画出 $\cos(x)$ 和 $\sin(x)$ 的图像，添加图例，标注 x 轴，区分线型和颜色。3. 编写一个脚本，画出函数图像，x, y 的取值范围为 -2 到 2，步长取

0.1。

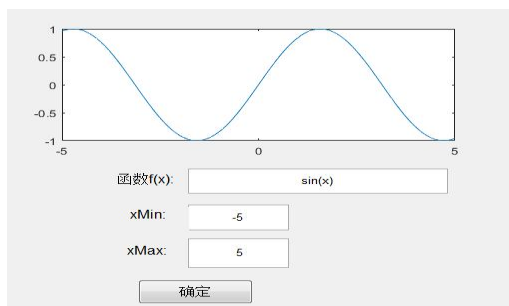
应得到的实验结果和数据：输出函数图形。

实验 4: gui 和 app

目的：掌握图形用户界面 GUI 设计的步骤，掌握 GUI 设计常用控件。

原理概述：利用 MATLAB 图形用户应用程序设计功能可以设计开发简单的 GUI 应用程序。

方法与手段：设计一个 MATLAB 应用程序，能够输入任意一个一元 MATLAB 函数表达式和自变量的取值范围，画出该函数的图形。界面参考下图：



应得到的实验结果和数据：应用程序运行正常，结果准确。

四、主要仪器设备

PC 机，MATLAB 软件开发工具。

五、教学形式

利用多媒体手段总体讲解有关实验的内容要求和注意事项，并且演示关于实验的相应程序，帮助学生从总体上把握每次实验的要求和细节，此后学生根据老师的讲解编制和调试程序，在调试的过程中教师给予积极的帮助和引导，当学生调试成功时，采用一定量的实验数据来验证程序的正确性和合理性，实验结束之后，学生将实验的总体过程详细地记录到实验报告中。

六、考核方式与成绩评定方法

实验成绩由实验报告成绩确定，实验报告成绩总分 100 分，每个实验项目报告 25 分。

七、教材及主要参考资料

- [1] 苏庆堂等编. MATLAB 原理及应用案例教程, 北京: 清华大学出版社, 2016
- [2] 薛定宇译. MATLAB 编程之父: 编程实践, 北京: 北京航空航天大学出版城, 2014
- [3] 赵骥等编. MATLAB 基础与实例教程, 北京: 清华大学出版社, 2018

八、说明

在执行大纲过程中应结合专业的特点和学生的实际情况，可适当调整授课内容与进度。

《Hadoop 技术基础》教学大纲

课程代码：176007

课程类别：专业基础

适用专业：数据科学与大数据技术

总学时/实验：40/8

总学分：2.5

执笔人：常亮

审定人：史健婷

审批人：李海华

一、课程性质及教学目标

Hadoop 技术基础为数据科学与大数据技术专业必修的一门专业基础课程。

通过本课程的学习，使学生熟练掌握大数据技术的分布式处理框架与开发环境，培养学生的设计和开发能力，在了解大数据存储、大数据访问、大数据采集、大数据管理和大数据的统计、分析原理的基础上，熟练掌握 Hadoop、HDFS、Hbase、Mapreduce、Zookeeper 等开发环境和技术的使用技能，使学生牢固掌握大数据开发工具和软件平台，为后续的学习和工作打下扎实的理论 and 实践基础。

二、教学内容及要求

（一）大数据和 Hadoop（2 学时）

1. Hadoop 与大数据的关系。
2. 为什么要学习 Hadoop，Hadoop 与云计算的关系。
3. 大数据的几个典型行业应用场景。
4. Hadoop 的发展及生态。

（二）Hadoop 基础知识（4 学时）

1. Hadoop 项目核心模块。
2. RPC 工作原理。
3. MapReduce 工作原理。
4. HDFS 工作原理。
5. YARN 工作原理。

（三）Hadoop 集群环境搭建与配置（6 学时）

1. 安装准备、集群部署、模式搭建。
2. 伪分布模式、全分布模式。
3. 集群测试、配置使用。

（四）Hadoop 分布式文件系统（4 学时）

1. HDFS 架构和原理。
2. HDFS 常用命令。
3. Shell 操作，HDFS 的 API。

（五）Hadoop 的 I/O 操作（4 学时）

1. I/O 序列化类型。
2. Hadoop 的压缩。
3. 基于文件的数据结构。

（六）MapReduce 分布式计算框架（4 学时）

1. 工作原理，编程组件。
2. 运行模式，性能优化策略。

数据科学与大数据技术专业教学大纲

3. 倒排索引，去重。

(七) MapReduce 高级编程 (4 学时)

1. 计数器，最值，全排序。

2. 经典案例-倒排索引。

3. 数据去重、TopN。

(八) Zookeeper 分布式服务 (4 学时)

1. Zookeeper 概述。

2. 数据模型，Watch 机制，选举机制。

3. Zookeeper 部署，Shell 操作。

4. 数据发布与订阅，统一命名服务，分布式锁。

本课程重点内容：大数据和 Hadoop、Hadoop 基础知识、Hadoop 集群环境搭建与配置、Hadoop 分布式文件系统。

本课程难点内容：Hadoop 的 I/O 操作、MapReduce 分布式计算框架、MapReduce 高级编程、Zookeeper 分布式服务。

三、过程考核方式

课程采取期末考核与平时表现相结合的方式进行，平时表现占40%，包括出勤情况，学习态度，课内测试等，期末考核考试占60%。

四、教学方法

“体验式”教学，学生带笔记本电脑进入课堂，教师以PPT和板书形式开展本课程教学。

五、各教学环节学时分配

授课内容	第一课堂	第二课堂	实验	实践实训	合计
大数据和 Hadoop	2				2
Hadoop 基础知识	4				4
Hadoop 集群环境搭建与配置	6		2		8
Hadoop 分布式文件系统	4		2		6
Hadoop 的 I/O 操作	4				4
MapReduce 分布式计算框架	4		2		6
MapReduce 高级编程	4		2		6
Zookeeper 分布式服务	4				4
合计	32		8		40

六、参考书

[1] 黑马程序员. Hadoop 大数据技术原理与应用. 北京：清华大学出版社, 2019

[2] 王宏志, 李春静. Hadoop 集群程序设计与开发. 北京：人民邮电出版社, 2018

[3] 尚沙勒·辛格, 张华. 精通 Hadoop3. 北京：清华大学出版社, 2020

七、说明

数据科学与大数据技术专业教学大纲

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《Hadoop 技术基础》（实验）教学大纲

课程代码：176007 课程类别：专业基础 适用专业：数据科学与大数据技术
课程属性：课内实验 课程总学时：40 总学分：2.5 实验学时：8 学分：0.5
执笔人：常亮 审定人：史健婷 审批人：李福军

一、实验性质及教学目标

《Hadoop 技术基础》（实验）属于课内实验，是《Hadoop 技术基础》理论课程的有效辅助，强调实践性和综合性。通过该课程的学习，学生将掌握 Hadoop 技术的基本知识和技能，具备使用 Hadoop 技术进行大数据存储和处理的能力，并为后续的学习和工作打下坚实的基础。

二、实验项目

序号	实验项目名称	学时	实验类型	实验要求
1	Hadoop 集群环境搭建与配置	2	验证性	必做
2	Hadoop 分布式文件系统	2	验证性	必做
3	MapReduce 分布式计算框架	2	验证性	必做
4	MapReduce 高级编程	2	验证性	必做

三、实验概述

实验 1：Hadoop 集群环境搭建与配置

目的：通过实验，学生能够深入理解 Hadoop 的分布式存储（HDFS）和分布式计算（MapReduce）架构。掌握 Hadoop 集群的搭建和配置方法，包括单机模式、伪分布式模式和完全分布式模式的搭建过程。

原理概述：Hadoop 是一个开源的分布式计算框架，它能够将大规模数据集分割成小的数据块，并将这些数据块存储在集群中的多个节点上。Hadoop 使用 MapReduce 编程模型，将数据的处理过程分为 Map 阶段和 Reduce 阶段。

方法与手段：1. 准备实验所需的硬件和软件环境，包括 Linux 操作系统、Java JDK、Hadoop 安装包等。2. 按照 Hadoop 的安装指南进行安装和配置，包括设置环境变量等。3. 根据实验要求搭建 Hadoop 集群，包括启动 NameNode、DataNode 等进程，以及配置 YARN 资源管理系统。

应得到的实验结果和数据：集群搭建成功。通过浏览器访问 Hadoop 的 Web 界面（如 NameNode 的 Web 界面），能够看到集群的状态信息，包括节点数量等。

实验 2：Hadoop 分布式文件系统

目的：通过实验，学生能够深入理解 HDFS 的分布式存储架构，掌握 HDFS 的基本操作，包括文件的创建、删除、查看、上传、下载等。

原理概述：HDFS 是一个被设计成适合运行在通用硬件上的分布式文件系统，具有高度容错性和高吞吐量。它采用主从架构，包括一个 NameNode 和多个 DataNode。NameNode 负责管理 HDFS 的命名空间和客户端对文件的访问，而 DataNode 则负责存储实际的数据块。HDFS 通过将大文件分割成多个小数据块，并将这些数据块存储在多个 DataNode 上，实现了数据的分布式存储和并行处理。

方法与手段：1. 启动 Hadoop 集群，包括启动 NameNode、DataNode 等进程，以及启动 YARN 资源管理系统（如果实验涉及 MapReduce 编程）。2. 使用 HDFS 命令行工具或编程 API 进行 HDFS 的基本操作，如创建目录、上传文件、查看文件列表、下载文件等。

应得到的实验结果和数据：通过查看 Hadoop 集群的启动日志和 Web 界面（如 NameNode 的 Web 界面），确认 HDFS 环境搭建成功，各节点正常运行。成功执行 HDFS 的基本操作，如创建目录、上传文件、查看文件列表、下载文件等，并能通过命令行或编程 API 查看操作结果。

实验 3: MapReduce 分布式计算框架

目的：通过实验，学生能够深入理解 MapReduce 的编程模型和执行流程，包括 Map 阶段和 Reduce 阶段的作用及数据转换过程。掌握使用 MapReduce 编写分布式计算程序的方法，包括编写 Mapper 和 Reducer 类、配置作业参数等。

原理概述：MapReduce 是一种编程模型，用于大规模数据集（大于 1TB）的并行运算。它将复杂的、运行于大规模集群上的并行计算过程高度地抽象为了两个函数：Map 和 Reduce。Map 负责把任务分解成多个任务，Reduce 负责把分解后多任务处理的结果汇总起来。MapReduce 通过输入分割、Map 映射、Shuffle 排序、Reduce 归约等步骤实现数据的分布式处理。

方法与手段：1. 根据实验要求，使用 Java 等编程语言编写 MapReduce 程序。程序需要包含 Mapper 类、Reducer 类以及作业配置部分。2. 将编写好的 MapReduce 程序打包成 jar 文件，并通过 Hadoop 命令行工具提交到 YARN 上运行。

应得到的实验结果和数据：MapReduce 作业能够成功提交到 YARN 上，并顺利完成执行。可以通过 Hadoop 的 Web 界面或命令行工具查看作业的执行状态和结果。检查 MapReduce 作业的输出结果是否符合预期。

实验 4: MapReduce 高级编程

目的：通过高级编程实验，深入理解 MapReduce 框架的高级特性和复杂应用场景，如数据去重、排序等。掌握 MapReduce 编程的高级技巧，包括复杂逻辑的实现、性能优化等，提升在大数据处理领域的编程能力。

原理概述：MapReduce 是一种用于大规模数据集并行处理的编程模型，其核心思想是将复杂的数据处理任务分解为可并行化的小任务，通过 Map 和 Reduce 两个基本操作实现数据的分布式处理。在高级编程实验中，通常会涉及更复杂的数据处理逻辑和更高级的数据结构，如自定义分区函数、排序算法、聚合操作等。

方法与手段：1. 根据实验要求，定义 Map 和 Reduce 函数，实现数据处理逻辑。2. 设置作业的配置参数，如输入/输出路径、Mapper 和 Reducer 的数量、内存限制等。3. 针对特定问题，优化 Map 和 Reduce 函数的实现，提高处理效率和准确性。

应得到的实验结果和数据：MapReduce 作业能够成功提交到 YARN 上，并顺利完成执行。可以通过 Hadoop 的 Web 界面或命令行工具查看作业的执行状态和结果。检查 MapReduce 作业的输出结果是否符合预期。

四、主要仪器设备

PC 机，虚拟机环境等。

五、教学形式

利用多媒体手段总体讲解有关实验的内容要求和注意事项，并且演示关于实验的相应程序，帮助学生从总体上把握每次实验的要求和细节，此后学生根据老师的讲解编制和调试程序，在调试的过程中教师给予积极的帮助和引导，当学生调试成功时，采用一定量的实验数据来验证程序的正确性和合理性，实验结束之后，学生将实验的总体过程详细地记录到实验报告中。

六、考核方式与成绩评定方法

实验成绩由实验报告成绩确定，实验报告成绩总分 100 分。

七、教材及主要参考资料

- [1] 黑马程序员. Hadoop 大数据技术原理与应用. 北京：清华大学出版社, 2019
- [2] 王宏志, 李春静. Hadoop 集群程序设计与开发. 北京：人民邮电出版社, 2018
- [3] 尚沙勒·辛格, 张华. 精通 Hadoop3. 北京：清华大学出版社, 2020

八、说明

在执行大纲过程中应结合专业的特点和学生的实际情况，可适当调整授课内容与进度。

《计算机网络》教学大纲

课程代码：172009 课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：56/8 总学分：3.5

执笔人：**聂红梅** 审核人：**史健婷** 审批人：**李海华**

一、课程性质及教学目标

本课程为数据科学与大数据技术专业开设的一门专业基础课程。

教学目标是网络技术基本理论与基本方法为主干，通过本课程的学习，使学生理解计算机网络的基本概念、基本原理、体系结构/参考模型，掌握典型网络技术以及典型网络协议等网络基础知识；培养学生具备网络应用、协议分析以及运用网络知识与技术解决实际问题的能力。

二、教学内容及要求

（一）计算机网络概论（10 学时）

包括内容：计算机网络概论、网络体系结构与网络协议

1. 计算机网络的形成与发展过程。
2. 计算机网络的定义与分类。
3. 计算机网络的组成与结构的基本概念。
4. 计算机网络拓扑构型的定义、分类与特点。
5. 分组交换的基本概念。
6. 网络体系结构与网络协议的基本概念。

（二）物理层（6 学时）

包括内容：物理层、数据通信的基本概念

1. 物理层与物理层协议的基本概念。
2. 数据通信的基本概念。
3. 传输介质类型及主要特性。
4. 数据编码的类型和基本方法。
5. 基带传输与频带传输的基本概念。
6. 多路复用技术的分类与特点。
7. 同步数字体系 SDH 的基本概念。
8. 接入技术的基本概念。

（三）数据链路层（10 学时）

包括内容：数据链路层

1. 误码率的定义与差错控制方法。
2. 数据链路层的基本概念。
3. MAC 层协议的基本概念。
5. IEEE 802.3 协议与以太网工作原理。
6. IEEE 802.11 协议与 WiFi 工作原理。

（四）第四单元：TCP/IP 协议（22 学时）

包括三部分内容：网络层、传输层、应用层

第一部分：网络层（10 学时）

数据科学与大数据技术专业教学大纲

1. 网络层与网络互联的基本概念。
2. Ipv4 协议的基本内容。
3. IP 地址、路由算法与路由协议的基本概念。
4. 地址解析 ARP 的基本概念。
5. IPv6 的基本内容。
6. 移动 IP 协议的基本概念。
7. ICMP 与 IGMP 协议的基本概念。

第二部分：传输层（6 学时）

1. 网络环境中分布式进程通信的基本概念。
2. 进程通信中客户/服务器模式的基本概念。
3. 传输层的基本功能与服务质量的的基本概念。
4. UDP 协议的基本内容。
5. TCP 协议的基本内容。

第三部分：应用层（6 学时）

1. 互联网应用发展与应用层协议的分类。
2. C/S 与 P2P 模式的特点。
3. DNS、DHCP 的工作原理。
4. SMTP、FTP 与 TELNET 的工作原理。
5. Web 与搜索引擎的工作原理。
6. 应用层协议的分析方法。

本课程重点内容：计算机网络的定义与分类、组成与结构、拓扑构型，分组交换；协议、层次、接口与网络体系结构的基本概念，OSI 参考模型及各层的基本服务功能，TCP/IP 参考模型的层次划分、各层的基本服务功能及主要协议；数据链路层的基本概念、误码率的定义与差错控制方法；网络层、传输层、应用层。

本课程难点内容：分组交换、循环冗余编码与差错控制机制、数据编码、多路复用技术、协议、层次、接口与网络体系结构的基本概念，OSI 环境中的数据传输过程、TCP/IP、IP 地址、Ethernet 局域网的基本工作原理、网桥的基本工作原理、路由算法与路由协议的基本概念。

三、过程考核方式

课程采取平时表现、实验与期末考试相结合的方式进行考核。平时表现（30%），实验（20%），期末考试（50%）。

四、教学方法

问题驱动、案例教学、过程管理、互动教学是本课程采用的主要教学方法。其中案例教学在涉及工程实践的内容部分采用，问题驱动、过程管理、互动教学在全课堂教学采用。积极采用现代化的教学手段教学，如多媒体、动画、网络教学平台等。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
计算机网络概论	10				10

数据科学与大数据技术专业教学大纲

物理层	6		2		8
数据链路层	10		2		12
TCP/IP 协议	22		4		26
合计	48		8		56

六、参考书

- [1] 吴功宜, 吴英. 计算机网络 (第 4 版). 北京: 清华大学出版社, 2017
- [2] 谢希仁. 计算机网络 (第 7 版). 北京: 电子工业出版社, 2017
- [3] James F. Kurose, Keith W. Ross 著, 陈鸣译. 计算机网络: 自顶向下方法 (原书第 7 版). 北京: 机械工业出版社, 2018

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《计算机网络》（实验）教学大纲

课程代码：172009 课程类别：专业基础 适用专业：数据科学与大数据技术
课程属性：课内实验 课程总学时：56 总学分：3.5 实验学时：8
执笔人：*聂红梅* 审定人：*史健婷* 审批人：*李海华*

一、实验性质及教学目标

《计算机网络》实验是数据科学与大数据技术专业基础课程的课内实验，主要培养学生计算机网络的三个基本能力：

一是以太网组网（2学时），包括双绞线制作与简单以太网组建，培养学生实际动手组网的能力。

二是协议分析（2学时），包括计算机网络中最主要的几个协议：IP、ARP、TCP，通过抓包，让学生分析相应协议，以更好掌握这些重要协议，并培养学生协议分析和通过协议分析解决实际问题的能力。

三是路由器与交换机配置（4学时），路由器与交换机是计算机网络工程中最关键的两种核心设备，路由器与交换机的配置是网络工程的基本技能。通过实验，学生能够掌握网络工程的基本技能，提高实际能力。

二、实验项目

序号	实验项目名称	学时	实验类型	实验要求
1	双绞线的制作与简单以太网组建	2	综合性	必做，4-5人一组
2	ARP协议分析与TCP协议特性分析	2	综合性	必做，4-5人一组
3	交换机路由器基本配置、VLAN基本配置	2	综合性	必做，4-5人一组
4	路由协议基本配置	2	综合性	必做，4-5人一组

三、实验概述

1. 双绞线的制作与简单以太网组建（2学时）

（1）实验目的

- ①掌握双绞线与RJ-45水晶头的接法，学会双绞线的测试方法。
- ②掌握简单以太网的组建方法，包括连接、对等网络配置、IP配置等。

（2）实验要求

简单以太网组建需要分组完成，每组不超过5人

（3）主要仪器

超5类双绞线、RJ-45水晶头、双绞线压线钳、双绞线测试器、PC机、小型非管理型交换机

（4）实验内容

- ①在双绞线的两头卡接水晶头，并测试其连接的正确性。
- ②简单以太网组建，包括：利用做好的双绞线进行物理拓扑连接、IP配、对等网配置、连通性测试、简单网络工具命令使用（Ping、IPconfig、arp）。

2. ARP协议分析与TCP协议特性分析（2学时）

（1）实验目的

- ①深入理解地址解析协议（ARP）的工作原理。理解IP和以太网协议的关系，掌握IP地址和MAC地址的映射机制，搞清楚IP报文是如何利用底层的以太网帧进行传输的。

②深入理解 TCP 的如下重要机制的工作原理：利用序号和确认号实现可靠数据传输，接收方通告的流量控制。掌握用 Wireshark 分析 TCP 踪迹文件的技能。

(2) 实验要求

需要分组局域网环境完成，并且局域网可以访问 Internet, 每组不超过 5 人。

(3) 主要仪器

超 5 类直通线、PC 机、小型非管理型交换机、Internet 访问设备。

(4) 实验内容

①查看本机硬件地址、使用 ARP 命令、分析 ARP 协议工作过程。

②俘获本机与远程服务器的 TCP 踪迹文件、熟悉 TCP 踪迹文件、分析 TCP 序列/应答编号和流量控制、分析应用层内容。

3. 交换机路由器基本配置、VLAN 基本配置 (2 学时)

(1) 实验目的

掌握交换机、路由器的基本配置，理解基于交换机端口的 VLAN 划分。

(2) 实验要求

使用网络模拟器软件 Packet Tracer 单机配置或实际配置，实际配置需要分组完成，每组不超过 5 人。

(3) 主要仪器

PC 机。实际配置还需路由器、交换机、超 5 类直通线和交叉线、配置线缆。

(4) 实验内容

①交换机、路由器的基本配置。

②跨交换机 VLAN 配置。

4. 路由协议基本配置 (2 学时)

(1) 实验目的

①掌握利用路由器的常用配置命令对路由器进行基本配置。

②掌握路由器的静态路由、RIP 协议、OSPF 协议配置，加深对静态路由、RIP 协议、OSPF 协议的理解。

(2) 实验要求

使用网络模拟器软件 Packet Tracer 单机配置或实际配置，实际配置需要分组完成，每组不超过 5 人。

(3) 主要仪器

PC 机。实际配置还需路由器、超 5 类直通线和交叉线、配置线缆。

(4) 实验内容

①利用路由器的常用配置命令对路由器进行基本配置。

②静态路由配置。

③RIP 协议配置。

④OSPF 协议配置。

四、主要仪器设备

数据科学与大数据技术专业教学大纲

PC 机每人 1 台，超 5 类双绞线每人 1 根、RJ-45 水晶头每人至少 2 个、双绞线压线钳每人 1 把、双绞线测试器每组 1 台、小型非管理型交换机每组 1 台、二层交换机每组 1 台、三层交换机每组 1 台、路由器每组 2 台、足够长直通线每台 PC1 根、交叉线每组 1 根、设备配置线每台交换机和路由器 1 根、每台 PC 安装抓包软件 Wireshark、每台 PC 安装网络模拟器软件 Packet Tracer。

五、教学形式

在网络实验室实际完成，其中：组网类实验每组实际动手完成。协议分析类实验每组搭建以太网环境，通过抓包软件 Wireshark 抓包并分析相应协议踪迹文件完成。设备配置类实验每组至少 2 名同学（自愿）实际配置，其他同学通过网络模拟器软件 Packet Tracer 完成。

六、考核方式与成绩评定方法

实验成绩 20 分，根据学生实验出勤、实验完成情况和实验报告情况综合考核。

七、教材及主要参考资料

- [1] 吴功宜, 吴英. 计算机网络（第 4 版）. 北京：清华大学出版社, 2017
- [2] 谢希仁. 计算机网络（第 7 版）. 北京：电子工业出版社, 2017
- [3] James F. Kurose, Keith W. Ross 著, 陈鸣译. 计算机网络：自顶向下方法（原书第 7 版）. 北京：机械工业出版社, 2018

八、说明

实验时，要根据实验设备情况做出微调，以保证实验开出率和实验效果。根据设备更新情况，要及时修改实验大纲。

《Python 程序设计》教学大纲

课程代码：175021 课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：32 总学分：2

执笔人：王光群 审定人：史健婷 审批人：李海华

一、课程性质及教学目标

Python 程序设计是数据科学与大数据技术专业必修课程，Python 是一种解释型、面向对象、动态数据类型的高级程序设计语言。注重锻炼学生的编程应用、数据分析能力等所需的基础知识和基本能力。

通过本课程的学习，学生可以掌握基础的 Python 开发方法，通过一系列的 Python 语言编程训练项目，使得学生能够理解 Python 的编程模式，验证、理解直至熟练运用课堂所学知识，培养学生具有一定的 Python 语言编程分析理解和应用实践能力。

二、教学内容及要求

（一）课程导论（2 学时）

1. Python 发展历史，特点，应用领域以及版本信息。
2. Python 开发环境的配置和常用工具。

（二）Python 基础语法和常用语句（8 学时）

1. 常用运算符的运算规则与优先次序；循环结构构造；不同循环结构的相互转换；不同结构的相互嵌套。
2. 算法与算法表达方法，用程序流程图来表达算法的三种基本结构的方法；基本数据类型，变量的定义与赋值，基本的输入/输出，能够构造分支程序和循环结构。

（三）列表、元组和字典（8 学时）

1. 掌握列表、元组以及字典之间的区别与联系，在实际程序设计中如何选择。
2. 了解列表、元组和字典，主要包括列表与集合的关系，列表的定义方法；列表的索引作用，使用索引，列表元素的方法；列表的循环遍历，列表的常见操作其中包括在列表中增加元素，在列表中查找元素，在列表中修改元素，在列表中删除元素。元组的介绍，字典的定义方法以及列表的遍历以及字典的常见操作。

（四）函数（8 学时）

1. 掌握函数的声明、定义与调用，变量作用域，函数的参数传递，函数的返回值，递归函数和匿名函数。
2. 了解函数，主要包括函数与程序的关系、函数的声明、定义与调用、变量的作用域的概念，局部变量的用法，递归函数和匿名函数的声明和定义，并通过函数返回值获取函数的运算结果，常用的日期和时间的函数。

（五）面向对象（6 学时）

1. 掌握构造方法和析造方法的使用，并能正确使用。
2. 了解面向对象，使用类描述显示世界实体（物理或逻辑）的方法；面向对象基本概念：对象、类、属性、方法、继承、派生；类与结构体的差异；Python 中用类创建对象，并添加属性、构造方法和析造方法的使用、运算符重载，会定制对象字符串的形式。

数据科学与大数据技术专业教学大纲

本课程重点内容：常用运算符的运算规则与优先次序；循环结构构造；不同循环结构的相互转换；不同结构的相互嵌套。使用列表索引操作列表元素。函数的声明、定义与调用，变量作用域，函数的参数传递，函数的返回值，递归函数和匿名函数。熟悉 self 的使用方法；掌握构造方法和析造方法的使用。

本课程难点内容：不同结构的相互嵌套，特别是多重循环的循环变量的变化。列表、元组以及字典之间的区别与联系，在实际程序设计中如何选择。实参与形参的关系，函数的返回值，递归函数。构造方法和析造方法的使用，并能正确使用。

三、过程考核方式

教师可根据学生的情况灵活采用考核的形式，主要考核学生的实际动手操作能力，可以采用闭卷考试形式或课题报告形式。

闭卷考试（或课程报告）形式，试卷成绩（或课程报告）占总成绩 70%，平时成绩占总成绩 30%：其中作业占 10%，课堂表现占 10%，回答问题情况占 10%。

四、教学方法

采用边学习边练习操作相结合。引导学生理解基本概念、理论和方法，掌握基本操作、基本技能和学习方法。鼓励学生自学和实践、主动获取知识，激发学生的学习主动性和兴趣。培养学生分析问题、解决问题的能力、自学能力、实践和创新能力。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
课程导论	2				2
Python 基础语法和常用语句	8				8
列表、元组和字典	8				8
函数	8				8
面向对象	6				6
合计	32				32

六、参考书

- [1] 董付国. Python 程序设计基础（第 2 版）. 北京：清华大学出版社, 2018
- [2] 黑马程序员. 解析 Python 网络爬虫. 北京：中国铁道出版社, 2018

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《算法分析与设计》教学大纲

课程代码：173013 课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：40 总学分：2.5

执笔人：丁文涛 审定人：史健婷 审批人：李海华

一、课程性质及教学目标

本课程是数据科学与大数据技术专业的一门专业基础课程。

教学目标是使学生了解计算机常用算法求解问题的特征，掌握几种重要的算法设计的方法：分治法，动态规划，贪心法，回溯法，分枝限界法，使学生掌握常用算法的基本思想，基本步骤和适用的问题，培养学生算法设计分析的能力。

二、教学内容及要求

（一）算法概述（2学时）

1. 了解算法与程序的概念。
2. 掌握算法复杂性分析及其有关的概念。

（二）递归与分治策略（8学时）

1. 理解递归的概念。
2. 了解分治法的基本思想。
3. 掌握线性时间选择，最接近点对问题的算法。
4. 理解合并排序和快速排序算法。

（三）动态规划（8学时）

1. 掌握动态规划算法的概念和步骤。
2. 掌握动态规划算法的基本要素。
3. 掌握最长公共子序列、0-1 背包问题的算法设计和分析。

（四）贪心算法（8学时）

1. 掌握贪心算法的概念，掌握贪心算法的基本要素。
2. 了解最优装载问题的算法分析，哈夫曼编码的算法分析。
3. 了解单源最短路径的 Dijkstra 算法的设计与分析。
4. 了解最小生成树的 Prim 和 Kruskal 算法的设计与分析。

（五）回溯法（8学时）

1. 掌握回溯法的算法框架。
2. 掌握装载问题，批处理作业调度，0-1 背包问题。
3. 图的 m 着色问题的算法设计与分析，了解 n 后问题的回溯法的算法设计与分析。

（六）分支限界法（6学时）

1. 掌握分支限界法的基本思想，掌握单源最短路问题。
2. 0-1 背包问题的分支限界法分析，了解旅行售货员问题的算法设计与分析。

本课程重点内容：分治法的一般方法，归并分类和快速分类算法，贪心设计策略的一般方法，归并算法，最小生成树算法，动态规划的一般方法，0/1 背包问题，贪心算法的概念和基本要素，回溯法的算法框架，分支限界法的基本思想。

数据科学与大数据技术专业教学大纲

本课程难点内容：快速分类算法，选择问题算法，最小生成树算法，最短路径算法，流水线调度问题，最优装载问题的算法分析，哈夫曼编码的算法分析，批处理作业调度，0-1 背包问题的算法设计与分析，单源最短路问题，0-1 背包问题的分支限界法分析。

三、过程考核方式

本课程采取学习表现与期末考核相结合的方式进行，学习表现占 30%，包括课堂表现，作业完成情况，期末理论知识开卷考试占 70%。

四、教学方法

利用多媒体教学，采取学练结合的方式进行教学。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
算法概述	2				2
递归与分治策略	8				8
动态规划	8				8
贪心算法	8				8
回溯法	8				8
分支限界法	6				6
合计	40				40

六、参考书

- [1] 王晓东编. 计算机算法设计与分析（第 4 版）. 北京：电子工业出版社, 2012
- [2] 吕国英编著. 算法设计与分析（第 2 版）. 北京：清华大学出版社, 2009
- [3] 吴哲辉等编. 算法设计方法. 北京：机械工业出版社, 2008

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《分布式计算系统原理》教学大纲

课程代码：176008

课程类别：专业基础

适用专业：数据科学与大数据技术

学时：40

学分：2.5

执笔人：常亮

审定人：史健婷

审批人：李福军

一、课程性质及教学目标

本课程是数据科学与大数据技术专业的一门必修课程。学生通过学习本门课程了解分布式系统的设计原理和实践知识，要求学生理解分布式系统的概念、原理、挑战与目标、以及中间件相关技术，掌握基本的分布式算法，建立软件的基本框架思想，具备设计、开发和评价以及应用分布式系统的基本能力，树立软件开发的规模经济分析意识。

通过学习使学生了解分布式计算系统的发展历史与挑战、基本概念、基本原理、分布式算法及中间件涉及到的技术，培养学生分布式系统的分析与实践能力，满足学生从事分布式系统设计、开发、评价及应用等工作的能力需求，培养学生独立分析问题、解决问题的能力。

二、教学内容及要求

（一）分布式概述（4 学时）

1. 了解分布式计算系统的产生背景、发展过程、目标和分布式系统的应用架构演变。
2. 掌握分布式系统的定义、主要特征及衡量指标。
3. 掌握分布式系统的硬件体系、软件体系结构及异构性。

（二）进程通信（4 学时）

1. 回顾进程与线程的基本知识。
2. 掌握掌握进程间通信的方式和程序接口原型、事件同步。
3. 掌握 Socket 机制、三种类型及 TCP Socket 和 UDP Socket 的编程实现。
4. 掌握代码迁移的知识。

（三）分布式系统通信（10 学时）

1. 掌握 RPC 的基本概念、原理、目标和 RPC 通信流程。
2. 对比 RPC，了解 RMI 相关知识。
3. 掌握面向消息通信的通信类型、消息中间件的定义。
4. 掌握 MQ 模型、体系结构、作用及面向消息的持久通信模式的基本原理。
5. 掌握 Kafka 实现原理和 Zookeeper 的工作机制、特点和应用场景。
6. 掌握面向流通信的数据流和传输模式。

（四）命名服务（8 学时）

1. 掌握命名服务中涉及到的基本概念。
2. 掌握无层次命名、结构化命名和基于属性命名的实现方法。

（五）同步与互斥（6 学时）

1. 了解时钟同步的基本知识和掌握典型时钟同步算法原理。
2. 掌握典型分布式互斥算法原理和了解互斥算法需要满足的条件和衡量互斥算法性能的指标。
3. 掌握分布式选举算法原理。
4. 掌握分布式共识和快照的基本知识。
5. 掌握分布式快照算法。

数据科学与大数据技术专业教学大纲

(六) 分布式数据存储 (4 学时)

1. 掌握 CAP 理论和 BASE 理论、分布式数据分布方式, 掌握数据复制类型及实现原理和一致性模型。
2. 掌握分布式缓存的基本概念和实现原理及应用。

(七) 分布式事务 (4 学时)

1. 掌握分布式事务的实现方法。
2. 掌握分布式锁的主流实现方法。

本课程重点内容: 软件体系结构中的中间件模型, IPC、Socket、TCP 与 UDP 编程、代码迁移, 消息中间件、RPC、MQ, Kafka 实现原理和 Zookeeper 的工作机制、特点, 无层次命名、结构化命名的实现方法, 时钟同步算法、分布式互斥算法、分布式选举算法和分布式快照算法, 数据分布方式、数据复制、一致性模型, 2PC、3PC、基于分布式消息的最终一致性方法, 分布式锁。

本课程难点内容: 分布式系统的异构性, Socket 机制、代码迁移, RPC 原理及通信流程、Kafka 和 ZK 分布式锁服务的实现原理, 基于 DHT 方法的无层次命名, 向量时钟、状态机复制, Quorum 机制、水平分区哈希算法、一致性模型, 分布式锁。

三、过程考核方式

课程最终成绩按百分制评定。成绩包括平时成绩和期末考试成绩, 平时成绩占 40%, 期末考试成绩占 60%, 平时成绩由课内作业、测试、日常评价等组成, 可根据情况适当修改考核方式和比例。

四、教学方法与手段

教学中采用案例式、启发式等多种教学方法, 以学生为主体, 以启发学生的思维为核心, 调动学生的学习主动性和积极性; 通过课内讨论和课外文献资料查阅等方式拓宽学生的视野和思维空间, 激发学生对分布式系统课程的学习兴趣, 加深对理论教学的理解, 并培养学生的自主学习和终身学习意识; 课堂教学采用多媒体教学手段, 配以形象直观的图表、生动的实例, 浅显通俗地对概念、原理、算法等进行解释, 以此增加教学的直观性, 让学生充分理解分布式系统的原理和实现机制。

五、各教学环节学时分配

教学内容	第一课堂	第二课堂	实验	实践实训	合计
分布式概述	4				4
进程通信	4				4
分布式系统通信	10				10
命名服务	8				8
同步与互斥	6				6
分布式数据存储	4				4
分布式事务	4				4
合计	40				40

六、参考书

- [1] 辛春生等. 分布式系统原理与范型 (第二版). 北京: 清华大学出版社, 2015
- [2] 林伟伟, 刘波. 分布式计算、云计算及大数据. 北京: 机械工业出版社, 2015

[3] 胡建平, 胡凯. 分布式计算系统导论-原理与组成. 北京: 清华大学出版社, 2014

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《大数据分析与应用》教学大纲

课程代码：176009 课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：40/8 总学分：2.5

执笔人：王光群 审定人：史健婷 审批人：李海华

一、课程性质及教学目标

《大数据分析与应用》是数据科学与大数据专业的一门专业必修课。

本课程注重培养学生在大数据采集、存储、计算、处理、分析、挖掘及可视化全过程的系统性认识和实践性应用，密切结合大数据技术体系的基础、前沿理论与实际案例，使学生在理解理论知识和实践的基础上掌握当前大数据分析的方法、技术和工具。

二、教学内容及要求

（一）数据分析概述（4学时）

1. 掌握数据分析的概念
2. 掌握数据分析的流程
3. 了解数据分析的应用场景
4. 了解数据分析的常用工具
5. 了解 Python 数据分析的优势
6. 了解 Python 数据分析的常用类库
7. 了解 Python 的 Anaconda 发行版
8. 在 Windows 操作系统上安装 Anaconda

（二）NumPy 数值计算基础（4学时）

1. 掌握创建数组对象
2. 了解生成随机数
3. 通过索引访问数组
4. 变换数组的形态
5. 掌握创建 NumPy 矩阵
6. 掌握 ufunc 函数
7. 读写文件

（三）Matplotlib 数据可视化基础（8学时）

1. 掌握 pyplot 的基础语法
2. 设置 pyplot 的动态 rc 参数
3. 了解绘制散点图
4. 了解绘制折线图
5. 了解绘制直方图
6. 了解绘制饼图

（四）pandas 统计分析基础（8学时）

1. 了解堆叠合并数据
2. 掌握主键合并数据
3. 了解重叠合并数据

数据科学与大数据技术专业教学大纲

4. 掌握检测与处理重复值
5. 掌握检测与处理缺失值
6. 了解检测与处理异常值
7. 离差标准化数据
8. 掌握标准差标准化数据
9. 了解小数定标标准化数据
10. 了解哑变量处理类别型数据, 离散化连续型数据。

(五) 使用 scikit-learn 构建模型 (8 学时)

1. 掌握加载 datasets 模块中的数据集
2. 掌握将数据集划分为训练集和测试集
3. 掌握使用 sklearn 转换器进行数据预处理与降维
4. 掌握使用 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, 评价回归模型

本课程重点内容: 掌握数据分析的概念、流程与应用场景; 掌握 NumPy 创建多维数组与生成随机数的方法, 掌握数组的索引与变换。掌握 pyplot 常用的绘图参数的调节方法, 掌握子图的绘制方法。掌握数据合并的原理与方法, 掌握数据清洗的基本方法。掌握 sklearn 转换器的使用方法, 掌握 sklearn 估计器的使用方法。

本课程难点内容: 掌握 NumPy 读写文件的方法和常用的统计分析的函数。掌握绘制图形的保存与展示方法, 掌握散点图和折线图的作用与绘制方法。掌握基本数据标准化的方法, 掌握常用的数据转换方法。掌握聚类模型的构建与评价, 掌握分类模型的构建与评价。

三、过程考核方式

考核方式: 根据大数据分析及应用课程特点和培养目标, 该课程的期末成绩评定由两部分组成, 即: 学生课堂表现 (包括问题讨论、回答问题和作业等方面) 占总成绩 40%、期末考试 (或课程报告) 占总成绩的 60%。

四、教学方法

采用边学习边练习操作相结合。引导学生理解基本概念、理论和方法, 掌握基本操作、基本技能和学习方法。鼓励学生自学和实践、主动获取知识, 激发学生的学习主动性和兴趣。培养学生分析问题、解决问题的能力、自学能力、实践和创新能力。

五、各教学环节学时分配

授课内容 (知识点)	第一课堂	第二课堂	实验	实践实训	合计
数据分析概述	4				4
NumPy 数值计算基础	4		2		6
Matplotlib 数据可视化基础	8		2		10
pandas 统计分析基础	8		2		10

数据科学与大数据技术专业教学大纲

使用 scikit-learn 构建模型	8		2		10
合计	32		8		40

六、参考书

[1] 黄红梅著. PYTHON 数据分析与应用. 北京: 人民邮电出版社, 2018

[2] 张良均. PYTHON 数据分析与挖掘实战. 北京: 机械工业出版社, 2016

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《大数据分析与应用》（实验）教学大纲

课程代码：176009 课程类别：专业基础 适用专业：数据科学与大数据技术
课程属性：课内实验 课程总学时：40 总学分：2.5 实验学时：8 学分：0.5
执笔人：王光群 审定人：史健婷 审批人：李海华

一、实验性质及教学目标

《大数据分析与应用》是数据科学与大数据技术专业的专业必修课之一，此门课程的主要目的在于让学生了解大数据技术对传统行业的冲击和影响。本课程在专业课程体系中占有较为重要的位置。大数据分析与应用的内课内实验是辅助理论课的上机实验，旨在强化基本理论，利用基本理论提高解决实际问题的能力。

二、实验项目

序号	实验项目名称	学时	实验类型	实验要求
1	猜数游戏	2	验证性	必做
2	分析国民生产总值特征分布	2	验证性	必做
3	掌握 DataFrame 的常规操作	2	验证性	必做
4	构建评价聚类模型	2	综合性	必做

三、实验概述

1. 实验项目 1 猜数游戏

实验目的与要求：熟悉 Python 语言编程环境和重要的 Python 库。

实验操作要点：利用 python 语言中程序控制语句如 for, if 语句等等实现猜数游戏。

2. 实验项目 2 分析国民生产总值特征分布

实验目的与要求：掌握 Matplotlib 绘图工具。

实验操作要点：设定颜色、标记、线型、刻度、标签、图例、注解，根据给定数据，绘制相应图表。

3. 实验项目 3 掌握 DataFrame 的常规操作

实验目的与要求：掌握 DataFrame、numpy、pandas 输入输出对象等常规操作。

实验操作要点：利用 DataFrame、numpy、pandas 库读写文本格式数据、逐块读写文本文件、将数据写出到文本格式、手工处理分隔符格式。

4. 实验项目 4 构建评价聚类模型

实验目的与要求：掌握建立聚类模型，使用 Python 编程实现聚类分析算法，并可视化展示。。

实验操作要点：聚类分析算法的选择、关键参数的设定、不同算法分类性能的对比、分类结果的可视化。

四、主要仪器设备

PC 机，python 语言软件环境

五、教学形式

教学形式：本门课程的实验采用提前预习的机制，要求学生提前熟悉实验题目。学生根据老师的讲解编制和调试程序，在调试的过程中教师给予积极的帮助和引导，当学生调试成功时，教师提供一定量的实验数据来验证程序的正确性和合理性，实验结束之后，学生将实验的总体过程详细地

记录到实验报告中。

六、考核方式与成绩评定方法

本实验课程不单独组织考试，而是将部分实验内容列入期末考试试卷中进行考试。另外，实验成绩将列入平时成绩中，实验成绩主要由实验报告成绩组成。

七、教材及主要参考资料

[1] 黄红梅著. PYTHON 数据分析与应用. 北京：人民邮电出版社, 2018

[2] 张良均. PYTHON 数据分析与挖掘实战. 北京：机械工业出版社, 2016

八、说明

应根据实际的实验教学学时的要求和学生们的具体特点与当时的条件，可适当调整实验授课内容与进度。

《NoSQL 数据库原理》教学大纲

课程代码：176010 课程类别：专业基础 适用专业：数据科学与大数据技术
总学时：56/8 总学分：3.5
执笔人：常亮 审定人：史健婷 审批人：李福军

一、课程性质及教学目标

本课程是数据科学与大数据技术专业本科生必修的一门专业基础课程。

NoSQL 数据库泛指非关系型数据库，是大数据场景下分布式数据存储和管理的核心技术。本课程旨在引导学生掌握 NoSQL 数据库的基本概念和技术，典型的 NoSQL 数据库的使用和开发方法。通过项目案例，深刻理解大数据环境下传统关系型数据库和 NoSQL 数据库的各自优势和不足，NoSQL 数据库的设计理念，能够针对实际应用场景，分析、选择和使用合适的数据库。

二、教学内容及要求

（一）绪论（4 学时）

1. 了解关系型数据库的发展历史和主要功能。
2. 了解 NoSQL 数据库的发展及其特色。
3. 了解本课程研究的对象、内容，明确本课程的地位。
4. 为什么需要 NoSQL 数据库以及它与传统关系型数据库的区别。
5. 了解 NoSQL 数据库分类，了解主流的 NoSQL 数据库产品和应用场景。

（二）简单而有效的键值型内存数据库 Memcached（4 学时）

1. Memcached 的特点。
2. memcached 安装与演示，memcached 运行相关命令。
3. 一致性 hash 算法描述，主从复制实验。

（三）键值型数据库 Redis（10 学时）

1. Redis 丰富的数据类型，Redis 命令语法，Redis 运行与测试。
2. Redis 的主从复制原理，Redis 测试主从复制。
3. Redis 集群，Redis 使用场景案例。
4. Redis 实现全文检索反向索引，购物篮分析。
5. Redis 的持久化，Redis 快照原理，AOF 追加日志。

（四）处理非结构化数据的 MongoDB（14 学时）

1. MongoDB：面向文档的 NoSQL、MongoDB 的集合。
2. 安装 MongoDB。MongoDB shell。
3. MongoDB 修改器。
4. CentOS 安装 Python，Pip 安装 pymongo 插件，python 测试修改器速度。
5. 瞬时完成与安全操作。MongoDB 创建账户。
6. MongoDB 冷、热备份，MongoRestore 数据恢复，MongoDB 主从复制。
7. 范式化与反范式化，MongoDB 的范式化与反范式化，关系数据库情形下的一致性，MongoDB 的模式设计。

（五）列式数据库 Cassandra（12 学时）

1. Cassandra 的安装，数据模型的演进，Cassandra 列族和列，Cassandra 的操作。

数据科学与大数据技术专业教学大纲

2. Cassandra 集群环境搭建和测试, 验证 Cassandra 集群各节点状态, cqlsh 的启动方式。
3. Cassandra 集群机制。
4. Cassandra 数据更新, Cassandra 读修复, Nodetool 的使用。

(六) 面向图的数据库 Neo4J (4 学时)

1. 图数据库的概念, 图数据库的应用场景, 图数据库 Neo4J 介绍与安装。
2. 图数据库 Neo4J 的配置与操作。

本课程重点内容: 学习 NoSQL 数据库的使用方法, 熟练地进行 NoSQL 数据库的部署和常规管理; 掌握 NoSQL 数据库的开发接口, 能够进行分布式数据管理项目的设计与开发。通过课程学习, 使学生掌握分布式数据管理的特点和常用技术, 具有一定的分布式数据库分析、设计、管理能力, 为未来从事大数据的管理和应用开发打下良好的基础。

本课程难点内容:

1. 了解 NoSQL 数据库的应用背景, NoSQL 与传统关系型数据库的差异, 理解 NoSQL 数据库的基本概念和方法, 掌握大数据领域数据管理、开发所需要的基础理论、工程基础知识和常规技术, 具备解决大数据领域复杂工程问题的基本能力。

2. 了解 memcached、Redis、Mongodb、Cassandra、Neo4J 五种典型的 NoSQL 数据库的各自特点和应用场景, 并能够熟练使用。针对具体工程项目, 能够恰当选择 NoSQL 数据库, 进行部署、使用和管理, 并运用其对大数据领域复杂工程问题进行模拟和实现。了解大数据领域数据存储与管理的前沿和发展动向。

3. 理解典型 NoSQL 数据库的设计原理, 熟悉开发接口, 能够分析、设计、实施分布式数据库, 具备以分布式数据库为核心的大数据应用系统的开发能力, 以及结合新思想、运用新技术解决实际问题的态度和意识。

三、过程考核方式

本课程采取平时作业(课程报告)+与课程相关的科研实践+平时表现, 进行考核, 平时作业+平时表现占 70%, 包括出勤情况, 作业完成情况; 与课程相关的科研实践占 30%。

四、教学方法

利用多媒体教学, 采取案例教学, PPT 结合投屏演示, 线上线下混合式, 利用腾讯课堂和腾讯会议进行线上的演示和答疑。

五、各教学环节学时分配

授课内容(知识点)	第一课堂	第二课堂	实验	实践实训	合计
绪论	4				4
简单而有效的键值型内存数据库 Memcached	4		1		5
键值型数据库 Redis	10		1		11
处理非结构化数据的 Mongodb	14		2		16
列式数据库 Cassandra	12		2		14
面向图的数据库 Neo4J	4		2		6

数据科学与大数据技术专业教学大纲

合计	48		8		56
----	----	--	---	--	----

六、参考书

- [1] Shashank Tiwari, 巨成. 深入 NoSQL. 北京: 人民邮电出版社, 2012
- [2] 佐佐木达也, 罗勇. NoSQL 数据库入门. 北京: 人民邮电出版社, 2012
- [3] 皮雄军编. NoSQL 数据库技术实战. 北京: 清华大学出版社, 2014

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《NoSQL 数据库原理》（实验）教学大纲

课程代码：176010 课程类别：专业基础 适用专业：数据科学与大数据技术
 课程属性：课内实验 课程总学时：56 总学分：3.5 实验学时：8 学分：0.5
 执笔人：常亮 审定人：史健婷 审批人：李福军

一、实验性质及教学目标

《NoSQL 数据库原理》（实验）是一门实践性很强的课程。在学习 NoSQL 数据库原理与应用的过程中，只有多阅读程序、多编写程序、多上机，才能真正掌握 NoSQL 数据库的方法和技巧。

所有实验都是与《NoSQL 数据库原理》课程内容相配套的，共分为四个部分：第一部分为 Memcached 和 Redis 键值型分布式数据库的综合实验；第二部分为 MongoDB 文档型分布式数据库的综合实验；第三部分为 Cassandra 列族分布式数据库的综合实验。第四部分为 Neo4J 知识图谱数据库的综合实验。

二、实验项目

序号	实验项目名称	学时	实验类型	实验要求
1	Memcached 和 Redis 的安装与部署	2	演示+验证	与课程报告相关（5 个数据库至少选做 1），实验内容不得雷同
2	Mongodb 的安装与部署	2	演示+验证	与课程报告相关（5 个数据库至少选做 1），实验内容不得雷同
3	Cassandra 的安装与部署	2	演示+验证	与课程报告相关（5 个数据库至少选做 1），实验内容不得雷同
4	Neo4J 的安装与部署	2	演示+验证	与课程报告相关（5 个数据库至少选做 1），实验内容不得雷同

三、实验概述

实验一：Memcached 和 Redis 的安装与部署

目的：实现 Memcached 和 Redis 的安装与部署。

原理概述：一致性 hash 算法描述，主从复制原理。

方法与手段：在 CentOS 环境下完成 Memcached 和 Redis 的安装和部署工作。

应得到的实验结果和数据：完成 Memcached 和 Redis 的安装和部署工作后，运行与测试正确后截图，做为实验结果放在课程报告里。

实验二：Mongodb 的安装与部署

目的：实现 Mongodb 的安装与部署。

原理概述：关系数据库情形下的一致性，MongoDB 主从复制原理。

方法与手段：在 CentOS 环境下，完成 Mongodb 的安装和部署工作，CentOS 环境下安装 Python，Pip 安装 pymongo 插件，python 编程测试。

应得到的实验结果和数据： Mongodb 的安装和部署工作，CentOS 环境下安装 Python 和 Pip；安装 pymongo 插件。上述工作完成后，运行与测试正确后截图，做为实验结果放在课程报告里。

实验三：Cassandra 的安装与部署

目的：实现 Cassandra 的安装与部署。

数据科学与大数据技术专业教学大纲

原理概述：数据模型的演进，Cassandra 集群机制。

方法与手段：在 CentOS 环境下，完成 Cassandra 的安装和 Cassandra 集群环境搭建。

应得到的实验结果和数据：Cassandra 的操作，验证 Cassandra 集群各节点状态，cqlsh 的启动方式。上述工作完成后，运行与测试正确后截图，做为实验结果放在课程报告里。

实验四：Neo4J 的安装与部署

目的：实现 Neo4J 的安装与部署。

原理概述：知识图谱与图数据库。

方法与手段：：在 CentOS 环境下，完成图数据库 Neo4J 安装。

应得到的实验结果和数据：完成图数据库 Neo4J 的配置后，运行与测试正确后截图，做为实验结果放在课程报告里。

四、主要仪器设备

1. OS: CentOS-6 或者 7 及以上版本
2. VMware Workstation Pro 12 及以上版本
3. 远程终端工具 Xshell 4 及以上版本
4. Java JDK1.8 及以上版本
5. 能够上网的 PC 机若干台

五、教学形式

利用多媒体教学，采取案例教学，PPT 结合投屏演示，线上线下混合式，利用腾讯课堂和腾讯会议进行线上的演示和答疑。实验可由课后或实验课堂完成。

六、考核方式与成绩评定方法

考核方式：实验演示操作+课程报告实验结果+答辩表现=实验内容考核的综合成绩

本实验课程不单独组织考核，而是将实验内容考核的综合成绩，列入到与课程相关的科研实践的成绩当中。实验演示操作+答辩表现成绩将也做为平时表现成绩的参考。

七、教材及主要参考资料

- [1] Shashank Tiwari, 巨成. 深入 NoSQL. 北京：人民邮电出版社, 2012
- [2] 佐佐木达也, 罗勇. NoSQL 数据库入门. 北京：人民邮电出版社, 2012
- [3] 皮雄军编. NoSQL 数据库技术实战. 北京：清华大学出版社, 2014

八、说明

在执行大纲过程中应结合专业的特点和学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《数据可视化技术》教学大纲

课程代码：176011 课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：32 总学分：2

执笔人：李彪 审定人：史健婷 审批人：李福军

一、课程性质及教学目标

《数据可视化技术》是数据科学与大数据技术专业的一门专业必修课程。该课程内容主要包括数据可视化基础理论、数据的统计图形展示、大数据可视化方法、数据可视化工具及其应用等内容。使学生掌握将数据转化为直观视觉信息的技能，进一步提升数据洞察力和分析能力，为大数据的处理与分析奠定坚实基础。

教学目标：了解数据及数据可视化的基本概念、一般原理和基本方法；针对不同类型的数据，能够选择适当的可视化方法表达和分析数据；培养常用数据可视化工具的应用能力；培养对数据可视化处理以解决相关复杂工程问题的创新能力和综合实践能力；能够在多学科背景下，完成数据可视化任务，培养学生的团队协作意识等。

二、教学内容及要求

本课程的教学涉及数据可视化的基本概念、一般原理、数据可视化方法和可视化常用工具及应用等内容。具体包括：数据可视化基础理论、数据的统计图形展示、大数据可视化方法、数据可视化工具及其应用等。

（一）数据可视化基础理论（4学时）

1. 了解数据可视化的分类与发展。
2. 理解数据可视化的基本概念。
3. 了解数据可视化方法的应用。
4. 认识数据可视化常用图形形式及数据可视化工具。
5. 了解数据可视化工具的选择和使用方法。

（二）数据的统计图形展示（6学时）

1. 掌握统计图形展示方法。
2. 掌握数据统计图形的绘制方法。
3. 了解数据统计图形的应用场景。
4. 能够对可视化结果解释分析。

（三）大数据可视化方法（10学时）

1. 了解大数据内涵、特征及大数据的类型。
2. 依据数据类型特征，能够使用恰当的可视化形式表现数据。
3. 掌握时间数据可视化、比例数据可视化、关系数据可视化、文本数据可视化及复杂数据可视化的方法。
4. 能够对可视化结果给予恰当的解释说明。

（四）数据可视化工具及应用（10学时）

1. 了解数据可视化常用工具及应用。
2. 掌握 SPSS 数据可视化常用方法。
3. 掌握 Matplotlib 数据可视化常用方法。

数据科学与大数据技术专业教学大纲

4. 能够阅读和理解可视化结果报告。

(五) 实战案例分析 (2 学时)

1. 能够针对实际数据案例, 选择恰当的工具实现数据可视化。

2. 能够对数据可视化结果给予合理的说明和解释, 得到适当的分析结论。

本课程重点内容: 数据可视化方法及应用; 常用统计图形及应用场景; 时间数据可视化、比例数据可视化、文本数据可视化方法及实现; 阅读和理解可视化结果。

本课程难点内容: 常用统计图形的应用场景; 时间数据可视化、文本数据可视化的方法; 阅读和理解可视化结果。

三、过程考核方式

课程的总成绩包括平时成绩 (40%) 和期末成绩 (60%)。平时成绩主要包括课堂互动表现、作业或测试等; 期末成绩主要以期末考试成绩体现。

四、教学方法

本课程是一门32学时的专业必修课, 以课堂教学为主。教学过程采用讲授法、讨论式教学法和案例教学相结合的方式完成教学任务。通过案例分析辅以课堂理论知识讲授的方式激发学生学习的积极性、培养学生综合运用所学知识探究新问题、发现规律的创新思维能力, 提升学生的团队协作意识。

五、各教学环节学时分配

授课内容 (知识点)	第一课堂	第二课堂	实验	实践实训	合计
数据可视化基础理论	4				4
数据的统计图形展示	6				6
大数据可视化方法	10				10
数据可视化工具及应用	10				10
实战案例分析	2				2
合计	32				32

六、参考书

[1] 姜枫等. 大数据可视化技术. 北京: 人民邮电出版社, 2019

[2] 黑马程序员. Python 数据分析与应用: 从数据获取到可视化. 北京: 中国铁道出版社, 2019

[3] 张文彤. SPSS 统计分析基础教程 (第 3 版). 北京: 高等教育出版社, 2018

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《机器学习基础》教学大纲

课程代码：176012 课程类别：专业基础 适用专业：数据科学与大数据技术
总学时：40/8 总学分：2.5
执笔人：高玉宇 审定人：史健婷 审批人：李福军

一、课程性质及教学目标

《机器学习基础》是数据科学、计算机科学与技术、人工智能等专业的专业必修课程，旨在介绍机器学习的基础理论、算法原理、实现技术及应用领域。

随着大数据时代的到来和计算能力的飞速提升，机器学习基础已成为推动科技进步、产业升级的重要力量，广泛应用于图像识别、语音识别、自然语言处理、推荐系统、金融风控、医疗健康等多个领域。本课程通过系统学习，使学生掌握机器学习的基础知识和核心技术，具备解决实际问题的能力，为后续的学习和研究打下坚实的基础。

二、教学内容及要求

（一）机器学习概述（4学时）

1. 机器学习的定义、发展历程与分类（监督学习、无监督学习、半监督学习、强化学习）。
2. 机器学习中的基本概念（数据集、特征、标签、模型、训练、测试）。
3. 机器学习流程与评估指标（准确率、召回率、F1分数、混淆矩阵、ROC曲线等）。

（二）线性回归（4学时）

1. 线性回归的基本概念。
2. 一元线性回归。
3. 多元线性回归。
4. 岭回归（Ridge Regression）和套索回归（Lasso Regression）等正则化方法。

（三）K-近邻算法（2学时）

1. K-近邻算法的基本思想。
2. 距离度量。
3. K值的选择。
4. K-近邻分类与回归。

（四）决策树模型（4学时）

1. 决策树的基本概念与原理（信息增益、基尼指数）。
2. ID3、C4.5、CART等决策树算法。
3. 决策树的构建与剪枝策略。
4. 决策树的应用案例。

（五）贝叶斯网络（4学时）

1. 贝叶斯定理与贝叶斯网络基础。
2. 贝叶斯网络的构建与推理（前向推理、后向推理、混合推理）。
3. 贝叶斯网络的学习（参数学习与结构学习）。
4. 贝叶斯网络的应用案例。

（六）支持向量机（4学时）

1. 支持向量机的基本思想。

2. 线性可分 SVM 与线性不可分 SVM（软间隔、核技巧）。

3. SVM 的求解算法（SMO 等）。

4. SVM 的参数选择与模型评估。

（七）集成学习（4 学时）

1. Bagging 及随机森林。

2. Boosting 及 AdaBoost。

3. 残差提升树。

4. GBDT 和 XGBoost。

（八）聚类分析（2 学时）

1. 聚类分析的基本概念与分类（K-means、层次聚类、DBSCAN 等）。

2. 聚类算法的原理与实现。

3. 聚类效果的评估方法（轮廓系数、Calinski-Harabasz 指数）。

4. 聚类分析的应用场景。

（九）神经网络（4 学时）

1. 神经网络基础（神经元模型、激活函数）。

2. 前馈神经网络与反向传播算法。

3. 神经网络的设计与训练技巧。

4. 神经网络的应用案例。

本课程重点内容：理解机器学习的基本概念与分类；掌握机器学习流程与评估方法；理解线性回归的基本思想；掌握一元和多元线性回归的模型构建方法；理解正则化技术的必要性和实现方式；理解 K-近邻算法的基本思想和工作原理，掌握距离度量的方法；掌握决策树的基本原理与构建方法，理解剪枝策略的重要性；理解贝叶斯网络的基本概念与推理机制，掌握其构建与推理方法；掌握 SVM 的基本原理与求解方法，理解核技巧在非线性问题中的应用；理解 Bagging 及随机森林的基本原理；掌握 Boosting 及 AdaBoost 算法的实现；理解聚类分析的基本原理，掌握几种常见聚类算法的实现；理解神经网络的基本原理，掌握前馈神经网络的构建与训练过程。

本课程难点内容：如何根据实际问题选择合适的机器学习方法和评估指标；多重共线性问题；如何根据数据集的特点和实际需求选择合适的 K 值，以达到最佳的分类或回归效果；如何根据数据特征选择合适的决策树算法及优化参数；贝叶斯网络的结构学习与复杂网络中的高效推理算法；核函数的选择与 SVM 参数调优；理解 Bagging 及随机森林的基本原理；如何根据数据特性选择合适的聚类算法及参数调优；反向传播算法的理解与实现，神经网络结构设计与参数调优。

三、过程考核方式

本课程最终成绩按百分制评定。成绩包括平时成绩和期末报告（期末考试）成绩，平时成绩占 40%，期末报告（期末考试）成绩占 60%，平时成绩由作业、测验等组成，可根据情况适当修改考核方式。

四、教学方法

本课程是一门共 40 学时的专业必修课，包括课堂教学 32 学时和实验 8 学时，课堂教学中采用案例式、启发式等多种教学方法，以学生为主体，以启发学生的思维为核心，调动学生的学习主动性和积极性；要求学生课外通过查阅文献等自学完成机器学习基础的实践内容，加深对理论教学的理解。

数据科学与大数据技术专业教学大纲

利用在线教学和利用在线教学与测试平台进行课后训练与实践水平锻炼，利用QQ与微信的在线答疑等。实验教学另见《机器学习基础》实验教学大纲。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
机器学习概述	4				4
线性回归	4				4
K-近邻算法	2		2		4
决策树模型	4		2		6
贝叶斯网络	4		2		6
支持向量机	4				4
集成学习	4				4
聚类分析	2		2		4
神经网络	4				4
合计	32		8		40

六、参考书

- [1] 赵卫东等. 机器学习. 北京：人民邮电出版社, 2018
- [2] 艾旭升等. 机器学习技术. 北京：电子工业出版社, 2020
- [3] 周志华. 机器学习. 北京：清华大学出版社, 2016

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《机器学习基础》（实验）教学大纲

课程代码：176012 课程类别：专业基础 适用专业：数据科学与大数据技术
课程属性：课内实验 课程总学时：40 总学分：2.5 实验学时：8 学分：0.5
执笔人：高玉宇 审定人：史健婷 审批人：姜福军

一、实验性质及教学目标

《机器学习基础》实验教学大纲旨在通过实践性强的课程实验，使学生掌握机器学习核心算法的原理与应用。实验性质强调动手操作与理论验证相结合，通过解决实际问题深化理解。教学目标在于培养学生数据预处理、模型构建、参数调优及结果评估等能力，同时激发创新思维，为后续大数据分析、专业实训和毕业设计的学习与研究奠定坚实基础。

二、实验项目

序号	实验名称	学时	实验类型	实验要求
1	K-近邻算法及应用	2	设计性	必做
2	朴素贝叶斯算法及应用	2	设计性	必做
3	决策树算法及应用	2	设计性	必做
4	K-means 算法及应用	2	设计性	必做

三、实验概述

实验一：K-近邻算法及应用

实验目的：

1. 熟悉 K-近邻的原理；
2. 掌握线性回归的实现；
3. 熟悉 matplotlib 的使用；
4. 熟悉 sklearn 的使用；
5. 应用 sklearn 中的 K-近邻分类函数对鸢尾花数据集进行分类和测试，计算在测试集上的准确率、召回率和精准率，绘制混淆矩阵。

原理概述：K-近邻（K-Nearest Neighbors, KNN）算法实验旨在通过实例理解监督学习中的基本分类与回归方法。KNN 基于实例学习，核心思想是在特征空间中，找出与待分类点距离最近的 K 个训练样本，根据这 K 个“邻居”的类别信息，通过多数投票（分类）或平均（回归）来决定待分类点的类别或预测值。实验将涵盖数据预处理、距离度量选择、K 值优化等关键步骤，直观展示 KNN 算法的原理及应用效果。

实验二：朴素贝叶斯算法及应用

实验目的：

1. 熟悉朴素贝叶斯分类的原理；
2. 掌握朴素贝叶斯的实现；
3. 熟悉 matplotlib 的使用；
4. 熟悉 sklearn 的使用；
5. 应用 sklearn 中的朴素贝叶斯分类函数对垃圾短信数据集进行分类和测试，计算在测试集上的准确率、召回率和精准率，绘制混淆矩阵。

数据科学与大数据技术专业教学大纲

原理概述：朴素贝叶斯算法实验旨在教授一种基于概率统计的分类方法。该算法假设特征之间相互独立，利用贝叶斯定理计算各类的后验概率，并选择概率最大的类作为预测结果。实验将涵盖数据处理、特征选择、先验概率计算等关键步骤，通过实例演示朴素贝叶斯算法在文本分类、垃圾邮件识别等领域的应用，加深学生对概率统计在机器学习中的理解。

实验三：决策树算法及应用

实验目的：

1. 熟悉决策树的原理；
2. 掌握决策树的实现；
3. 熟悉 sklearn 的分类和回归使用；
4. 应用 sklearn 中的决策树回归函数对波士顿房价预测数据集进行学习和测试，并计算测试集上的平均绝对误差。

原理概述：决策树算法实验旨在通过构建决策树模型，教授学生一种直观易懂的分类与回归方法。实验将展示如何根据数据特征选择最优划分属性，递归构建决策树，并学习剪枝技术防止过拟合。通过实例操作，学生将理解决策树的工作原理，及其在分类、预测任务中的广泛应用，如信用评估、医疗诊断等。

实验四：K-means 算法及应用

实验目的：

1. 熟悉 K-means 的原理；
2. 掌握 K-means 的实现；
3. 应用 sklearn 中的 K-means 分类函数对鸢尾花数据集进行分类和测试，与 K-近邻算法在测试集上进行准确率、召回率和精准率的比较，绘制混淆矩阵。

原理概述：K-means 算法实验旨在教授学生一种常用的聚类分析方法。该算法通过迭代方式，将数据划分为 K 个簇，使得同一簇内数据相似度高，不同簇间相似度低。实验将涵盖初始聚类中心选择、簇内样本分配、聚类中心更新等关键步骤，通过实践让学生理解 K-means 的聚类原理及其在数据挖掘、图像处理等领域的应用。

四、主要仪器设备

1. 操作系统：

- (1) 类型：Windows、Linux 或 macOS 等，视学生熟悉程度和实验需求而定。
- (2) 要求：稳定、支持常用的机器学习库和工具。

2. 编程语言及环境：

- (1) 主要语言：Python（因其丰富的机器学习库和社区支持），Java、R 等也可作为补充。
- (2) 开发环境：集成开发环境（IDE）如 PyCharm、Visual Studio Code、Jupyter Notebook 等，便于代码编写、调试和结果展示。

3. 机器学习库和工具：

- (1) 核心库：scikit-learn（Python，提供大量机器学习算法实现）、TensorFlow 或 PyTorch（用于深度学习）。
- (2) 数据处理和可视化工具：pandas（数据清洗和分析）、NumPy（数值计算）、Matplotlib（绘图）、Seaborn（高级绘图）等。

五、教学形式

在每次实验前，通过课堂讲授或线上视频的形式，介绍本次实验所涉及的机器学习算法的基本原理、关键步骤和注意事项。帮助学生建立对实验内容的理论框架，为后续的实践操作提供理论指导。同时进行辅导答疑，最后进行实验总结。

六、考核方式与成绩评定方法

每次完成实验内容，进行总结分析，写出实验报告。按照学生每次的预习情况、实验态度、实验内容完成情况及实验报告书写情况等方面进行考核，实验总成绩占期末成绩的10%。具体分配如下：

1. 平时成绩（包括学生的实验态度、实验完成情况等）：占实验总成绩的50%；
2. 实验报告：占实验总成绩的50%。

七、教材及主要参考资料

[1] 吕云翔等. 机器学习原理及应用. 北京：机械工业出版社, 2020

八、说明

应根据实际教学学时的要求，根据实际实验条件，结合专业特点和学生的实际情况安排实验。按照本大纲的总体框架和思路，制定出符合专业特色的教学计划和教学进度安排，有计划地安排实验的授课顺序，以符合学生接受的需求。

《Spark 大数据处理技术》教学大纲

课程代码：176013 课程类别：专业基础 适用专业： 数据科学与大数据技术
总学时/实验： 48/8 总学分： 3
执笔人：史健婷 审定人：丁文妙 审批人：李福军

一、课程性质及教学目标

《Spark 大数据处理技术》是数据科学与大数据技术专业的一门专业必修课。

本课程主要介绍 Spark 的原理和架构，主要以 Scala 作为开发 Spark 应用程序的编程语言，系统讲解 Spark 编程的基础知识、设计与运行原理、Spark 环境搭建和使用方法、RDD 编程、Spark SQL、Spark Streaming、Structured Streaming、Spark MLlib 等，并结合实际案例，进行分析。通过理论讲授+实践操作，让同学们掌握 Spark 大数据处理技术的应用。

二、教学内容及要求

（一）大数据处理与分析技术概述（2 学时）

1. 大数据的技术背景，计算机软硬件的发展历程，IT 基础设施发展的历程。
2. 大数据的基本概念，大数据处理与分析关键技术的基础知，大数据技术的不同层面及其功能。
3. 大数据的几种经典计算模式。

（二）编程语言基础（8 学时）

1. 编程语言概述。
2. 选择与循环，分支结构，迭代器。
3. 函数：参数，return 语句，变量作用域。

（三）Spark 基础（8 学时）

1. Spark 基本概念、历史，与 Hadoop 的对比。
2. Spark 生态系统，Spark 运行架构，Spark 的部署方式和安装方式。
4. 在 spark-shell 中运行代码，编写 Spark 独立应用程序。

（四）Spark 编程模型（8 学时）

1. MapReduce 基本模型概念。
2. 编程常用函数，RDD 编程基础，键值对 RDD，数据读写。
3. Spark SQL 简介，DataFrame 概述，DataFrame 的创建。

（五）Spark Streaming（8 学时）

1. 流计算简介，Spark Streaming，DStream 操作概述。
2. 基本输入源，高级数据源，转换操作，输出操作。
3. Structured Streaming 简介。
4. 机器学习工作流。

（六）Spark MLlib（6 学时）

1. Spark MLlib 简介。
2. 特征抽取、转化和选择、聚类算法、机器学习参数调优。

本课程重点内容：大数据处理与分析技术概述、编程语言基础、Spark 基础、特征抽取、DStream 操作、DataFrame 的创建。

本课程难点内容：Spark 编程模型、Spark Streaming、Spark MLlib、机器学习参数调优。

数据科学与大数据技术专业教学大纲

三、过程考核方式

课程采取学习表现与结课报告相结合的方式进行，学习表现占 40%，包括出勤情况，作业完成情况，课堂表现等，结课报告占 60%。

四、教学方法与手段

课堂教学中采用案例式、讨论等多种教学方法，教学过程中要突出教学互动，发挥学生的主观能动性，以“体验式”教学进行讲授、交流、讨论、实践，以项目驱动的方式引入问题，解决问题来带入知识点，将知识点应用于实际编程，实现将理论融入实践，使学生理解知识点的同时，能遇到问题-分析问题-解决问题，鼓励学生自学和实践、主动获取知识，激发学生的学习主动性和兴趣。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
大数据处理与分析技术概述	2				2
编程语言基础	8		2		10
Spark 基础	8				8
Spark 编程模型	8		2		10
Spark Streaming	8		2		10
Spark MLlib	6		2		8
合计	40		8		48

六、参考书

- [1] 林子雨, 郑海山, 赖永炫. Spark 编程基础 (Python 版). 北京: 人民邮电出版社, 2020
- [2] Holden Karau 等. Spark 快速大数据分析. 北京: 人民邮电出版社, 2018
- [3] 许利杰, 方亚芬. 大数据处理框架 Apache Spark 设计与实现. 北京电子工业出版社, 2020

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《Spark 大数据处理技术》（实验）教学大纲

课程代码：176013 课程类别：专业基础 适用专业：数据科学与大数据技术 课程属性：课内实验

总学时：48 总学分：3 实验学时：8 学分：0.5

执笔人：史健婷 审定人：丁文妙 审批人：李福军

一、实验性质及教学目标

《Spark 大数据处理技术》实验教学大纲旨在通过课程实验项目，使学生掌握 Spark 平台进行大数据处理与分析的基本技能。实验性质强调动手操作与理论验证相结合，通过解决实际问题深化理论学习。教学目标在于培养学生对大规模结构化和非结构化数据的基本处理与分析，进而建构大数据处理与分析的基础知识体系，为后续专业实训和毕业设计的学习与研究奠定坚实基础。

二、实验项目

序号	实验名称	学时	实验类型	实验要求
1	Scala 编程基础	2	设计性	必做
2	Spark 环境搭建和 RDD 编程	2	设计性	必做
3	Spark SQL 和 Spark Streaming	2	设计性	必做
4	Spark MLlib	2	设计性	必做

三、实验概述

实验一：Scala 编程基础

实验目的：

1. 掌握 Scala 语言的基础语法；
2. 理解函数式编程思想；
3. 熟悉面向对象编程特性的使用；
4. 掌握集合库的操作。

原理概述：Scala 编程基础实验旨在帮助学生全面掌握 Scala 语言的基础知识和编程技能，深入理解函数式编程和面向对象编程的思想，通过循环、函数定义和调用、类、继承、特质等实验操作，提高学生将 Scala 应用于实际项目开发的能力。

实验二：Spark 环境搭建和 RDD 编程

实验目的：

1. 掌握 Spark 环境的搭建方法；
2. 掌握在 Spark 平台提交和运行代码的方法；
3. 掌握 RDD 键值对的创建方法；
4. 掌握 RDD 的转换操作和行动操作。

原理概述：Spark 环境搭建主要涉及 Spark 集群的部署和配置。Spark 支持多种部署模式，包括单机模式、集群模式（如 Standalone 模式、YARN 模式、Mesos 模式等）。在部署过程中，需要配置 Spark 的核心组件（如 Driver、Executor 等）、内存和 CPU 资源分配、网络通信设置等。此外，还需要配置与 Hadoop 等大数据技术的集成，以确保 Spark 能够访问和处理存储在 Hadoop HDFS 等系统中的数据。

RDD 是 Spark 中的核心抽象，代表了一个不可变的、可分区并行计算的分布式数据集。RDD 的设

计基于两个关键原则：一是数据容错性，通过数据依赖关系（Lineage）实现数据重建；二是数据局部性，通过尽可能在数据所在节点上执行计算来减少数据移动。RDD 的编程模型基于转换（Transformation）和行动（Action）两种操作。转换操作是惰性的，不会立即执行计算，而是构建了一个计算图（DAG），描述了 RDD 之间的依赖关系。行动操作则触发计算图的执行，将结果输出到外部系统或变量中。这种设计使得 RDD 能够高效地处理大规模数据集，并支持复杂的迭代计算。通过对 RDD 编程的实验项目，帮助学生在操作时有效提高大数据分析的效率。

实验三：Spark SQL 和 Spark Streaming

实验目的：

1. 掌握通过 Spark SQL 进行数据查询、数据转换和数据聚合等操作；
2. 掌握 Spark SQL 与 MySQL 的集成使用；
3. 理解 DStream 的概念及其与 RDD 的关系；
4. 掌握从多种数据源（如 Kafka、Flume、TCP 套接字等）接入实时数据流的方法。

原理概述：Spark SQL 是 Spark 用来处理结构化数据的一个模块，它提供了一个编程抽象叫做 DataFrame，并且作为分布式 SQL 查询引擎的作用。教授学生利用 Spark SQL 将 SQL 查询转换为一系列 RDD 操作来实现数据的处理，并通过优化器来提高执行效率。Spark SQL 支持多种数据源，包括 JDBC、JSON 等，学生可以方便地从不同来源读取数据。

Spark Streaming 是对核心 Spark API 的一个扩展，教授学生掌握其数据处理流程和方法，可以帮助学生实现对实时数据流的流式处理，并具有很好的可扩展性、高吞吐量和容错性。

实验四：Spark MLlib

实验目的：

1. 掌握 Spark MLlib 提供的 API 使用方法；
2. 利用逻辑回归等机器学习方法实现分类问题；

原理概述：Spark MLlib 是基于 Apache Spark 的机器学习库，旨在提供可扩展的机器学习算法和工具。Spark MLlib 建立在 Spark 的分布式计算框架之上，充分利用了 Spark 的内存计算能力和高效的通信机制。教授学生通过 MLlib 库调用多种常用学习方法（如逻辑回归、聚类等），帮助学生在大数据环境下进行高效的机器学习任务提供了有力的支持。

四、主要仪器设备

1. 操作系统：Windows 或 Linux 等，视学生熟悉程度和实验需求而定。
2. 编程语言及环境：
 - （1）主要语言：Java 或 Python 等。
 - （2）开发环境：头歌实践教学平台或集成开发环境（IDE）如 PyCharm 等，便于代码编写、调试和结果展示。

五、教学形式

在实验课上，通过课堂讲授的形式，介绍本次实验所涉及的 Spark 技术的基本原理、关键步骤和注意事项。帮助学生建立对实验内容的理论框架，为后续的实践操作提供理论指导。同时进行辅导答疑，最后进行实验总结。

六、考核方式与成绩评定方法

每次完成实验内容，进行总结分析，写出实验报告。按照学生每次的实验态度、实验内容完成

情况及实验报告书写情况等方面进行考核，具体分配如下：实验态度 20%，实验内容完成情况 40%，实验报告书写情况 40%。实验总成绩占期末成绩的 20%。

七、教材及主要参考资料

[1] 林子雨, 赖永炫, 陶继平. Spark 编程基础 (Scala 版). 北京: 人民邮电出版社, 2018

八、说明

应根据实际教学学时的要求和实际实验条件，结合专业特点和学生的实际情况安排实验。根据学生的接受程度，有计划地安排实验的授课顺序和授课内容。实验成绩占期末成绩的比例可由任课教师根据实际情况调整。

《专业外语》教学大纲

课程代码：176014

课程类别：专业基础

适用专业：数据科学与大数据技术

总学时：32

总学分：2

执笔人：史健婷

审定人：丁文抄

审批人：李福军

一、课程性质及教学目标

本课程是数据科学与大数据技术专业的一门专业基础课程。

通过该课程的学习，学生可掌握一定数量的计算机专业英语词汇和大数据专业英语词汇，使学生能够阅读、理解和翻译计算机使用过程中的英文提示信息 and 大数据专业的英文资料，从而加深学生对计算机相关课程和大数据知识的理解，并为学生在国际学术环境下跨文化交流打下坚实的基础。

二、教学内容及要求

(一) 计算机屏幕英语和计算机产品说明书 (6 学时)

1. 计算机专业英语知识。
2. 屏幕英文提示信息和产品说明书。

(二) 计算机专业基础知识 (10 学时)

1. 计算机专业基础知识内容的专业英语词汇。
2. 计算机硬件系统，操作系统，数据结构，数据库，多媒体和计算机网络的阅读分析。

(三) 大数据专业基础知识 (8 学时)

1. 大数据专业英语词汇。
2. 大数据技术的阅读分析。

(四) 英语科技文章的阅读和写作 (8 学时)

1. 大数据专业英语科技文章的阅读技巧。
2. 英文论文检索，英语科技文章的阅读，英语应用文写作。
3. 大数据开发文档写作。

本课程重点内容：计算机屏幕英语、数据结构、计算机网络等阅读和翻译、大数据基础知识阅读理解。

本课程难点内容：系统指令解释及注释、数据分析方法和技术相关的缩略词和应用介绍、大数据开发文档说明书、数据仓库与数据挖掘常用词汇、英语应用文写作。

三、过程考核方式

本课程的考核方式采取平时学习表现与期末报告相结合的方式，学习表现占 50%，主要包括课堂表现（课上提问、阅读和翻译）占 15%、小组汇报展示占 15%、作业完成情况占 20%；期末报告占 50%，报告包括翻译专业领域文献、写英文摘要等。

四、教学方法

针对本科生授课的要求和专业外语广泛应用的特点，主要运用讲授法、讨论法、启发式、实例式和翻转课堂等教学法。

教学中以多媒体教学为主，辅以板书。

五、各教学环节学时分配

数据科学与大数据技术专业教学大纲

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
计算机屏幕英语和计算机产品说明书	6				6
计算机专业基础知识	10				6
大数据专业基础知识	8				10
英语科技文章的阅读和写作	8				10
合计	32				32

六、参考书

- [1] 王莹等编. 计算机专业英语. 西安: 西安电子科技大学出版社, 2018
- [2] 郭文芳等编. 计算机专业英语. 北京: 电子工业出版社, 2017
- [3] 王小刚. 计算机专业英语 (第4版). 北京: 机械工业出版社, 2015

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《R 语言统计分析》教学大纲

课程代码：176015 课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：40 总学分：2.5

执笔人：陈学刚 审定人：史健婷 审批人：李海华

一、课程性质及教学目标

《R 语言统计分析》为数据科学与大数据技术专业的专业基础课。该课程以《数理统计》、《大数据导论》为基础，主要学习 R 语言的原理和功能。R 语言具有承上启下的作用，学生可在学习 R 语言的同时，融合已学习过的大数据相关技术，对复杂数据进行获取和统计，利用统计结果从科学的角度理解数据。同时，培养学生用数据分析和可视化技术实现数据的可视化，解决不同难度的数据分析问题。为进一步学习数据仓库与挖掘、大数据编程等专业课程奠定良好的基础。

本课程的教学目标是通过 R 语言的语法规则、数据结构、数据清洗、数据变换、数据整合、数据可视化的学习，较好地训练学生利用计算机对数据进行分析 and 展现，使学生掌握数据管理和数据分析的技术，为培养学生开发大数据分析系统能力打下良好的基础。具体目标可分解如下：

(1) 能够应用各种数据源导入数据、整理数据及管理数据。

(2) 能够利用 R 语言实现常见的柱状图、饼图等单变量和多变量的图形绘制方法，并将图形保存为各种格式的文件。

(3) 能够利用稳定的、先进的数据分析方法，实现对比较复杂数据的处理。

二、教学内容及要求

(一) R语言介绍 (2学时)

1. R 语言的诸多特性，R 语言开发环境的获取方法。
2. 如何用网上的拓展包增强 R 的基本功能，用户界面，交互和批处理方式运行程序。

(二) 创建数据集 (4学时)

1. 数据集的概念，数据结构。
2. 使用键盘及第三方数据处理平台实现数据的输入。
3. 数据集的标注和处理数据对象的实用函数等。

(三) 图形的初阶 (6学时)

1. 使用图形和符号、线条、颜色、文本属性等图形参数。
2. 添加文本、自定义文本和图例。
3. 文本标注和图形组合。

(四) 基本数据管理 (6学时)

1. 创建新变量，变量的重编码和重命名。
2. 缺失值、日期值及类型转换。
3. 数据排序、数据集的合并。
4. 数据集取子集。
5. 使用 SQL 语句操作数据框等。

(五) 高级数据管理 (6学时)

1. 数据处理难题的解决办法，数值和字符处理函数。
2. 控制流和用户自编函数，整合和重构等。

数据科学与大数据技术专业教学大纲

（六）基本图形（8学时）

1. 基本图形绘制知识。
2. 条形图、饼图、直方图、核密度图、箱线图和点图等绘制。

（七）基本统计分析（8学时）

1. 描述性统计分析，频数表和列联表、相关、t 检验。
2. 组件差异的非参数检验以及组件差异的可视化等。

本课程重点内容：R 语言拓展包的使用；R 中的数据结构；设置图形参数；创建新变量、变量的重编码和重命名；缺失值、日期值及类型转换；数据排序、数据集的合并；数据集取子集；数值和字符处理函数；数据的整合；条形图、直方图、点图、箱线图的使用；频数表和列联表、相关性检验。

本课程难点：图形参数设定；变量重编码；缺失值识别和处理；数值处理函数；数据的整合；箱线图；组件差异的可视化。

三、过程考核方式

课程最终成绩按百分制评定。成绩包括平时成绩和期末考试成绩，平时成绩占 40%，期末考试成绩占 60%，平时成绩由课堂表现（10%）、作业（20%）、课堂实验组成（10%）。

四、教学方法

课程采用多媒体方式进行教学，其中实践教学采用案例式、项目式等方式进行，在教学中使用 R 语言软件进行相关理论的分析与实践。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
R 语言简介	2				2
创建数据集	4				4
图形初阶	6				6
基本数据管理	6				6
高级数据管理	6				6
基本图形	8				8
基本统计分析	8				8
合计	40				40

六、参考书

- [1] 王翔等. R 语言数据可视化与统计分析基础. 北京：机械工业出版社, 2019
- [2] 祖尔等. R 语言数据操作. 西安：西安交通大学出版社, 2011
- [3] 汤银才. R 语言与统计分析. 北京：高等教育出版社, 2008
- [4] 刘鹏等. R 语言. 北京：清华大学出版社, 2020
- [5] 薛毅等. 统计建模与 R 语言. 北京：清华大学出版社, 2007

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《复变函数》教学大纲

课程代码：175011 课程类别：专业基础 适用专业：数据科学与大数据技术
总学时：40 总学分：2.5
执笔人：张兴华 审定人：史健婷 审批人：李福军

一、课程性质及教学目标

本课程是数据科学与大数据技术专业大学本科生选修的专业基础课程。是理工科相关专业的一门重要的数学类基础课程。

复变函数论的基本理论由柯西积分理论、魏尔斯特拉斯级数理论、黎曼的几何理论三大部分组成。课程的教学目标是使学生掌握复变函数的基本理论和方法，获得独立地分析和解决某些相关理论和实际问题的能力，为有关后续课程的学习和进一步扩大数学知识面而奠定必要的数学基础。包括复数运算、解析函数、初等函数、复变函数的积分理论、级数展开及留数理论等内容。在传授基础理论和基本技能的同时，注重工程背景和实际应用，加强学生在应用方面的培养。扩展学生的学习思路，使他们了解更多的和现代生活息息相关的数学应用知识。

二、教学内容及要求

（一）复数与复变函数（4学时）

1. 复数及其运算、几何表示。
2. 复平面上的点的类型划分、点集、区域、曲线，区域的连通性。
3. 复变函数、极限、连续。
4. 复球面与无穷远点。

（二）解析函数（8学时）

1. 解析函数的概念、柯西-黎曼条件、函数可微与解析的充要条件。
2. 初等单值解析函数、幂函数、指数函数、三角函数。
3. 初等多值解析函数、根式函数、对数函数以及一般幂函数与一般指数函数。

（三）复变函数积分（8学时）

1. 复积分的概念及其性质。
2. 柯西积分定理及其推广（单连通，复连通），复合闭路定理以及闭路变形定理，原函数概念。
3. 柯西积分公式及其推论，解析函数的无穷可微性公式以及应用，刘维尔定理、摩勒拉定理、代数学基本定理。
4. 调和函数概念，解析函数与调和函数的关系。

（四）解析函数的幂级数表示法（6学时）

1. 复级数的基本性质（复数项级数和函数项级数的性质）。
2. 幂级数：Abel 定理，和函数的解析性，Taylor 展开式，解析函数的级数展开举例。牢记一些初等函数的泰勒展式和推导。
3. 解析函数零点的孤立性，解析函数的唯一性定理，最大模原理。

（五）解析函数的罗朗展式与孤立奇点（6学时）

1. 解析函数的罗朗展式（圆环域和孤立奇点的邻域内），罗朗级数与泰勒级数之间的关系。
2. 解析函数的孤立奇点，可去奇点、极点、本性奇点的划分方法，三类孤立奇点的判别条件。
3. 解析函数在无穷远点邻域的性质。

数据科学与大数据技术专业教学大纲

4. 整函数与亚纯函数概念、简单性质。

(六) 留数理论及其应用 (8 学时)

1. 留数的概念, 留数定理, 无穷远点的留数。

2. 用留数定理计算积分。

本课程重点内容: 复变函数极限、连续的定义和性质, 基本初等解析函数, 柯西积分定理、柯西积分公式、解析函数无穷可微性公式, 复数项级数和复函数项级数的收敛性判别, 留数定义和计算。

本课程难点内容: 初等多值解析函数、复积分的计算、解析函数和调和函数的关系、泰勒级数和罗朗级数、零点和孤立奇点、留数理论及应用。

三、过程考核方式

课程一般采用平时表现与期末考试相结合的方式进行考核。平时表现占 30%, 包括课堂表现 15%, 作业完成情况 15%; 期末考试 70%。也可以只采用笔试考核的方式。

四、教学方法与手段

教学方法以课堂的讲授为主, 课前预习、教师课后辅导、答疑为辅助手段。鼓励学生充分利用现代教学技术, 到网上收集, 查找本学科的相关资料, 巩固学习成果, 开阔视野。

五、各教学环节学时分配

授课内容 (知识点)	第一课堂	第二课堂	实验	实践实训	合计
复数与复变函数	4				4
解析函数	8				8
复变函数的积分	8				8
解析函数的幂级数表示法	6				6
解析函数的洛朗展式与孤立奇点	6				6
留数理论及其应用	8				8
合计	40				40

六、参考书

[1] 方企勤. 复变函数. 北京: 北京大学出版社, 2017

[2] 李红. 复变函数与积分变换 (第 5 版). 北京: 高等教育出版社, 2018

[3] 庞学诚. 复变函数学习指导书. 北京: 科学出版社, 2019

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《数据分析》教学大纲

课程代码： 175035 课程类别： 专业基础 适用专业： 数据科学与大数据技术

总学时： 32 总学分： 2

执笔人： 李彪 审定人： 史健婷 审批人： 李福军

一、课程性质及教学目标

《数据分析》是数据科学与大数据技术专业的一门重要的专业课程。本课程的学习为学生学习大数据相关的后继课程打下良好的理论基础，也为与数据分析相关方向的高质量就业积累知识储备。

教学目标：了解数据分析的基本原理、基本概念，掌握数据分析的基本方法和基本步骤；培养学生的抽象思维能力、逻辑推理能力和自学能力，培养学生综合运用所学知识分析和解决实际问题的意识和能力；通过引入生活相关的数据案例，培养学生的主人翁意识和责任意识，使学生充分认识中国国情和自己应履行的社会责任，同时培养学生的团队协作意识和实践创新能力。

二、教学内容及要求

本课程主要围绕数据分析问题，系统研究数据搜集、整理，辅以科学、合理的分析，并依据数据做出行业研究、评估和预测技术等。课程的教学内容较全面地涵盖了数据文件的建立、数据管理及预处理、描述性统计分析、假设检验、相关分析、方差分析、线性回归分析、聚类分析等。主讲教师可以根据学情分析、技术发展情况和授课实际等对教学内容进行适当删减或补充。

（一）数据统计分析概述及软件简介（2学时）

1. 了解数据科学的概念和内涵。
2. 了解数据统计分析基本理论、基本分析方法。
3. 掌握数据分析流程。
4. 了解数据统计分析软件的发展、软件安装与启动方法。
5. 掌握数据统计分析软件的工作界面、主要菜单与功能说明。

（二）数据文件的建立、管理及数据预处理（4学时）

1. 掌握 SPSS 数据分析的基本操作。
2. 了解变量视图和数据视图。
3. 了解数据的读取、保存方法。
4. 掌握数据文件的管理、数据预处理方法及其实现。

（三）描述性统计分析（4学时）

1. 掌握数据描述的数值方法和数据分布的分析方法。
2. 了解数据的图形化描述方法。
3. 掌握描述性统计分析的实现方法。
4. 了解常用统计图形的绘制方法。
5. 能够对描述性统计分析结果给与恰当的解释说明。

（四）参数检验（4学时）

1. 了解参数检验的概念和基本原理。
2. 了解 T 检验应用的前提条件。
3. 掌握 T 检验方法的基本原理及操作方法。
4. 了解非参数检验的基本原理。

5. 能够阅读和理解分析结果。

（五）方差分析（4 学时）

1. 了解方差分析的基本思想。
2. 掌握单因素方差分析的原理和操作方法。
3. 了解多因素方差分析的原理和操作方法。
4. 能够阅读和理解分析结果。

（六）相关分析（4 学时）

1. 了解相关分析的概念和基本原理。
2. 掌握双变量相关分析方法。
3. 掌握相关系数的计算方法及其显著性检验。
4. 了解偏相关分析与距离分析的概念及实现方法。
5. 能够阅读和理解分析结果。

（七）线性回归分析（6 学时）

1. 了解线性回归分析的概念和基本原理。
2. 掌握一元线性回归分析方法及其实现。
3. 掌握多元线性回归分析方法。
4. 了解多元线性回归分析模型诊断的方法。
5. 能够阅读和理解分析结果。

（八）聚类分析（2 学时）

1. 了解聚类分析的概念和聚类分析的一般步骤。
2. 掌握常见的聚类统计量。
3. 掌握常见聚类分析方法及其应用。

（九）实战案例分析（2 学时）

1. 了解案例分析的基本步骤。
2. 了解分析目标的确定原则。
3. 能够选择恰当的分析方法完成案例分析任务。
4. 能够阅读和理解分析结果，并形成适当的分析结论。

本课程重点内容：数据统计分析的基本步骤；数据文件的管理与预处理方法及实现；描述性统计分析方法及实现；相关分析、方差分析、线性回归分析等的实现；能够针对案例数据的背景，选择恰当的方法进行数据分析；能够对分析的结果给予恰当的说明和解释，并形成适当的分析结论等。

本课程难点内容：应用统计分析软件实现描述性统计分析；相关分析、方差分析、线性回归分析等方法的实现及应用；对分析结果给予恰当的说明和解释，并形成适当的分析结论等。

三、过程考核方式

课程的总成绩包括平时成绩（40%）和期末成绩（60%）。平时成绩主要包括课堂互动表现、作业或测试；期末成绩主要以期末考试成绩体现。

四、教学方法

本课程是一门32学时的选修课，以课堂教学为主。课程主要采用讲授法、案例式教学的方法完成基本理论、基本概念部分的教学；结合情境创设和讨论式教学方法，提升学生的学习积极性，促使学生理论与实践相结合，使学生能够主动探究问题、发现规律。

数据科学与大数据技术专业教学大纲

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
数据统计分析概述及软件简介	2				2
数据文件的建立、管理及数据预处理	4				4
描述性统计分析	4				4
参数检验	4				4
方差分析	4				4
相关分析	4				4
线性回归分析	6				6
聚类分析	2				2
实践案例分析	2				2
合计	32				32

六、参考书

- [1] 张文彤. SPSS 统计分析基础教程（第3版）. 北京：高等教育出版社, 2018
- [2] 王国才等. 数据分析基础——基于Excel和SPSS. 上海：上海交通大学出版社, 2018
- [3] 李洪成, 江宏华. SPSS数据分析教程. 北京：人民邮电出版社, 2012

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《数字图像处理》教学大纲

课程代码：172037 课程类别：专业基础 适用专业：数据科学与大数据技术
总学时：40 总学分：2.5
执笔人：高本宇 审定人：史健婷 审批人：李福军

一、课程性质及教学目标

《数字图像处理》是数据科学与大数据技术专业的一门重要专业选修课程，旨在培养学生掌握数字图像处理的基本理论、核心算法及其在实际应用中的技能。本课程融合了计算机科学、数学、等多个学科的知识，是连接数据科学与现实世界图像信息处理的桥梁。

通过本课程的学习，学生不仅能够深入理解图像数据的基本特性、处理方法和分析技术，还能将所学知识应用于图像处理、计算机视觉、医学影像分析、遥感图像处理、安全监控、艺术创作等多个领域，为解决复杂的数据科学问题提供有力支持。

二、教学内容及要求

（一）绪论（4学时）

1. 图像信号的表示；
2. 数字信号处理技术的主要内容；
3. 数字图像信号处理技术的主要特点；
4. 图像系统的线性模型；
5. 图像处理系统的构成；
6. 图像通信系统；
7. 人眼视觉特性；
8. 图像质量的评价方法。

（二）MATLAB图像处理工具箱（4学时）

1. 图像类型及类型转换；
2. 图像的读、写操作；
3. 图像的显示操作；
4. MATLAB工具箱的使用。

（三）图像变换（6学时）

1. 一维离散傅立叶变换；
2. 二维离散傅立叶变换；
3. 离散傅立叶变换的性质；
4. 离散K-L变换；
5. 一维离散余弦变换；
6. 二维离散余弦变换；
7. 变换的矩阵表达式；
8. 基本图像和基本频谱；
9. 离散沃尔什和哈达玛变换。

（四）图像增强（6学时）

1. 灰度变换法；

2. 直方图修正法;
3. 直方图规定化;
4. 图像的同态增晰;
5. 图像的平滑;
6. 图像的锐化;
7. 图像的伪彩色处理。

(五) 图像形态学 (6 学时)

1. 集合的定义;
2. 二值形态学的运算;
3. 膨胀和腐蚀;
4. 开运算和闭运算;
5. 击中与未击中;
6. 灰度图像的形态学处理。

(六) 图像分割 (6 学时)

1. 图像分割的定义与分类;
2. 基于边缘的图像分割;
3. 边缘检测;
4. 边缘跟踪;
5. 基于区域的分割;
6. 阈值分割法;
7. 区域生长法;
8. 分裂合并法;
9. 运动图像分割方法。

(七) 图像复原 (4 学时)

1. 连续图像退化的数学模型;
2. 几个典型的退化模型;
3. 离散图像退化的数学模型;
3. 循环矩阵的对角化;
4. 对角化在降质模型中的应用;
5. 逆滤波;
6. 无约束图像复原;
7. 匀速直线运动引起的图像模糊的复原。

(八) 图像编码与压缩技术 (4 学时)

1. 图像信息的冗余;
2. 信息熵与压缩比;
3. 统计编码;
4. 预测编码;
5. 变化编码;

数据科学与大数据技术专业教学大纲

6. 静止图像压缩编码标准等；

7. 编码的matlab实现。

本课程重点内容：图像信号的表示；数字图像信号处理技术的主要特点；MATLAB图像处理工具箱的使用方法；二维离散傅立叶变换；离散傅立叶变换的性质；直方图修正法；直方图规定化；图像的平滑；膨胀和腐蚀；开运算和闭运算；边缘检测；边缘跟踪；阈值分割方法；区域生长分割方法；离散图像退化的数学模型；无约束图像复原的病态性质；行程编码、霍夫曼编码、算术编码。

本课程难点内容：人眼视觉特性；图像类型的转换方法；基本图像和基本频谱；图像的均衡化和规定化；灰度图像的形态学处；分裂合并分割方法；运动图像分割方法；无约束图像复原的病态性质；匀速直线运动引起的图像模糊的复原；霍夫曼编码。

三、过程考核方式

本课程最终成绩按百分制评定。成绩包括平时成绩和期末考试（期末报告）成绩，平时成绩占40%，期末考试（期末报告）成绩占60%，平时成绩由作业、测验等组成，可根据情况适当修改考核方式。

四、教学方法

本课程是一门共40学时的专业选修课，课堂教学中采用案例式、启发式等多种教学方法，以学生为主体，以启发学生的思维为核心，调动学生的学习主动性和积极性；要求学生课外通过查阅文献等自学完成数字图像处理的实践内容，加深对理论教学的理解。利用在线教学和利用在线教学与测试平台进行课后训练与实践水平锻炼，利用QQ与微信的在线答疑等。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
绪论	4				4
MATLAB 图像处理工具箱	4				4
图像变换	6				6
图像增强	6				6
图像形态学	6				6
图像分割	6				6
图像复原	4				4
图像编码与压缩技术	4				4
合计	40				40

六、参考书

- [1] 王慧琴. 数字图像处理与应用（MATLAB版）. 北京：人民邮电出版社, 2019
- [2] 阮秋琦等译. 数字图像处理的MATLAB实现（第2版）. 北京：清华大学出版, 2013
- [3] 杨杰. 数字图像处理及MATLAB实现. 北京：电子工业出版社, 2013

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《人工智能》教学大纲

课程代码：175030

课程类别：专业基础

适用专业：数据科学与大数据技术

总学时：40

总学分：2.5

执笔人：高本宇

审定人：史健婷

审批人：李福军

一、课程性质及教学目标

《人工智能》课程是数据科学与大数据技术专业的一门专业选修课程，旨在为学生构建人工智能领域的基础理论框架与实践技能体系。

本课程融合了计算机科学、数学、统计学、认知科学等多个学科的知识，通过深入浅出的方式，使学生掌握人工智能的基本原理、核心技术、应用方法及前沿趋势。它不仅是对学生已学数据分析、机器学习、数据挖掘等课程知识的深化与拓展，也是培养未来数据科学家、人工智能工程师等高级专业人才的重要基石。

二、教学内容及要求

主讲教师可以根据学情、技术发展情况和授课实际等，对教学内容进行适当删减或补充。

（一）人工智能概述（4学时）

1. 能够了解人工智能的产生背景；
2. 能理解人工智能（artificial intelligence, AI）和预测的关系。

（二）编程环境和基础（4学时）

1. 掌握 Anaconda 的安装及使用；
2. 了解 Python 是什么，掌握 python 基础语法；
3. 掌握 NumPy 基础语法；
4. 掌握 sklearn 的安装。

（三）搜索算法（4学时）

1. 了解七桥问题、旅行商问题和迷宫问题；
2. 掌握 DFS 算法；
3. 掌握 BFS 算法；
4. 使用搜索算法解决七桥问题、旅行商问题和迷宫问题。

（四）进化算法（4学时）

1. 理解基本的遗传算法（genetic algorithm, GA）的原理；
2. 理解进化算法（evolutionary algorithm, EA）的原理和过程；
3. 理解多目标优化方法及过程；
4. 了解常见的几种进化算法。

（五）统计学习（6学时）

1. 理解机器预测中分类和回归的概念；
2. 学会区分有监督学习和无监督学习情况；
3. 学会常见的机器学习模型原理和应用。

（六）神经网络（6学时）

1. 理解神经网络的发展和计算机的发展之间的关系；
2. 掌握神经网络中的非线性、通用拟合性质和深度特征提取性质；

3. 理解神经网络的反向传播算法原理；
4. 深度学习库构建和训练基本的神经网络模型。

（七）深度学习（4学时）

1. 了解卷积操作的基本概念和计算过程；
2. 理解 CNN 中浅层及深层的卷积层提取到的图像特征的层次的区别；
3. 掌握如何使用 python 编码实现 CNN 模型；
4. 理解 RNN 的原理；
5. 理解梯度消失和梯度爆炸的概念及产生原因；
6. 理解 LSTM 的原理；
7. 掌握如何使用 python 编码实现 LSTM 模型；
8. 了解 RNN 在自然语言处理领域以外的应用。

（八）深度学习模型（4学时）

1. 了解图像分类、检测和分割的基本概念；
2. 了解常用的图像分类、检测和分割数据集；
3. 掌握利用 python 搭建、训练及评估深度神经网络模型的编程实现；
4. 了解知名的图像分类、检测和分割模型。

（九）智能体学习（4学时）

1. 理解强化学习发展、组成和基本原理；
2. 理解强化学习的分类、值函数和策略学习等方法；
3. 理解大规模分布式强化学习框架；
4. 了解 AlphaGo 的运行原理；
5. 了解智能体学习的高级强化学习方法；
6. 了解多智能体强化学习的基本原理。

本课程重点内容：人工智能的定义和研究内容；NumPy 基础语法，sklearn 库；DFS 算法，BFS 算法；遗传算法应用，进化算法应用；分类和回归的概念，有监督学习和无监督学习；神经网络的反向传播算法，深度学习库构建和训练基本的神经网络模型；CNN 模型、RNN 模型、LSTM 模型；利用 python 搭建、训练及评估深度神经网络模型的编程实现；强化学习的分类、值函数和策略学习等方法、大规模分布式强化学习框架。

本课程难点内容：NumPy 基础语法；搜索算法原理；遗传算法的原理，进化算法的原理；常见的机器学习模型原理；神经网络的反向传播算法原理；RNN 的原理、LSTM 的原理；利用 python 搭建深度神经网络模型；强化学习基本原理、智能体强化学习的基本原理。

三、过程考核方式

本课程最终成绩按百分制评定。成绩包括平时成绩和结课报告（期末考试）成绩，平时成绩占 40%，结课报告（期末考试）成绩占 60%，平时成绩由作业、测验等组成，可根据情况适当修改考核方式。

四、教学方法

本课程是一门共 40 学时的专业选修课，课堂教学中采用案例式、启发式等多种教学方法，以学生为主体，以启发学生的思维为核心，调动学生的学习主动性和积极性；要求学生课外通过查阅文

数据科学与大数据技术专业教学大纲

献等自学完成人工智能的实践内容，加深对理论教学的理解。利用在线教学和利用在线教学与测试平台进行课后训练与实践水平锻炼，利用QQ与微信的在线答疑等。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
人工智能概述	4				4
编程环境和基础	4				4
搜索算法	4				4
进化算法	4				4
统计学习	6				6
神经网络	6				6
深度学习	4				4
深度学习模型	4				4
智能体学习	4				4
合计	40				40

六、参考书

- [1] 刘刚. 人工智能导论. 北京: 北京邮电出版社, 2020
- [2] 尚文倩. 人工智能. 北京: 清华大学出版社, 2017
- [3] 王万良. 人工智能导论 (第4版). 北京: 高等教育出版社, 2017

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《深度学习基础》教学大纲

课程代码：176022

课程类别：专业基础

适用专业：数据科学与大数据技术

总学时：32

总学分：2

执笔人：高玉军

审定人：史健婷

审批人：高玉军

一、课程性质及教学目标

《深度学习基础》是数据科学与大数据技术专业的一门专业选修课程，旨在为学生奠定深度学习领域的理论基础与实践能力。随着大数据时代的到来，深度学习作为人工智能领域的重要分支，已成为处理复杂数据、实现高效模式识别与智能决策的关键技术。

本课程通过系统介绍深度学习的基本原理、核心算法、常用模型及其实际应用，旨在培养学生的数据科学思维、算法设计与实现能力，以及利用深度学习技术解决实际问题的能力，为后续的高级课程及专业研究打下坚实基础。

二、教学内容及要求

（一）深度学习概述（2学时）

1. 了解人工智能、机器学习与深度学习；
2. 学习并掌握深度学习的发展、已经取得的应用进展；
3. 掌握人工智能与深度学习的未来。

（二）PyTorch 快速入门（4学时）

1. 掌握深度学习的主要框架；
2. 熟悉PyTorch安装；
3. 掌握张量的计算与操作；
4. 掌握PyTorch中的自动微分；
5. 掌握torch.nn模块；
6. 掌握PyTorch中数据操作和预处理。

（三）PyTorch 深度神经网络及训练（6学时）

1. 掌握随机梯度下降算法的原理及实现；
2. 熟悉并掌握PyTorch中的优化器及应用；
3. 熟悉并掌握PyTorch中的损失函数及应用；
4. 掌握过拟合及防止过拟合的方法及应用；
5. 熟悉并掌握网络超参数的初始化方法及应用；
6. 掌握PyTorch中定义网络的方式；
7. 熟悉并掌握PyTorch模型保存和加载方法及实现。
8. 掌握深度模型评估方法及实现。

（四）基于 PyTorch 的相关可视化工具（2学时）

1. 了解网络结构的可视化技术；
2. 掌握训练过程的可视化方法；
3. 掌握使用 Visdom 进行可视化。

（五）全连接神经网络（4学时）

1. 了解全连接神经网络结构；

2. 熟悉全连接神经网络模型的搭建、训练和可视化；

3. 掌握全连接神经网络回归模型及实现。

(六) 卷积神经网络 (6 学时)

1. 了解卷积神经网络基本单元；

2. 熟悉并掌握经典深度卷积神经网络模型；

3. 掌握卷积神经网络识别应用；

4. 掌握对预训练好的卷积神经网络微调方法；

5. 掌握卷积神经网络进行情感分类方法；

6. 学会掌握预训练好的卷积神经网络的方法及可视化。

(七) 循环神经网络 (4 学时)

1. 了解常见的循环神经网络结构；

2. 掌握循环神经网络手写体分类；

3. 掌握 LSTM 进行中文新闻分类；

4. 掌握 GRU 网络进行情感分析。

(八) 图像风格迁移与生成对抗网络 (4 学时)

1. 了解常见的图像风格迁移方式；

2. 掌握固定风格固定内容的普通风格迁移实战；

3. 掌握固定风格任意内容的快速风格迁移实战；

4. 掌握分布式处理技术和 CPU 上使用预训练好 GPU 模型；

5. 掌握生成对抗网络的结构及实现。

本课程重点内容：深度学习的应用；torch.nn 模块，PyTorch 中数据操作和预处理；PyTorch 中的优化器及应用，PyTorch 中的损失函数及应用，网络超参数的初始化方法及应用；训练过程的可视化方法；全连接神经网络模型的搭建、训练和可视化；卷积神经网络识别应用，预训练好的卷积神经网络微调方法；循环神经网络结构；固定风格固定内容的普通风格迁移实战，固定风格任意内容的快速风格迁移实战

本课程难点内容：PyTorch 安装；梯度下降算法的原理；使用 Visdom 进行可视化；全连接神经网络模型的可视化；预训练好的卷积神经网络的方法及可视化；LSTM 进行中文新闻分类；生成对抗网络的结构及实现。

三、过程考核方式

本课程最终成绩按百分制评定。成绩包括平时成绩和期末报告（期末考试）成绩，平时成绩占 40%，期末报告（期末考试）成绩占 60%，平时成绩由作业、测验等组成，可根据情况适当修改考核方式。

四、教学方法

本课程是一门 32 学时的专业选修课，课堂教学中采用案例式、启发式等多种教学方法，以学生为主体，以启发学生的思维为核心，调动学生的学习主动性和积极性；要求学生课外通过查阅文献等自学完成深度学习基础的实践内容，加深对理论教学的理解。利用在线教学和利用在线教学与测试平台进行课后训练与实践水平锻炼，利用 QQ 与微信的在线答疑等。

五、各教学环节学时分配

数据科学与大数据技术专业教学大纲

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
深度学习概述	2				2
PyTorch 快速入门	4				4
PyTorch 深度神经网络及训练	6				6
基于 PyTorch 的相关可视化工具	2				2
全连接神经网络	4				4
卷积神经网络	6				6
循环神经网络	4				4
图像风格迁移与生成对抗网络	4				4
合计	32				32

六、参考书

- [1] 孙玉林, 余本国著. PyTorch 深度学习入门与实践. 北京: 中国水利水电出版社, 2020
- [2] 李沐. 动手学深度学习. 北京: 人民邮电出版社, 2019
- [3] 张敏. PyTorch 深度学习实战. 北京: 电子工业出版社, 2020

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《软件工程技术》教学大纲

课程代码: 176023 课程类别: 专业基础 适用专业: 数据科学与大数据技术

学时: 40 学分: 2.5

执笔人: 王光群 审定人: 史健婷 审批人: 李福军

一、课程性质及教学目标

《软件工程技术》是数据科学与大数据技术专业的一门必修课程。学生学习完本课程后应深入理解计算机软件的概念和特点、软件开发面临的困难和挑战、软件工程的基本思想和原则;系统掌握软件开发、管理和维护的过程、方法学和工具,包括需求分析、软件设计、系统建模、代码实现、软件测试、质量保证、软件维护、项目管理等,并能运用它们来开发具有一定规模和复杂性的高质量软件系统。随着软件作用和地位的不断增强以及软件成为重要的信息基础设施,软件开发和运维变得日趋重要,因此《软件工程技术》课程在计算机大类专业的人才培养体系中具有核心和关键性的地位。

通过本课程的学习,使学生掌握软件工程的基本概念和基础理论知识,达到熟练地运用各种软件实用开发方法和基本技术方法。了解软件工程各领域的发展动向,从事各种类型软件项目的工程化开发;培养具有独立工作能力的高素质高水平的软件管理及开发人才。

二、教学内容及要求

(一) 软件工程基础(10学时)

1. 了解软件的概念、特点和分类。
2. 掌握软件危机基本概念及基本知识。
3. 掌握软件工程概念、目标和原则。
4. 掌握软件生存期方法学及软件过程模型和开发方法。

(二) 软件需求分析(6学时)

1. 了解软件需求工程基础知识。
2. 掌握软件需求获取的过程和描述。
3. 掌握软件需求分析的任务、过程和UML模型。
4. 了解需求规格说明与评审。

(三) 软件概要设计(4学时)

1. 掌握软件设计的过程和原则。
2. 掌握面向对象软件设计方法学。
3. 掌握软件体系结构设计的过程、UML模型。
4. 了解概要设计文档与评审。

(四) 详细设计(6学时)

1. 掌握用户界面设计的任务、原则和过程。
2. 掌握详细设计的过程和UML模型。
3. 了解详细设计规格说明与评审。

(五) 软件实现(2学时)

1. 掌握编写代码的任务、方法和高质量编码。
2. 掌握代码缺陷的基础知识和代码重用的重要作用。

数据科学与大数据技术专业教学大纲

（六）软件测试（4学时）

1. 掌握软件测试的基本概念。
2. 掌握软件测试的步骤及基本方法。
3. 掌握软件测试的策略和软件测试技术。
4. 掌握面向对象软件测试方法。

（七）软件部署、软件维护与演化（4学时）

1. 了解程序设计语言基础知识。
2. 掌握软件部署的概念及方法。
3. 掌握软件维护和演化的概念。
4. 软件维护技术、过程及软件可维护性。

（八）软件项目管理（4学时）

1. 掌握软件项目及其特点。
2. 掌握软件项目管理的对象和要素。
3. 掌握软件项目管理的常见方法。

本课程重点内容：软件生存周期方法学和软件过程模型，软件需求获取的过程和描述，软件需求分析的任务、过程和UML模型，软件设计的过程和原则。对象软件设计方法学，软件体系结构设计的过程和UML模型，详细设计的过程和UML模型，高质量编码与代码缺陷，软件测试的步骤及基本方法、软件测试技术和面向对象软件测试方法，软件部署、维护和演化的概念、软件维护技术、过程及软件可维护性，软件项目管理的对象和要素，软件项目管理的常见方法。

本课程难点内容：软件过程模型，软件需求分析的UML模型，软件体系结构设计的UML模型，详细设计的UML模型，高质量编码，软件测试技术，软件的可维护性，软件度量和关键路径分析。

三、过程考核方式

课程最终成绩按百分制评定。成绩包括平时成绩和期末考试成绩，平时成绩占 30%，期末考试成绩占 70%，平时成绩由日常评价、课内测试等组成，可根据情况适当修改考核方式和比例。

四、教学方法和手段

本课程将采用课堂教学和课后实践相结合的讲授方式。理论教学尽可能地采用图表加案例驱动的方式，这样可以更好地帮助学生深入理解和认识软件工程的基本思想、原则、方法、技术和工具。这种方式能够让学生既能了解掌握理论，又能看到具体案例的应用。使学生上课不感觉枯燥，可以逐步培养学习兴趣，同时做到了理论与应用的结合。

五、各教学环节学时分配

授课内容	第一课堂	第二课堂	实验	实践实训	合计
软件工程基础	10				10
软件需求分析	6				6
软件概要设计	4				4
软件详细设计	6				6
软件实现	2				2
软件测试	4				4

数据科学与大数据技术专业教学大纲

软件维护与演化	4				4
软件项目管理	4				4
合计	40				40

六、教材及参考书目

- [1] 毛新军. 软件工程：从理论到实践. 北京：高等教育出版社，2022
- [2] 齐治昌. 软件工程（第4版）. 北京：高等教育出版社，2019
- [3] 钱乐秋. 软件工程（第3版）. 北京：清华大学出版社，2016

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《数值分析》教学大纲

课程代码：172022 课程类别：专业基础 适用专业：数据科学与大数据技术
总学时：40 总学分：2.5
执笔人：高本军 审定人：史健婷 审批人：李福军

一、课程性质及教学目标

《数值分析》是数据科学与大数据技术专业的一门专业基础课程，旨在为学生提供解决科学与工程计算问题中数值方法的基本理论、算法原理及其实践应用的系统知识。本课程是数学、计算机科学、工程科学等多学科交叉融合的产物，它不仅是数学分析理论在计算机科学中的具体应用，也是现代科学研究和工程技术领域中不可或缺的工具。

通过本课程的学习，学生能够掌握数值计算的基本原理、方法和技术，为解决复杂的数据处理、模拟仿真、优化决策等大数据相关问题奠定坚实的数学基础与计算能力。

二、教学内容及要求

（一）科学计算与MATLAB（4学时）

1. 科学计算的意义；
2. 误差基础知识；
3. MATLAB 软件。

（二）线性方程组的直接解法（6学时）

1. 高斯消去法；
2. 矩阵的三角分解；
3. QR分解和奇异值分解。

（三）多项式插值与样条插值（6学时）

1. 多项式插值；
2. 拉格朗日插值；
3. 牛顿插值；
3. 埃尔米特插值；
4. 三次样条插值。

（四）函数逼近（4学时）

1. 内积与正交多项式；
2. 最佳一致逼近与切比雪夫展开；
3. 最佳平方逼近；
4. 曲线拟合的最小二乘法。

（五）数值积分与数值微分（6学时）

1. 几个常用积分公式及其复合积分公式；
2. 变步长方法与外推加速技术；
3. 牛顿-科茨公式；
4. 高斯公式；
5. 多重积分的计算；
6. 数值微分。

数据科学与大数据技术专业教学大纲

(六) 线性方程组的迭代解法 (6学时)

1. 范数和条件数;
2. 基本迭代法;
3. 不定常迭代法。

(七) 非线性方程求根 (4学时)

1. 非线性方程求根的基本问题;
2. 二分法;
3. 不动点迭代方法;
4. 牛顿法;
5. 割线法;
6. 非线性最小二乘问题。

(八) 矩阵特征值与特征向量的计算 (4学时)

1. 幂方法;
2. QR方法。

本课程重点内容：误差的来源；有效数字；矩阵的三角分解原理；拉格朗日插值、牛顿插值、埃尔米特插值原理；最佳平方逼近、曲线拟合的最小二乘法原理；几个常用积分公式及其复合积分公式；范数和条件数，基本迭代法原理；二分法，不动点迭代方法，牛顿法，割线法；幂方法。

本课程难点内容：QR分解和奇异值分解的推导；三次样条插值的推导；最佳一致逼近与切比雪夫展开的原理；变步长方法与外推加速技术、牛顿-科茨公式的推导；不定常迭代法的推导；不动点迭代方法的推导；QR方法的推导。

三、过程考核方式

本课程最终成绩按百分制评定。成绩包括平时成绩和期末考试成绩，平时成绩占40%，期末考试成绩占60%，平时成绩由作业、测验等组成。

四、教学方法

本课程是一门共40学时的专业选修课，课堂教学中采用案例式、启发式等多种教学方法，以学生为主体，以启发学生的思维为核心，调动学生的学习主动性和积极性；要求学生课外通过查阅文献等自学完成数值分析的实践内容，加深对理论教学的理解。利用在线教学和利用在线教学与测试平台进行课后训练与实践水平锻炼，利用QQ与微信的在线答疑等。

五、各教学环节学时分配

授课内容 (知识点)	第一课堂	第二课堂	实验	实践实训	合计
科学计算与 MATLAB	4				4
线性方程组的直接解法	6				6
多项式插值与样条插值	6				6
函数逼近	4				4
数值积分与数值微分	6				6
线性方程组的迭代解法	6				6

数据科学与大数据技术专业教学大纲

非线性方程求根	4				4
矩阵特征值与特征向量的计算	4				4
合计	40				40

六、参考书

- [1] 同济大学计算数学教研室编著. 现代数值计算 (第二版). 北京: 人民邮电出版社, 2014
- [2] 李庆扬. 数值分析. 北京: 清华大学出版社, 2008
- [3] 王仁宏. 数值逼近. 北京: 高等教育出版社, 2012

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《最优化方法》教学大纲

课程代码：176025

课程类别：专业基础 适用专业：数据科学与大数据技术

总学时：40

总学分：2.5

执笔人：付喜辉

审定人：史健婷

审批人：李福军

一、课程性质及教学目标

本课程为数据科学与大数据技术专业大学本科生必修的一门专业基础课程。

教学目标是使学生掌握“最优化”各主要分支的基本概念、数学模型及其求解方法，通过计算机上机，使学生能把实际问题构成数学模型、选择适当方法、运用计算机求出最优解或满意解，以提高学生分析和解决实际问题的能力，也为进一步学习后继课程打下基础。

二、教学内容及要求

（一）线性规划与单纯法（12学时）

1. 理解线性规划的数学模型。
2. 掌握线性问题标准型及解的概念。
3. 掌握线性规划问题图解法。
4. 掌握线性规划问题解的性质。
5. 掌握线性规划的单纯型法。
6. 了解求解线性规划的人工变量法。

（二）对偶问题与灵敏度分析（8学时）

1. 了解对偶线性规划问题。
2. 掌握对偶问题的基本性质。
3. 掌握对偶单纯型法。
4. 理解线性规划的灵敏度分析。

（三）整数规划（2学时）

1. 了解整数规划的数学模型。
2. 理解分枝定界法和割平面法。

（四）无约束非线性最优化方法（10学时）

1. 理解无约束问题基本概念。
2. 掌握一维搜索方法。
3. 理解最速下降法。
4. 了解Newton法。
5. 了解拟Newton法。

约束问题的非线性最优化方法（8学时）

1. 掌握约束非线性规划问题的最优性条件。
2. 了解罚函数法。
3. 了解乘子法。
4. 了解可行方向法。

本课程重点内容：线性规划的单纯型法、对偶问题的基本性质及对偶单纯型法、一维搜索算法、非线性规划问题的最优性条件。

数据科学与大数据技术专业教学大纲

本课程难点内容：线性规划的单纯型法、非线性规划问题的最优性条件。

三、过程考核方式

课程最终成绩按百分制评定。成绩包括平时成绩和期末考试成绩，平时成绩占 40%，期末考试成绩占 60%，平时成绩由课堂表现、作业、测验等组成，可根据情况适当修改考核方式。

四、教学方法

利用多媒体教学，采取学练结合的方式进行教学。

五、各教学环节学时分配

授课内容（知识点）	第一课堂	第二课堂	实验	实践实训	合计
线性规划与单纯型法	12				12
对偶问题与灵敏度分析	8				8
整数规划线性规划	2				2
无约束非线性最优化方法	10				10
约束非线性最优化方法	8				8
合计	40				40

六、参考书

- [1] 王开荣. 最优化方法. 北京：科学出版社, 2012
- [2] 李学文. 最优化方法. 北京：北京理工大学出版, 2018

七、说明

在执行大纲过程中应结合学生的实际情况，可适当调整授课内容与进度，考核方式的百分比。

《云计算与大数据》教学大纲

课程代码：176026 课程类别：专业基础 适用专业：数据科学与大数据技术
学时：32 学分：2
执笔人：张兴华 审定人：史健婷 审批人：李福军

一、课程性质及教学目标

本课程是数据科学与大数据技术专业的一门选修课程。云计算与大数据课程是介绍目前信息技术领域热点“云计算大数据”的现状及其发展，探讨云计算大数据的主要技术及发展。课程主要内容包括：云计算与大数据概述、虚拟化技术、Google云计算及开源云平台等。

通过本课程的学习，让学生对云计算与大数据有一个系统的、全面的了解，掌握云计算与大数据的基础知识和理论，掌握部署一个私有云所必需的基本理论、基本知识和基本技能，为更深入地学习和今后的实践打下良好的基础。

二、教学内容及要求

（一）云计算大数据技术概览（4学时）

1. 掌握云计算的技术特征。
2. 了解云计算的起源和大规模云计算数据中心相关知识。
3. 掌握云计算与其它热门技术的知识。
4. 了解主流的商用云计算平台和开源云计算平台。
5. 掌握云计算的三元认识论内容。
6. 了解云计算的三元方法论内容。

（二）虚拟化技术（10学时）

1. 掌握虚拟化的定义、分类、架构和特性。
2. 掌握服务器虚拟化（CPU、内存和I/O）实现原理。
3. 掌握网络虚拟化、存储虚拟化和桌面虚拟化的实现技术。
4. 了解超融合平台的构成与应用和PVE基础知识。

（三）Google云计算原理与应用（4学时）

1. 掌握GFS的系统架构、容错机制及系统管理技术。
2. 掌握MapReduce的编程模型和实现机制。
3. 掌握Paxos算法的设计原理和实现步骤。
4. 掌握Chubby的系统设计思想、文件系统与通信协议及Paxos算法在Chubby中的应用。

（四）云计算架构（2学时）

1. 了解云计算的本质。
2. 掌握云计算的架构。
3. 掌握云计算的服务模型。
4. 掌握云计算服务部署模式。

（五）云计算安全（4学时）

1. 掌握云计算安全的基本概念和术语。
2. 掌握云计算安全风险分析及安全防护模型。

数据科学与大数据技术专业教学大纲

3. 掌握等保 2.0 下的云安全要求。

(六) 云计算数据中心 (4 学时)

1. 了解云数据中心的特征。
2. 掌握云数据中心的网络部署。
3. 了解云数据中心的绿色节能技术及自动化管理技术。
4. 掌握云数据中心的容灾备份。

(七) OpenStack 开源平台 (4 学时)

1. 了解 OpenStack 的产生背景和提供的主要服务。
2. 掌握计算服务 Nova 组件的构成和体系结构。
3. 掌握对象存储服务 Swift 组件的构成和基本原理。
4. 掌握镜像服务 Glance 组件的作用和组成部分。

本课程重点内容：云计算的三元方法论，服务器虚拟化（CPU、内存和 I/O）实现原理，网络虚拟化、存储虚拟化的实现技术，GFS 的系统架构、MapReduce 实现机制、Paxos 算法、Chubby 的系统设计思想、文件系统与通信协议，云计算架构，服务模型和部署模型，云安全防护模型，云数据中心的网络部署和容灾备份，计算服务 Nova 组件、对象存储服务 Swift 组件。

本课程难点内容：CPU、内存和 I/O 虚拟化的实现原理，网络虚拟化的实现技术，Paxos 算法，GFS 的容错机制及 Chubby 的通信协议，云计算架构，云安全防护模型，云数据中心的网络部署，Nova 组件的体系结构，Swift 组件的基本原理。

三、过程考核方式

课程最终成绩按百分制评定。成绩包括平时成绩和期末考试成绩，平时成绩占 30%，期末考试成绩占 70%，平时成绩由课内测试、课外实验等组成，可根据情况适当修改考核方式。

四、教学方法与手段

本课程是一门 32 学时的专业选修课，对授课内容以启发式讲授教学为主，采用多媒体 PPT 的教学方法，增加课堂信息量，浅显通俗地对概念、定义和原理进行解释，以增加教学的直观性，激发学生的学习积极性；教学过程中注意各个知识点的关联性，以使學生更好地理解课程内容；对课程中关键性概念、设计思想方面的问题可辅以课堂讨论的形式；不断完善课件的制作，使其更加适应教学的需要。

五、各教学环节学时分配

教学内容	第一课堂	第二课堂	实验	实践实训	合计
云计算技术概览	4				4
虚拟化技术	10				10
Google 云计算原理与应用	4				4
云计算架构	2				2
云计算安全	4				4
云计算数据中心	4				4
OpenStack 开源平台	4				4

数据科学与大数据技术专业教学大纲

合计	32				32
----	----	--	--	--	----

六、参考书

- [1] 王伟. 云计算原理与实践. 北京: 人民邮电出版社, 2018
- [2] 刘鹏. 云计算(第三版). 北京: 电子工业出版社, 2015
- [3] 徐小龙. 云计算技术及性能优化. 北京: 电子工业出版社, 2017

七、说明

在执行大纲过程中应结合学生的实际情况, 可适当调整授课内容与进度, 考核方式的百分比。

《认识实习》教学大纲

课程代码：176150 课程类别：专业能力 适用专业：数据科学与大数据技术
学时：1周 学分：1
执笔人：李彪 审定人：史健婷 审批人：李福军

一、实习性质及教学目标

认识实习是数据科学与大数据技术专业重要的实践教学环节之一。实习过程中有针对性的调研、参观实际工程现场，培养学生观察问题、思考问题的能力，为进一步提高学生对数据科学与大数据技术专业的认识、提升学生学习兴趣、树立学生的专业自豪感及后续专业知识的学习打下良好的基础。

教学目标：拓展学生的专业视野，使学生对数据科学与大数据技术专业及其相关行业有较深入的认识，激发学生学习兴趣，增强学生的专业自豪感；了解专业发展前景以及数据科学与大数据技术专业相关岗位的新技术、新知识，培养学生的观察问题、思考问题和解决问题的能力；培养学生数据思维、学科精神和工匠精神等。

二、实习教学基本环节

序号	实习环节名称	学时/天
1	实习准备及动员	1
2	参观校外实践基地，了解企业的工作模式和文化	1
3	讲座：大数据领域行业分析，了解大数据行业的当前情况和发展趋势	1
4	大数据实际项目的介绍和演示，了解项目的开发过程	1
5	实习总结并进行实习答辩	1
合计		5

三、实习各环节基本内容及要求

（一）实习环节

实习环节 1：实习准备及动员

基本内容：讲解实习任务、实习安排和实习期间的要求；介绍实习期间安全等方面的注意事项；讲解实习日记和实习总结报告的撰写方法等。

实习环节 2：参观校外实践基地，了解企业的工作模式和文化

基本内容：参观校外实践基地，介绍本专业相关企业的工作模式和文化；介绍数据科学相关专业的专业发展前景。

实习环节 3：讲座：大数据领域行业分析，了解大数据行业的当前情况和发展趋势

基本内容：大数据领域行业分析；介绍实习企业在大数据领域采用的新技术、新知识的运用现状及产品的研发概况；讲解与本专业相关领域的理论研究及发展。

实习环节 4：大数据实际项目的介绍和演示，了解项目的开发过程

基本内容：大数据实际项目的介绍和演示，介绍企业产品设计的分工及常用工具的使用。

实习环节 5：实习总结并进行实习答辩

基本内容：指导学生结合认识实习情况，撰写实习总结报告；进行认识实习答辩。

（二）实习要求

数据科学与大数据技术专业教学大纲

1. 实习前与合作企业沟通落实学生实习时间、实习内容；召开实习动员会，对学生进行实习安全教育，明确实习任务、实习安排和实习期间的要求。
2. 全面认识实习企业在数据科学领域的典型产品研发流程，了解数据科学相关的新技术、新知识的运用及使用情况。
3. 了解企业产品的分工设计，深入理解团队合作精神。
4. 通过实习总结报告的撰写，锻炼学生的文字组织能力。
5. 根据实习内容，按时做好实习日记；实习总结要求内容充实、重点突出、格式规范、逻辑性强的总结报告。

四、教学参考资料与教学形式

1. 教学参考资料：企业或公司的介绍和产品研发资料。
2. 教学形式：集中参加企业和公司现场实践学习。

五、考核方式与成绩评定方法

考核方式：教师根据学生参加实习过程的表现及实习日记、实习总结报告和实习答辩情况评定成绩。

评分标准：

1. 认真实习，遵守纪律，能按时参加实习，实习日记和实习总结报告质量高，全面完成实习大纲的要求，有独特见解或合理化建议可被评为优秀（90~100分）。
2. 实习期间纪律较好，能够按照要求较好地完成实习环节的内容，实习日记和实习总结报告能较好地满足要求，质量较高，有一定的独立工作能力者，成绩为良好（80~89分）。
3. 实习期间能够按照要求完成实习环节的内容，能完成实习日记和实习总结报告，但质量一般者，成绩评定中等（70~79分）。
4. 基本能够按照要求完成实习环节的内容，能完成实习日记和实习总结报告，但质量一般者，成绩评定为及格（60~69分）。
5. 实习日记不合格，或实习期间表现差，有严重违反纪律行为者，成绩评定为不及格（0~59分）。

六、说明

实习大纲在实施过程中，指导教师可以根据实习单位的具体条件，对实习内容及考核方式做适当必要的修改。

《专业实训（基础）》教学大纲

课程代码：176151 课程类别：专业能力 适用专业：数据科学与大数据技术
学时：4周 学分：4
执笔人：陈学刚 审定人：史健婷 审批人：李海华

一、实训的性质、目的与任务

专业实训（基础）是数据科学与大数据技术专业重要的实践教学环节之一。通过专业实训（基础），提高学生实际开发能力。实训以实践能力为导向，结合实际问题，通过具体项目使学生实际参与到项目的实施过程中，将所学理论知识同应用有机的结合起来，锻炼学生分析、解决较复杂问题的能力，提高学生独立开发大数据相关系统的能力。

二、实践教学的基本环节及内容

序号	实训环节名称	学时/周
1	需求分析：对所完成的专业项目进行调研，需求收集，需求整理，需求分析，明确系统的需要	1
2	系统分析与设计：根据选定的对系统进行功能模块划分与设计。通过实际运行系统与并进行测试，发现系统的不足及设计中的错误，修正系统使其符合预期设计功能	1
3	数据预处理与试运行	1
4	撰写实训报告：按格式要求写出完整、规范的报告并打印。其中模块图、流程图要清楚、规范	1
合计		4

三、实训各环节基本内容及要求

（一）基本内容

1. 实训动员

对学生进行实训之前的动员，主要讲解实训的目的、有关的实训要求、主要的实训内容、安全和注意事项以及考核方式等。

2. 需求分析

对所完成的专业项目进行调研，需求收集，需求整理，需求分析，明确系统的需要。

3. 系统分析与设计

根据选定的对系统进行功能模块划分与设计。

4. 系统实现与测试

开发完成设计阶段划分好的功能模块，可利用仿真数据，或真实项目数据进行系统测试。软件设计与实现要符合软件工程规范。

5. 撰写实训报告

按格式要求写出完整、规范的报告并打印。其中模块图、流程图、数据表等要清楚、规范。

6. 实训答辩

对学生的实训工作进行评估和考核。

数据科学与大数据技术专业教学大纲

(二) 实训要求

1. 实训主要内容是设计实现一个软件系统，如系统要有信息的查询、增加、修改、删除等管理系统的主要功能。

2. 具体题目自拟，如：“xx 公司进销存管理系统”、“xx 网上书店系统”、“基于 ios 的 xxAPP 应用”等，名字要具体到哪个公司或单位。

3. 开发平台和开发工具自选，建议在 JavaEE 平台下基于 SSH 框架进行开发。

4. 系统至少有 4 个模块，信息管理系统的每个模块要有信息的查询、增加、修改、删除等基本功能。

5. 在整个实训过程中，要求学生每周向指导教师进行口头汇报，汇报内容包括但不限于完成的工作、遇见的问题、解决问题的方法及有何进步等。

四、主要仪器设备

服务器、电脑及相应软件。

五、教学参考资料与教学形式

1. 教学参考资料

[1] 姜桂洪. MySQL 数据库应用与开发. 北京：清华大学出版社, 2020

[2] 王翔. R 语言数据可视化与统计分析基础. 北京：机械工业出版社, 2019

[3] 王泰峰. 软件项目综合开发综合实训--Java 篇. 北京：人民邮电出版社, 2006

2. 教学形式

课堂讲授结合上机实践。课堂讲授时，教学方法上尽量采用启发式、讨论式教学，综合利用实验室环境条件，结合 PPT 和项目介绍、实践操作进行讲解，通过实训项目开发，培养学生的实践动手能力和解决问题的能力。

六、考核方式与成绩评定方法

成绩根据实训报告评定。主要考察的方面包括但不限于设计工作量、设计内容、方案、计算可靠性、文字和图表质量及学生的基础理论、专业知识水平、设计完成情况等。不同分数对应等级见下表。

成绩（分数）	评定等级
0 分-59 分	不及格
60 分-69 分	及格
70 分-79 分	中等
80 分-89 分	良好
90 分-100 分	优秀

七、说明

大纲在执行过程中，可根据教学需要适当调节专业实训进度。成绩评定形式可为百分制分数，或依据对应分数给定成绩等级。

《Spark 大数据处理技术课程设计》教学大纲

课程代码：176152 课程类别：专业能力 适用专业：数据科学与大数据技术
学时：3 周 学分：3
执笔人：史健婷 审定人：丁文妙 审批人：李福军

一、课程性质及教学目标

《Spark 大数据处理技术课程设计》是数据科学与大数据技术的专业能力课程之一。该课程设计的目标是从大数据处理与分析的实践出发、让学生掌握基于 Spark 平台进行大数据处理与分析的基本技能，能够实现对大规模结构化和非结构化数据的基本处理与分析，进而建构大数据处理与分析的基础知识体系。

本门课程在培养学生专业素质和思维能力的同时，通过将大数据技术和社会需求紧密结合，引导学生勤于学习、善于学习、乐于钻研、勇于探索、立志学好科学技术回馈国家、回馈社会。

二、课程设计教学环节

序号	课程设计环节名称	学时/天
1	Linux 系统的安装和常用命令	2
2	Spark 编程环境	3
3	Spark 编程基础	3
4	Spark 应用案例实战	3
5	Spark 优化和调优	4
合计		15

三、课程设计各环节基本内容及要求

（一）Linux 系统的安装和常用命令

目的：掌握 Linux 虚拟机的安装方法，利用虚拟机软件加载镜像文件，熟悉 Linux 系统的基本用法。

原理概述：Linux 虚拟机环境是 Spark 数据处理的基础要求，学习安装方法并利用虚拟机软件加载镜像文件不仅在 Spark 课程中需要，Linux 系统基本操作的学习更有益于未来工作。

方法与手段：教师利用教师机演示 Linux 安装过程，将步骤详解发给学生，学生通过查找资料等方式利用 PC 机自行操作完成安装。

（二）Spark 编程环境

目的：掌握 Spark 编程环境的搭建和配置，包括安装 Spark、配置环境变量、启动 Spark Shell 等。

原理概述：Spark 数据分析需要 Spark 编程环境的支持，后续通过 Spark shell 或独立 Spark 程序运行代码，实现数据分析结果。

方法与手段：主要通过教师讲解并演示环境搭建过程，学生通过线上资源和课程的学习完成环境搭建和简单程序的编写。

（三）Spark 编程基础

目的：掌握 Spark 基础编程的方法和技巧，包括数据的读取、转换、分析和输出等，通过示例案例展示 Spark 编程的具体操作和效果。

数据科学与大数据技术专业教学大纲

原理概述：把理论课中讲述的 RDD 键值对和 Spark SQL 等技术相贯通，掌握数据的读取转换和输出等方法。

方法与手段：以教师讲述和演示为主，学生动手实践实现不同案例。

（四）Spark 应用案例实战

目的：基于 Spark 编程实践，设计实际的大数据处理案例，帮助学生加深理解和熟悉 Spark 的实际应用场景，加强数据处理的能力。

原理概述：将生活中的实际案例引入课程设计，搭建 Spark 环境，利用编程基础进行 Spark 数据分析，数据量达到一定量级，学生能更深刻的感受 Spark 技术的处理优势。

方法与手段：以教师演示为主导，学生动手实现，并采取多种案例数据，鼓励学生创新实现。

（五）Spark 优化和调优

目的：了解 Spark 应用的性能瓶颈和调优问题，介绍提高 Spark 运行效率的方法和工具，为学生量身定制解决问题的思路和方案。

原理概述：Spark 框架相比 Hadoop 技术提高了数据处理和分析的效率，但在 Spark 运行过程中仍存在可提升效率的操作，如 RDD 持久化等。锻炼学生持续改进和优化的能力，突破性能瓶颈寻求调优方案。

方法与手段：以教师引导思维为开端，学生自主研讨并实践，展示运行结果进行 PK，激发学生的学习积极性。

四、主要仪器设备配置

PC 机，Java 编程环境

五、教学参考资料与教学形式

1. 教学参考资料

- [1] 林子雨, 郑海山, 赖永炫. Spark 编程基础 (Python 版). 北京: 人民邮电出版社, 2020
- [2] Holden Karau 等. Spark 快速大数据分析. 北京: 人民邮电出版社, 2018
- [3] 许利杰, 方亚芬. 大数据处理框架 Apache Spark 设计与实现. 北京电子工业出版社, 2020

2. 教学形式

课堂讲授结合上机操作，结合学生知识结构，全程按计划指导，以教师引导和学生操作为主，讲授为辅。课堂讲授时，教学方法上尽量采用启发式、讨论式教学，综合利用实验室环境条件，结合 PPT 和项目介绍、实践操作进行讲解，通过实训项目开发，培养学生的实践动手能力和解决问题的能力。

六、考核方式及成绩评定办法

本课程设计的学生成绩满分为 100 分，考核包含课程报告和答辩情况，答辩时需展示设计内容。

1. 课程报告（60 分）：占总成绩的 60%，包括报告的设计方案、流程、文字描述合理性、语句通顺性（表达是否清晰准确）；报告组织结构、图表和格式规范性等。

2. 答辩情况（40 分）：占总成绩的 40%，包括作品演示（系统运行情况）和讲解情况和回答问题情况及反映出的知识掌握情况等。

七、说明

可根据学生实际专业知识素养，结合知识体系的更新和变化实时调整授课内容及各部分内容的授课课时、考核方式和比例。

《大数据分析与应用课程设计》教学大纲

课程代码：176153 课程类别：专业能力 适用专业：数据科学与大数据技术
总学时：3周 总学分：3
执笔人：丁文妙 审定人：史健婷 审批人：李海军

一、课程性质及教学目标

《大数据分析与应用课程设计》是数据科学与大数据技术专业本科生必修的实践类课程，是锻炼学生大数据开发技术理论与实践融合与应用的综合运用能力。

根据现实大数据的应用场景，获取相应的数据集，按照大数据分析的基本流程，综合采用大数据的常用分析算法及技术和相应的工具，构建一套大数据应用系统，并对数据进行有针对性的分析，生成分析结果图表。通过课程设计的综合训练，培养学生实际分析问题、解决问题的能力，以及编程能力和综合运用能力。该目标分解为以下二个子目标：

- (1) 能够掌握大数据采集、存储、处理分析和可视化的主要开发方法和技术。
- (2) 能够根据具体的业务需求，设计和开发大数据系统。

二、课程设计教学基本环节

序号	课程设计基本环节	学时/天
1	需求分析、确定框架与工具	2
2	算法模型与数据采集、处理、分析	6
3	系统实现、测试和报告撰写	6
4	课程设计答辩	1
合计		15

三、课程设计各环节基本内容及要求

(一) 基本环节 1：需求分析、确定框架与工具

教学目标：根据所选题目进行需求分析，明确开发内容和所用开发框架与工具。

教学内容：复习软件工程的需求分析方法，介绍大数据领域常用的开发框架与工具。

学时分配：2天

学生设计任务：完成需求分析，确定开发框架与工具。

教学方法：教师讲授，学生查找资料与实践。

(二) 基本环节 2：算法模型与数据采集、处理、分析

教学目标：确定算法与模型，并能够进行数据采集、处理和分析。

教学内容：根据案例，介绍大数据常用的算法与模型、大数据数据处理的基本流程和方法。

学时分配：6天

学生设计任务：确定算法与模型，数据集的采集、处理、分析与可视化。

教学方法：教师讲授，学生查找资料与实践。

(三) 基本环节 3：系统实现、测试和报告撰写

教学目标：培养学生大数据分析的综合运用能力。

教学内容：根据案例，演示完整系统的综合实现过程。

学时分配：6天

数据科学与大数据技术专业教学大纲

学生设计任务：系统实现，并进行测试、优化和完善，完成课程设计报告的撰写。

教学方法：教师讲授，解答疑难和学生实践。

（四）基本环节 4：课程设计答辩

教学目标：培养学生知识总结、临场应变能力和表达能力。

教学内容：学生通过PPT展示实践流程、实践技术和实践成果，教师依据实践内容进行提问，并给出进一步完善及开发建议。

学时分配：1天

学生设计任务：完成课程设计报告，制作答辩PPT，按要求完成答辩环节。

教学方法：师生互动交流。

四、教学参考资料与教学形式

1. 教学参考资料

[1] 千锋教育. PYTHON快乐编程-数据分析与实战. 北京：清华大学出版社, 2021

[2] 阿里云计算有限公司. 大数据分析与应用(中级). 北京：高等教育出版社, 2021

2. 教学形式

课堂讲授结合上机实践。课堂讲授时，教学方法上尽量采用启发式、讨论式教学，综合利用实验室环境条件，结合PPT和项目案例介绍，培养学生的实践动手能力和解决问题的能力。

五、考核方式与成绩评定方法

本课程设计的学生成绩满分为100分，从以下四个方面进行综合考核与评定：

1. 设计态度（10分）：占总成绩的10%，包括出勤、设计态度和认真程度等情况。

2. 课程作品（15分）：占总成绩的20%，包括设计工作量及设计难度、设计方案、研究方法和设计程序完成情况等。

3. 课程报告（35分）：占总成绩的30%，包括报告的文字描述合理性、语句通顺性（表达是否清晰准确）；报告组织结构、图表和格式规范性等。

4. 答辩情况（40分）：占总成绩的40%，包括作品演示（系统运行情况）和讲解情况和回答问题情况及反映出的知识掌握情况等。

六、说明

本课程设计应在相关的专业基础课程、专业课程之后进行，以便将这些知识进行综合运用，为课程设计奠定坚实的基础。本课程设计教学时数为3周，其中必须保证上机学时不小于1周。

《专业实训（高级）》教学大纲

课程代码：176160 课程类别：专业能力 适用专业：数据科学与大数据技术

总学时：4周 学分：4

执笔人：陈学研 审定人：史健婷 审批人：李福军

一、实训的性质、目的与任务

专业实训（高级）是数据科学与大数据技术专业重要的实践教学环节之一，属于专业能力课。本课程针对大四学生，通过专业实训（高级），提高学生软件开发能力、数据分析能力、大数据平台离线环境搭建能力。实训以培养实践能力为导向，结合实际问题，通过具体项目使学生在参加项目实施的过程中，将理论所学知识同应用有机的结合起来，锻炼学生分析、解决较复杂问题的能力，提高学生独立开发较大型软、硬件系统、数据分析系统能力，建立数据仓库的能力。

二、实践教学的基本环节及内容

序号	实训环节名称	学时/周
1	需求分析与模型建立： 对所要完成的专业项目进行调研、需求收集、需求整理、需求分析、明确系统需要和系统边界。并依据需求分析形成的元数据，建立数据模型，规范数据模型	1
2	系统分析与设计： （1）若项目为开发系统软件，在需求分析基础上，依据第1周所建立的数据模型，根据选定题目对系统进行功能模块划分与设计。进行系统测试，发现系统的不足及设计中的错误，修正系统使其符合预期设计功能。 （2）若项目为数据分析与可视化，可在当前周进行数据获取与处理 （3）若项目为建离线数据仓库，可在当前周利用Map-Reduce技术实现数据分发；利用HDFS实现文件系统中数据的访问；利用Hive完成数据提取、转化和加载等工作	1
3	试运行与调试： （1）若项目为开发系统软件，在第2周基础上，进一步确定系统功能模块科学合理，并调试系统bug，开发UI，完善系统 （2）若项目为数据分析与可视化，可在当前周进行数据可视化，及对可视化数据进行分析，并在分析基础上得出结论，或者做出相应的决策 （3）若项目为数据仓库搭建，则需完成对数据的清洗或处理，并将数据导入数据仓库。实现数据分发、处理和存储	1
4	撰写实训报告： 按格式要求写出完整、规范的报告并打印。其中模块图、流程图要清楚、规范，涉及的表格内容需全面、完整，数据类型及数据间的联系要清晰，体现的技术路线需清晰、合理、可行	1
合计		4

三、实训各环节基本内容及要求

（一）基本内容

1. 实训动员：对整个大四年级的学生进行实训之前的动员，主要讲解工程实训的目的、有关实训要求、主要的实训内容、安全和注意事项以及考核方式等。

2. 需求分析：对所要完成的专业项目进行调研，需求收集，需求整理，需求分析，明确系统的需要。

3. 系统分析与设计：根据选定题目对系统进行功能模块划分与设计。

4. 系统实现与测试：开发完成设计阶段划分好的功能模块，可利用仿真数据或真实项目数据进行系统测试。软件设计与实现要符合软件工程规范。

5. 撰写实训报告：按格式要求写出完整、规范的报告并打印。其中模块图、流程图、数据表等要清楚、规范。

6. 实训答辩：对学生的实训工作进行评估和考核。

（二）实训要求

1. 实训主要内容是设计实现一个应用软件或数据分析与可视化平台，如为应用软件要有信息的查询、增加、修改、删除等管理系统的主要功能；如为可视化平台，则需有数据可视化效果、可视化分析结论、数据增删改的功能；如为离线数仓则需基于批处理模式，实现对大量实时数据或历史数据的批处理或分析功能，满足离线数据仓库数据质量高、准确性强、可靠性好的基本要求。

2. 具体题目自拟，如：“xx 公司进销存管理系统”、“xx 网上书店系统”、“基于 ios 的 xxAPP 应用”等，名字要具体到哪个公司或单位；数据分析与可视化的题目可为“xx 数据的分析与可视化”、“xx 公司 xx 数据的数据可视化算法研究”等；“xx 离线数据仓库的搭建”。

3. 开发平台和开发工具自选，建议在 JavaEE 平台下基于 SSH 框架进行开发；若为数据分析与可视化，可选择 Python 或 Spark、R 语言等分析工具；若为数据仓库，则需搭建 Hadoop 环境，并进一步配置 Map-Reduce、Hive、Flume、Kafka 或 Yarn 等环境。

4. 系统至少有 4 个模块，信息管理系统的每个模块要有信息的查询、增加、修改、删除等基本功能。

5. 在整个实训过程中，要求学生每周向指导教师进行口头汇报，汇报内容包括但不限于阶段完成的工作、遇到的问题、问题解决的方法、收获和提高等方面。

四、主要仪器设备

服务器、电脑及相应软件

五、教学参考资料与教学形式

1. 教学参考资料

[1] 姜桂洪. MySQL 数据库应用与开发. 北京：清华大学出版社, 2020

[2] 王翔. R 语言数据可视化与统计分析基础. 北京：机械工业出版社, 2019

[3] 王泰峰. 软件项目综合开发综合实训-Java 篇. 北京：人民邮电出版社, 2006

2. 教学形式

课堂讲授结合上机实践。课堂讲授时，教学方法上尽量采用启发式、讨论式教学，综合利用实验室环境条件，结合 PPT 和项目介绍、实践操作进行讲解，通过实训项目开发，培养学生的实践动手能力和解决问题的能力。

六、考核方式与成绩评定方法

数据科学与大数据技术专业教学大纲

成绩根据实训报告评定。主要考察的方面包括但不限于设计工作量、设计内容、方案、计算可靠性、文字和图表质量及学生的基础理论、专业知识水平、设计完成情况等。不同分数对应等级如下。

成绩（分数）	评定等级
0分-59分	不及格
60分-69分	及格
70分-79分	中等
80分-89分	良好
90分-100分	优秀

七、说明

大纲在执行过程中，授课教师可根据教学需要适当调节专业实训进度、考核方式和比例。

《生产实习》教学大纲

课程代码：176162 课程类别：专业能力 适用专业：数据科学与大数据技术
总学时：3周 总学分：3
执笔人：史健婷 审定人：丁文抄 审批人：李福军

一、实习性质及教学目标

生产实习是数据科学与大数据技术专业教学计划中非常重要的实践教学环节之一。通过实习，学生能够做到理论联系实际，掌握专业基本技能，从而培养学生实践动手能力和创新能力，提升学生的专业应用与实践技能。

生产实习的教学目标就是让学生了解和掌握本专业业务范围内现代化生产方式、管理方式、工艺过程和技术方法，培养和提高学生从事专业技术工作的基本技能，增强学生实际分析问题、综合解决问题的能力。

二、实习教学基本环节

序号	实习各环节名称	学时/天
1	实习准备及动员	1
2	去企业和公司实践学习，动手参与实践项目	13
3	实习总结	1
合计		15

三、实习各环节基本内容及要求

（一）实习基本内容

1. 实习动员

主要讲解实习目的、实习要求、实习内容、实习日记和总结的撰写、安全和注意事项、实习考核方式等。

2. 去企业和公司实践学习

数据科学与大数据技术专业的生产实习主要采取去企业和公司实践学习。实习的主要内容包括：

①在了解科研院所及企事业单位对数据科学知识应用与研究情况的基础上，确定这方面的研究方向；

②掌握计算机局域网规划、设计、安装和调试等基本技能；

③掌握大数据技术应用的开发方法、技术和过程；

④了解和掌握用计算机进行自动控制的具体过程和所用技术与方法；

⑤了解和掌握计算机网络编程和网络安全技术；

⑥掌握用软件工程思想开发软件的方法和过程；

⑦掌握大数据技术在国内公司和企业的应用；

⑧掌握数据科学领域的理论研究情况及今后的发展趋势。

3. 实习总结

指导学生根据3周的实习的情况撰写实习总结；与学生沟通实习的不足和改进措施；根据学生的实习表现和实习日记、总结给出实习成绩。

（二）实习要求

数据科学与大数据技术专业教学大纲

1. 实习前一学期与合作企业沟通落实学生实习时间、实习内容；召开实习动员会，对学生进行实习安全教育，明确实习任务、实习进行安排和实习期间的要求。

2. 全面认识实习企业在 IT 领域的典型产品研发流程，了解 IT 领域的新技术、新知识的运用及使用情况。

3. 了解现代化工厂企业的组织机构、生产管理、规章制度和运行机制。

4. 了解现代化数据分析系统的构成和设计。

5. 了解企业产品的分工设计，深入理解团队合作精神。

6. 根据实习内容每天按时做好内容详实的日记，实习结束后编写一份内容丰富、重点突出、格式规范、逻辑性强的生产实习报告；实习结束后可根据具体情况安排例如笔试的实习考核。

四、教学参考资料与教学形式

1. 教学参考资料：企业和公司的介绍和产品研发资料。

2. 教学形式：企业和公司现场实践学习，由企业和公司的研发工程师指导。

五、考核方式及成绩评定

1. 生产实习成绩应按生产实习大纲的要求，根据学生在生产实习过程的表现、实习日记、实习报告及教师的考核进行评定，并按 90~100 分、80~89 分、70~79 分、60~69 分、0~59 分五等级加以区分。

2. 认真实习，遵守纪律，能按时到场和结束，无无故缺勤者，实习日记、报告质量高，全面完成实习大纲的要求，有独特见解或合理化建议可被评为 90~100 分。

3. 实习期间纪律较好，较好完成实习大纲的要求，实习日记、报告能较好地满足要求，质量较高，有一定的独立工作能力者，成绩为 80~89 分。

4. 实习期间能完成大纲的要求，能完成实习日记、报告，但质量一般者，成绩评定为 70~79。

5. 基本完成实习大纲的要求，能基本完成实习日记、报告，但质量一般者，成绩评定为 60~69。

6. 实习日记或报告不合格，或实习期间表现差，有严重违反纪律行为者，成绩评定为“0~59”。

六、说明

实习大纲在实施过程中，指导教师可以根据实习单位的具体条件，对实习内容及考核方式有权做必要的修改。

《毕业实习》教学大纲

课程代码：176163 课程类别：专业能力 适用专业：数据科学与大数据技术
总学时：2周 总学分：2
执笔人：张兴华 审定人：史健婷 审批人：李福军

一、实习性质及教学目标

毕业实习是学生在修完人才培养方案中规定的全部理论课程并且完成其它教学环节的基础上进行的综合性实习，是学生毕业设计开始前的教学环节，毕业实习任务可以结合毕业设计选题进行。通过毕业实习，巩固和深化学生所学的专业基本理论、基本方法、基本技能，进一步将理论联系实际，培养工程意识和适应工作环境的能力，是重要的实践教学环节。该目标分解为如下子目标：

1. 通过毕业实习使学生了解实习单位的一般情况和社会需求，增加对本专业学科范围的感性认识。初步了解所学专业在国民经济建设中的地位、作用和发展趋势。了解本专业业务范围内的企业生产组织形式、管理方式、主要技术和方法。

2. 了解与计算机专业相关的工作及其重要作用，深入了解数据科学与大数据技术专业及其应用技术，了解计算机技术、数据科学和大数据技术相关背景知识。使学生对所学专业的意义和特点有更为全面的认识。

3. 培养学生理论联系实际、从实际出发分析问题、研究问题和解决问题的能力，将学生所学理论知识系统化。

二、实习教学基本环节

序号	实习环节名称	学时
1	实习准备及动员	1天
2	去企业和公司实践学习，动手参与实践项目	8天
3	实习总结	1天
合计		10天

三、实习各环节基本内容及要求

(一) 实习各环节基本内容

1. 实习动员

主要讲解实习目的、实习要求、实习内容，实习日记和总结的撰写，安全和注意事项，实习考核方式等。

2. 选择适合自己的企业和公司进行实践学习，动手参与实践项目。

数据科学与大数据技术专业的毕业实习主要采取去企业和公司实践学习，主要实习的内容包括：

- (1) 了解企业的经营范围和企业文化。
- (2) 了解实习单位对计算科学知识研究与应用的情况，以及数据科学与大数据技术专业的工程应用领域，
- (3) 了解其对相关的社会、健康、安全、法律、文化的影响，以及应承担的责任。
- (4) 了解和掌握软件的开发方法、技术和过程，了解企业级的大数据分析与应用的项目开发需要哪些技能。
- (5) 了解计算机网络编程和网络安全技术，掌握计算机局域网规划、设计、安装和调试等基

本技能。

- (6) 学习实习单位利用大数据技术发现、分析和解决具体行业的实际工程技术问题的方法。
- (7) 了解本专业市场需求，深入理解团队合作精神。
- (8) 了解数据科学与大数据技术理论研究情况及今后新技术的发展趋势。

3. 实习总结

指导学生根据 2 周的实习的情况撰写实习总结；与学生沟通实习的不足和改进措施；根据学生的实习表现和实习日记、总结给出实习综合成绩。

(二) 实习要求

1. 实习前一学期与合作企业沟通落实实习时间、实习内容。
2. 召开实习动员会，对学生进行实习安全教育，明确实习任务、实习进程安排和实习期间的要求。
3. 遵守学校和实习单位的组织管理、规章制度和运行机制。
4. 保护实习单位的商业机密和技术机密。
5. 根据实习内容每天按时做好内容详实的日记，实习结束后撰写一份内容丰富、重点突出、格式规范、逻辑性强的实习总结报告。

四、教学参考资料与教学形式

1. 教学参考资料

企业和公司的介绍和产品研发资料。

2. 教学形式

企业和公司现场实践学习，由企业和公司的研发工程师指导。

五、考核方式及成绩评定

1. 毕业实习成绩应按毕业实习大纲的要求，教师根据学生在实习过程的表现、实习日记、实习报告进行考核评定，并按 90~100 分、80~89 分、70~79 分、60~69 分、0~59 分五等加以区分。

2. 认真实习，遵守纪律，能按时到场和结束，无无故缺勤者，实习日记、报告质量高，全面完成实习大纲的要求，有独特见解或合理化建议可被评为 90~100 分。

3. 实习期间纪律较好，较好完成实习大纲的要求，实习日记、报告能较好地满足要求，质量较高，有一定的独立工作能力者，成绩为 80~89 分。

4. 实习期间能完成大纲的要求，能完成实习日记、报告，但质量一般者，成绩评定为 70~79。

5. 基本完成实习大纲的要求，能基本完成实习日记、报告，但质量一般者，成绩评定为 60~69。

6. 实习日记或报告不合格，或实习期间表现差，有严重违反纪律行为者，成绩评定为“0~59”。

六、说明

实习大纲在实施过程中，指导教师可以根据实习单位的具体条件，对实习内容及考核方式有权做必要的修改。

《毕业设计（论文）》教学大纲

课程代码：176164

课程类别：专业能力

适用专业：数据科学与大数据技术

总学时：16周

学分：16学分

执笔人：史健婷

审定人：丁文妙

审批人：李福军

一、毕业设计（论文）性质及教学目标

毕业设计（论文）是实现专业培养目标的重要阶段，是对学生学习、研究与实践成果的全面总结；是对学生综合素质与工程实践能力培养效果的全面检验；是重要的实践环节之一。

通过毕业设计（论文），培养学生的软、硬件系统开发和设计能力、大数据技术应用能力以及设计内容（论文）撰写能力，同时提高学生综合运用所学知识和技能去分析、解决实际问题的能力，检验学生的学习效果等。通过毕业设计（论文），应使学生在以下几方面能力得到训练和提高：

1. 综合运用所学的大数据专业理论知识和技术，培养解决实际问题的综合能力；
2. 掌握文献检索、资料查询的基本方法以及获取新知识的能力；
3. 具备熟练应用计算机（包括常用语言、工具软件及专用软件）的技能，具有较强的算法设计、算法分析与编程能力。

二、毕业设计（论文）教学基本环节

序号	毕业设计环节名称	学时
1	选题环节	3周
2	开题答辩环节	1周
3	中期检查环节	2周
4	毕业设计（论文）环节	10周
合计		16周

三、毕业设计各环节基本内容及要求

为保证毕业设计的质量，加强指导与监控工作，毕业设计共分为：选题、开题答辩、中期检查、毕业答辩和归档五个环节。

1. 选题环节：毕业设计（论文）的选题原则上是一人一题，以利于培养学生独立学习及工作的能力。对于工程实践题目，可根据个人参与完成的部分模块进行选题设计，但工作量必须满足要求，并且最终毕业设计需要拿出作品实物并能进行演示；对于极个别大型系统，可考虑多人一题。指导教师必须严格区分每人工作方向，毕业设计不能互相抄袭。题目要求师生共同酝酿、商讨，所选择的毕业设计（论文）题目必须符合本专业选题要求，任务书的设计内容要具体，设计目标要明确。

2. 开题答辩环节：重点考察学生是否充分了解课题的设计内容和设计目标，工作计划是否可行，是否具备课题所要求的客观条件（实验、设备、资料等），主要设计或研究内容是否科学、合理。

3. 中期检查环节：答辩委员会对学生毕业设计（论文）开展情况进行中期检查，包括：工作进度、遇到困难及解决办法、后期规划等内容，答辩委员会根据学生完成情况给出后期工作的指导性建议。

4. 毕业设计（论文）环节：学生必须按计划规定的时间内完成毕业设计（论文），并撰写毕业设计说明书（论文），按照学校要求完成查重任务并查重合格。经指导教师评阅通过并签字后，方可获得参加毕业答辩资格，主审教师审查毕业设计（论文）时，发现问题并没有达到要求的可以退回

数据科学与大数据技术专业教学大纲

毕业设计给指导教师，也可终止参加毕业答辩资格，直到满足要求方可重新获得答辩资格。

答辩学生应核照答辩时间准时参加答辩，遵守答辩会场的有关纪律，每个学生的答辩时间为30分钟，需要以PPT的形式报告自己所设计的内容，并出示有关图表、数据、实物等，个人讲解10分钟，之后回答答辩组成员所提出的问题，要求问题在4个以上。

答辩结束之后，将毕业设计（论文）的所有材料按专业、班级送到院（系）统一保管。

四、考核方式与成绩评定方法

毕业设计成绩由过程性考核成绩（40%）和结果考核成绩（60%）组成。其中过程性考核包括开题答辩、中期答辩、过程评价三个环节，结果考核以答辩环节为主，考核毕业设计说明书（论文）和答辩表现。

五、时间安排

本毕业设计（论文）安排在第8学期，共16周。

六、说明

1. 答辩小组由4名以上教师组成。
2. 评阅人由中级及中级以上职称的教师担任。

七、其他说明

此大纲在教学执行过程中根据具体情况适当调整。