

## 【智能信息处理技术】

## 【Intelligent Information Processing Technology】

## 一、基本信息

课程代码：【2050242】

课程学分：【4】

面向专业：【网络工程】等

课程性质：【系级选修课◎】

开课院系：【信息技术学院 网络工程系】

使用教材：

教材【《Python 数据分析与机器学习（微课视频版）》杨年华 清华大学出版社 2023.01】

参考书目【《大数据分析：Python 爬虫、数据清洗和数据可视化》黄源 清华大学出版社 2022.09】

【《大数据与人工智能技术》吕云翔清华大学出版社 2022.09】

【《Python 数据分析与可视化案例教程》余本国 人民邮电出版社 2022.06】

课程网站网址：【i1.gench.edu.cn/BB】等

先修课程：无

## 二、课程简介

本课程采用理论与实践相结合的教学方式，通过学习 Python 语法基础、numpy 数据处理、matplotlib 数据可视化技术、pandas 数据预处理与分析技术、sklearn 机器学习等内容，培养学生数据处理的能力，掌握机器学习的基本流程，并能够利用机器学习的方法挖掘数据中的有用价值，使学生更好地适应大数据和人工智能时代的工作与学习。

## 三、选课建议

本课程作为信息技术学院网络工程专业的专业选修课程，建议学生在学习了程序设计基础、数据结构等课程基础上在高年级选修。

## 四、课程与专业毕业要求的关联性

毕业要求	指标点	关联
毕业要求 1 工程知识：能够将数学、自然科学、工程基础和专业知识用于解决复杂网络工程问题	指标点 1-1 能够将数学与物理的知识、方法与思想，用于网络工程过程中所需要的抽象思维与逻辑分析。	
	指标点 1-2 能够将算法、数据结构与程序设计等知识与方法，用于进行计算思维，用于基本算法问题的分析、设计与实现。	
	指标点 1-3 能够将数字逻辑电路、计算机组成结构、操作系统、数据库系统等知识与方法用于进行计算机系统工作原理或机理的分析与理解。	●
	指标点 1-4	

	能够将网络体系结构、网络协议、网络互联等网络工程基础知识，用于复杂网络系统的工作原理或机理的分析与理解。	
	<p>指标点 1-5</p> <p>能够将网络互联、信息安全、网络测试、网络编程、网络规划与设计等网络工程专业知识，用于进行网络系统的规划、设计、部署、开发、测试与运维。</p>	
<p>毕业要求 2</p> <p>问题分析：能够应用数学、自然科学和工程科学的基本原理，识别、表达、并通过文献研究分析复杂网络工程问题，以获得有效结论</p>	<p>指标点 2-1</p> <p>能够应用数学、物理和工程科学的基本原理，进行复杂网络工程问题的识别、分析与表达。</p>	
	<p>指标点 2-2</p> <p>能够通过文献与信息资源的有效收集与研读，获得可用的知识、技术或方法，辅助进行复杂网络工程问题的研究、分析与解决。</p>	
<p>毕业要求 3</p> <p>设计解决方案：能够设计针对复杂网络工程问题的解决方案，包括满足特定需求的网络系统设计方案、网络工程实施方案和网络测试方案，并能够在设计环节中体现创新意识，考虑社会、健康、安全、法律、文化以及环境等因素</p>	<p>指标点 3-1</p> <p>能够针对复杂网络工程问题，通过有效的需求调查与研究、技术分析与设计、设备与产品选型，规划与设计满足特定需求的网络系统解决方案，并具有对解决方案进行部署与实施、测试与验证的能力。</p>	●
	<p>指标点 3-2</p> <p>针对复杂的网络工程问题，能够关注社会、健康、安全、法律、文化以及环境等因素的影响，并在解决方案的设计与实施环节中予以考虑。</p>	
	<p>指标点 3-3</p> <p>能够在网络系统规划、设计、部署、开发、运维和测试等过程中，就多元需求、目标与影响因素，综合运用网络工程和相关学科或领域的知识、技术与方法，通过系统性的分析与研判、合理的规划与设计、有效的统筹与协调，给出独到的或具有一定创新性的解决思路、方法或方法。</p>	
<p>毕业要求 4</p> <p>研究：能够基于科学原理并采用科学方法对复杂网络工程问题进行研究，包括设计实验、分析与解释数据、并通过信息综合得到有效的结论</p>	<p>指标点 4-1</p> <p>能够基于工程科学与网络系统工作原理，结合网络工程专业基础知识与方法，就复杂网络系统中涉及的局部性功能或性能问题进行研究，设计相关的实验方案，并对实施结果或数据进行有效分析和合理解释。</p>	
	<p>指标点 4-2</p> <p>能够基于工程科学与网络系统工作原理，运用网络工程技术领域的专门知识与方法，就复杂网络系统中涉及的领域性功能或性能问题进行研究，设计相关的实验方案，并对结果或数据进行有效分析与合理解释。</p>	
	<p>指标点 4-3</p> <p>能够基于工程科学与网络系统工作原理，综合运用网络工程多技术领域的知识与方法，进行复杂网络工程系统中所涉及的综合性功能或性能问题进行研究，设计相关的实验方案，对实施结果或数据进行分析，并通过信息综合得到合理有效的结论。</p>	
<p>毕业要求 5</p> <p>使用现代工具：能够针对复杂网络工程问题，开发、选择与使用恰当的技术、资源、现代工程工具和信息技术工</p>	<p>指标点 5-1</p> <p>能够选择和利用基本的信息技术工具和网络工程工具，结合适当的技术与资源，用于复杂网络工程问题的预测与分析。</p>	
	<p>指标点 5-2</p> <p>能够针对复杂网络工程问题，选择恰当的虚拟仿真工具或方法，对网络系统或其解决方案进行必要的模拟与预测，并能够理解仿真模拟系统与真实</p>	

具, 包括对复杂工程问题的预测与模拟, 并能够理解其局限性	系统之间的差异。	
毕业要求 6 工程与社会: 能够基于网络工程相关背景知识进行合理分析, 评价网络工程实践和复杂网络工程问题解决方案对社会、健康、安全、法律以及文化的影响, 并理解应承担的责任	指标点 6-1 具有网络工程实习和社会实践的经历。	
	指标点 6-2 能够了解互联网和信息安全相关的法律、法规, 并理解网络系统或网络工程实践对于社会、健康、安全、法律以及文化的可能影响。	
	指标点 6-3 能够基于网络工程专业知识, 结合相关的应用背景知识, 评价网络系统解决方案或网络工程实践对于社会、健康、安全、法律以及文化的可能影响, 并理解应承担的责任。	
毕业要求 7 环境和可持续发展: 能够理解和评价针对复杂网络工程问题的工程实践对环境、社会可持续发展的影响	指标点 7-1 具有环境保护的自觉和可持续发展意识, 了解环境保护与社会可持续发展相关的方针与政策、法律与法规。	
	指标点 7-2 能够理解和评价针对复杂网络工程问题的网络系统解决方案或网络工程实践对环境、社会可持续发展的影响。	
毕业要求 8 职业规范: 具有人文社会科学素养、社会责任感, 能够在网络工程实践中理解并遵守工程职业道德和规范, 履行责任	指标点 8-1 具有基本的人文社会科学素养。	
	指标点 8-2 能够运用哲学的知识与方法认识、分析社会现象, 具有思辨能力与批判精神。	
	指标点 8-3 具有良好的社会公德与社会责任感, 富于爱心, 懂得感恩。	
	指标点 8-4 能够理解并遵守网络工程的相关职业道德和规范, 能够在网络工程实践中承担质量、安全、服务和环保等方面的社会责任。	
毕业要求 9 个人和团队: 能够在多学科背景下的团队中承担个体、团队成员以及负责人的角色	指标点 9-1 具备良好的身体素质和明确的个体意识, 具有在团队框架下承担个体责任、发挥个体作用的能力。	
	指标点 9-2 具备良好的团队意识、团队合作与沟通、团队协调或组织能力, 能够在多学科背景下的团队中根据需要承担成员或负责人的角色。	
毕业要求 10 沟通: 能够就复杂网络工程问题与业界同行及社会公众进行有效沟通和交流, 包括撰写报告和设计文稿、陈述发言、清晰表达或回应指令, 并具备一定的国际视野, 能够在跨文化背景下进行沟通和交流	指标点 10-1 具备沟通交流的基本技巧与能力, 良好的口头与书面表达能力, 有效表达自己思想与意愿的能力, 倾听与理解他人需求和意愿的能力, 适应工作与人机环境变化的能力。	
	指标点 10-2 能够依照相关的工程标准或行业规范, 进行网络工程相关技术问题及文档(如需求分析报告、系统设计方案、系统实施方案等)的撰写与交流表达。	
	指标点 10-3 具备一门外国语言的基本听、说、读、写、译能力, 能够阅读网络工程专业领域的外文资料, 具备一定的国际视野, 对专业领域相关的新技术具有敏感性。	

毕业要求 11 项目管理:理解并掌握 工程管理原理与经济 决策方法,并能在多学 科环境中应用	指标点 11-1 具有基本的成本管理意识,在设计针对复杂网络工程问题的解决方案时, 能够对经济与成本因素加以必要的考量。	
	指标点 11-2 能够理解 IT 项目管理的知识、原理与方法,并在多学科背景的网络工程 项目或实践中进行应用。	
毕业要求 12 终身学习:具有自主学 习和终身学习的意识, 有不断学习和适应发 展的能力	指标点 12-1 具有持续更新知识、提升能力与素质的终身学习意识,养成自主学习的习 惯。	
	指标点 12-2 具有跟踪网络技术发展、增强自我竞争力、适应持续发展所需的自主学习 能力与自我挑战能力。	

## 五、课程目标/课程预期学习成果

序号	课程预期 学习成果	课程目标 (细化的预期学习成果)	教与学 方式	评价 方式
1	指标点 1-3: 能够将数字逻辑电路、计算机组成结构、操作系统、数据库系统等知识与方法用于进行计算机系统工作原理或机理的分析与理解。	1. 学会信息处理工具 python,具有能够处理信息的能力,包括数据的导入导出,数据的抽取、拆分、合并、匹配,以及缺失数据的处理。	讲课 实验 课外阅读	实验 课堂表现
2	指标点 3-1: 能够针对复杂网络工程问题,通过有效的需求调查与研究、技术分析与设计、设备与产品选型,规划与设计满足特定需求的网络系统解决方案,并具有对解决方案进行部署与实施、测试与验证的能力。	2.学会数据可视化的相关技术,具有熟练使用 python 进行数据可视化的能力。	讲课 实验	实验 课堂表现 课程报告
		3.学会常用的数据分析方法,具有熟练使用 python 对数据进行独立分析的能力。	讲课 实验	实验 课堂表现 课程报告

## 六、课程内容

### 第1单元 Python语言与开发环境概述 (4课时)

知道Python语言的特点、下载与安装、交互与文件两种使用方式。掌握代码的风格、模块的概念及其导入方式、帮助的使用、Anaconda环境的安装与使用。

**重点:** 常用Python内置函数和模块的导入与使用以及代码编写规范。

**理论课时数:** 2

**实践课时数:** 2

### 第2单元 Python语言基础 (8课时)

知道控制台的输入与输出、标识符、变量、赋值语句，常用数据类型、运算符、表达式，分支结构，循环结构，常用组合类型（列表、元组、字符串、字典、集合），推导式、用于组合类型的常用内置函数、正则表达式。

**重点：**列表、元组、字典、集合的访问、切片、计算等；if选择、for循环、while循环。

**难点：**range对象在循环中的使用，正则表达式。

**理论课时数：**4

**实践课时数：**4

### **第3单元 函数（4课时）**

理解函数的定义、调用、函数的返回值、位置参数与关键参数，默认参数、个数可变的参数、参数与返回值类型注解，lambda表达式、函数式编程的常用类与函数。

**重点：**函数的定义方式；类的定义与使用。

**难点：**lambda表达式声明匿名函数。

**理论课时数：**2

**实践课时数：**2

### **第4单元 自定义类与对象（4课时）**

理解Python中的对象与方法、类的定义与对象的创建，类的继承。

**重点：**类与对象的定义。

**难点：**类的继承。

**理论课时数：**2

**实践课时数：**2

### **第5单元 numpy数据处理基础（6课时）**

知道numpy数据结构、数据的准备、数组的索引与切片，改变数组的形状、对角线上元素的替换、插入维度、数组的基本运算、数组的排序、数组的组合、数组的分割，随机打乱数组中的元素顺序、多维数组的展开、其他适用于数组的函数与对象、利用numpy进行统计分析、数组在其他文件中的存取。

**重点：**numpy数组的定义与操作。

**难点：**利用numpy进行统计分析。

**理论课时数：**2

**实践课时数：**4

### **第6单元 matplotlib数据可视化基础（6课时）**

知道绘制基本图形、绘制多轴图，坐标轴的刻度标签、主次刻度、网格设置、移动坐标轴、文字说明和注释、显示图片、日期作为横坐标，绘制横线与竖线、绘制其他二维图表、绘制三维图表。

**重点：**饼图、散点图、折线图以及柱图的画法。

**难点：**三维图表的绘制方法。

**理论课时数：**2

**实践课时数：**4

### **第7单元 pandas数据处理与分析（8课时）**

理解数据结构与基本操作、文件与数据库中存取DataFrame对象，常用函数与方法，DataFrame

数据清洗与处理、时间处理，移动数据与时间索引、统计分析、pandas中的绘图方法。

**重点：**Dataframe（数据框）的使用（增、删、改、查）。

**难点：**移动数据与时间索引、统计分析、pandas中的绘图方法。

**理论课时数：**4

**实践课时数：**4

### **第8单元 机器学习方法概述与数据加载（4课时）**

知道机器学习概述、sklearn简介与安装，加载数据、划分训练集与测试集，sklearn中机器学习的基本步骤示例、sklearn编程接口风格。

**重点：**加载数据、划分训练集与测试集。

**理论课时数：**2

**实践课时数：**2

### **第9单元 数据预处理（4课时）**

知道特征的离散化，识别与处理异常值、特征的Min-Max缩放，特征值的标准化、特征值的鲁棒缩放、无序分类数据的热编码、有序分类数据的编码、每个样本特征值的正则化。

**重点：**识别与处理异常值、特征的Min-Max缩放。

**难点：**特征值的标准化与正则化。

**理论课时数：**2

**实践课时数：**2

### **第10单元 模型评估与轨道（4课时）**

理解泛化、过拟合、欠拟合的概念；模型评估指标，交叉验证，轨道的创建与使用。

**重点：**泛化、过拟合、欠拟合的概念。

**难点：**泛化、过拟合、欠拟合的概念。

**理论课时数：**2

**实践课时数：**2

### **第11单元 有监督学习之分类与回归（4课时）**

理解分类与回归概述、线性回归，逻辑回归与岭回归实现线性分类、支持向量机用于分类和回归，朴素贝叶斯分类、决策树用于分类和回归。

**重点：**分类与回归的概念。

**难点：**朴素贝叶斯分类。

**理论课时数：**2

**实践课时数：**2

### **第12单元 集成学习（2课时）**

知道投票法集成、bagging/pasting集成，提升法集成、堆叠法集成（2课时）。

**重点：**集成学习的概念与分类。

**理论课时数：**2

### **第13单元 无监督学习之聚类与降维（4课时）**

知道用k-均值基于相似性聚类、层次聚类、基于密度的聚类，聚类性能评估、无监督的降维。

**重点：**用k-均值基于相似性聚类。

**难点：**无监督的降维。

理论课时数：2

实践课时数：2

#### 第14单元 超参数调优与模型选择（2课时）

知道基于循环语句的网格搜索、划分验证集避免过拟合、带交叉验证的网格搜索、带交叉验证的随机搜索，搜索多个不同特征的空间、对轨道中的超参数进行搜索、搜索算法和超参数。

重点：超参数调优的基本方法。

难点：模型选择。

理论课时数：2

### 七、课内实验名称及基本要求

序号	实验名称	主要内容	实验时数	实验类型	备注
1	数据结构与控制流程	超市打印小票	2	设计型	
2	Numpy 的数组方法的使用	图片翻转裁剪压缩处理	6	综合型	
3	正则表达式	对输入的身份证、手机、Email 等信息格式进行验证	4	设计型	
4	Matplotlib 数据可视化	根据给定数据，绘制南丁格尔玫瑰图	4	设计型	
5	Pandas 数据处理	输出符合条件的属性内容	6	综合型	
6	综合数据处理与分析	股票数据下载与统计分析	10	综合型	

### 八、评价方式与成绩

总评构成 (X)	评价方式	占比
X1	课程报告	40%
X2	实验	30%
X3	考勤与课堂表现	30%

撰写人：李洋

系主任审核签名：王瑞

审核时间：2023 年 2 月