

# 《专业技能实训 III》

## 课程教学大纲

课程编号：

学 分：3 学分

课 时：54 课时 先修课程：Python 程序设计基础

适用专业：计算机类相关专业

### 一、课程的性质与目标

《专业技能实训 III》综合实训课程是一门面向有 Python 基础人群的进阶课程，该课程通过理论与实践结合的模式，循序渐进讲解了 Python 数据分析的相关知识，包括科学计算库 NumPy、数据分析库 pandas、数据预处理、分组与聚合、数据可视化、文本数据采集与分析、机器学习等内容。通过本课程的学习，可以让学生掌握 Python 数据分析等综合编程核心知识。

### 二、课程设计思路和教学要求

课程设计理念：本课程注重知识的全面性和系统性，重视学习方法的传授。本课程配置了丰富的章节案例，可以强化学生对知识的运用，力求让学生打下一个牢靠的基本功，具备举一反三和独立解决实际问题的能力。

课程思路：本课程遵循知识点先易后难的顺序安排章节，采用“理论知识+要点分析+示例演示+章节案例”的讲解形式，既有理论知识的介绍，又提供了充足的任务实践，保证学生在理解核心知识的前提下可以真正地学有所得。在讲解晦涩难懂的知识点时，会有形象的场景举例或者示意图，帮助学生更好地理解与吸收知识。本课程最后还开发了一个实用性强的综合项目，通过项目将所学知识串连起来运用，培养学生分析问题和解决问题的能力。

操作系统：Windows 7 以上

开发工具：Anaconda3 2022.10(可选)、Python IDLE

### 三、课程的主要内容及基本要求

#### 第 1 章 数据分析概述

学习单元	第 1 章 数据分析概述	课时	2	
学习目标	1. 了解数据分析产生的背景，能够简述数据分析产生的背景 2. 了解数据分析的概念及层次，能够说出数据分析的概念以及数据分析的 4 个层次 3. 了解数据分析的应用领域，能够列举至少 3 个数据分析的应用领域 4. 熟悉数据分析的流程，能够归纳数据分析的基本流程 5. 了解 Python 做数据分析的优势，能够说出 Python 在数据分析方面有哪些优势 6. 了解 Anaconda 工具，能够说出 Anaconda 工具的特点 7. 掌握 Anaconda 的安装与使用，能够独立在计算机中安装 Anaconda 工具，并通过 Anaconda 工具安装、更新、卸载包			
学习内容	知识点	掌握程度	重点	难点
	数据分析产生的背景	了解		
	什么是数据分析	了解		
	数据分析的应用场景	了解		
	数据分析的流程	熟悉		
	为什么选择 Python 做数据分析	熟悉		
	Anaconda 概述	了解		
	Anaconda 的安装	掌握	√	
	通过 Anaconda 管理包	掌握		√
	启动 Anaconda 自带的 Jupyter Notebook	掌握		
	Jupyter Notebook 界面详解	熟悉		
	Jupyter Notebook 的基本使用	掌握		√

## 第 2 章 科学计算库 NumPy

学习单元	第 2 章 科学计算库 NumPy	课时	6
学习目标	1. 了解 NumPy 数组的相关概念，能够说出什么是 NumPy 数组、维度、轴和秩 2. 熟悉 NumPy 数组的属性，能够归纳 ndim 和 shape 属性的作用 3. 掌握数据的创建方式，能够灵活创建一维数组和二维数组 4. 掌握数组的数据类型，能够查看与转换数组的数据类型 5. 掌握数组的索引和切片操作，能够灵活地通过不同形式的索引获取数组元素		

	<p>6. 掌握数组的算术运算，能够实现数组与数组或数组与标量的算术运算</p> <p>7. 掌握数组的通用函数，能够熟练地使用一元通用函数和二元通用函数进行数学运算</p> <p>8. 掌握数组的重塑操作，能够通过 reshape()方法实现数组的重塑操作</p> <p>9. 掌握数组的转置操作，能够通过 T 属性或 transpose()方法实现数组的转置操作</p> <p>10. 掌握数组的其他操作，能够实现数组的条件逻辑、统计运算、排序操作</p> <p>11. 熟悉线性代数模块，能够通过 linalg 模块的功能完成矩阵操作</p> <p>12. 掌握随机数模块，能够通过 random 模块的功能生成包含随机数的数组</p>			
<b>学习内容</b>	<b>知识点</b>	<b>掌握程度</b>	<b>重点</b>	<b>难点</b>
	NumPy 数组的相关概念	了解		
	NumPy 数组的属性	熟悉		
	创建数组	掌握	√	
	查看数据类型	掌握		
	转换数据类型	掌握		
	数组的索引方式	熟悉		
	整数索引和切片	掌握	√	
	花式索引	掌握	√	
	布尔索引	掌握	√	
	形状相同的数组间的算术运算	掌握	√	
	形状不同的数组间的算术运算	掌握	√	√
	数组与标量的算术运算	掌握		
	通用函数	掌握		
	数组的重塑	掌握		
	数组的转置	掌握		√
	条件逻辑	掌握		
	统计运算	掌握		
	数组元素排序	掌握		
	检索数组元素是否满足条件	掌握		
	查找数组的唯一元素	掌握		
	判断元素是否在其他数组中	掌握		
	线性代数模块	熟悉		√
随机数模块	掌握			
案例：计算股票收益率和波动率	掌握			

### 第3章 数据分析库 pandas 基础

学习单元	第3章 数据分析库 pandas 基础	课时	6	
学习目标	1. 掌握 Series 的结构和创建方式，能够通过 Series 类的构造方法创建 Series 类的对象			
	2. 掌握 DataFrame 的结构和创建方式，能够通过 DataFrame 类的构造方法创建 DataFrame 类的对象			
	3. 熟悉索引对象的类型和特点，能够归纳索引对象的类型和特点			
	4. 掌握重置索引的方式，能够通过 <code>reindex()</code> 方法重置 Series 或 DataFrame 类对象的索引			
	5. 掌握索引与切片的基本用法，能够通过索引或切片获取 Series 或 DataFrame 的数据			
	6. 掌握 <code>loc</code> 和 <code>iloc</code> 属性的基本用法，能够使用 <code>loc</code> 和 <code>iloc</code> 属性获取 Series 或 DataFrame 的数据			
	7. 掌握读写数据的方式，能够熟练地读取网页表格的数据，以及读写 CSV 文件、TXT 文件、Excel 文件、数据库			
	8. 掌握数据的排序方式，能够根据索引或值对 Series 或 DataFrame 进行排序			
	9. 掌握 pandas 的算术运算，能够通过算术运算符或算术方法对 DataFrame 中的数据进行算术运算			
	10. 掌握 pandas 的统计计算，能够通过统计方法对 DataFrame 中的数据进行统计计算			
	11. 掌握 pandas 的统计描述，能够通过 <code>describe()</code> 方法描述 DataFrame 类的对象的统计指标			
	12. 掌握分层索引的相关操作，能够熟练地创建有分层索引的 DataFrame，并使用分层索引获取数据			
	学习内容	知识点	掌握程度	重点
Series		掌握	√	
DataFrame		掌握	√	
索引对象		了解		
重置索引		掌握		
通过索引和切片获取数据		掌握	√	
通过 <code>loc</code> 和 <code>iloc</code> 属性获取数据		掌握	√	
读写 CSV 和 TXT 文件的数据		掌握	√	
读写 Excel 文件的数据		掌握	√	
读取网页表格的数据		掌握		

	读写数据库	掌握		√
	按索引排序	掌握		
	按值排序	掌握		
	算术运算与数据对齐	掌握		
	统计计算	掌握		
	统计描述	掌握		
	创建分层索引	掌握		
	创建有分层索引的对象	掌握		
	使用分层索引获取数据	掌握		√
	交换索引层级的顺序	熟悉		√
	分层索引排序	熟悉		√
	案例：陕西高考分数线统计分析	掌握		

## 第 4 章 数据预处理

学习单元	第 4 章 数据预处理	课时	8
学习目标	<ol style="list-style-type: none"> <li>1. 掌握缺失值的检测方式，能够通过 <code>isnull()</code>和 <code>notnull()</code>函数检测数据中是否存在缺失值</li> <li>2. 掌握缺失值的处理方式，能够通过 <code>dropna()</code>或 <code>fillna()</code>方法删除缺失值或填充缺失值</li> <li>3. 掌握重复值的检测方式，能够通过 <code>duplicated()</code>方法检测数据中是否存在重复值</li> <li>4. 掌握重复值的处理方式，能够通过 <code>drop_duplicates()</code>方法删除重复值</li> <li>5. 熟悉异常值的检测方式，能够通过 <math>3\sigma</math> 原则和箱形图检测数据中是否存在异常值</li> <li>6. 掌握异常值的处理方式，能够通过 <code>replace()</code>方法替换数据中的异常值</li> <li>7. 熟悉数据类型的转换方式，能够通过 <code>astype()</code>方法或 <code>to_numeric()</code>函数转换数据类型</li> <li>8. 掌握数据合并的相关操作，能够根据需求选择适合的方案实现数据合并的操作</li> <li>9. 掌握数据重塑的相关操作，能够根据需求选择适合的方案实现数据重塑的相关操作</li> <li>10. 掌握数据转换的相关操作，能够根据需求选择适合的方案实现数据转换的相关操作</li> </ol>		

学习内容	知识点	掌握程度	重点	难点
	缺失值的检测	掌握	√	
	缺失值的处理	掌握	√	
	重复值的检测	掌握	√	
	重复值的处理	掌握	√	
	异常值的检测	熟悉		√
	异常值的处理	掌握		
	转换数据类型	掌握		
	堆叠合并	掌握	√	
	主键合并	掌握	√	√
	根据索引合并	掌握		
	合并重叠数据	熟悉		
	重塑分层索引	掌握		
	轴向旋转	掌握		
	面元划分	掌握		√
	哑变量处理	熟悉		√
案例：预处理二手房数据	掌握			

## 第 5 章 数据聚合和分组运算

学习单元	第 5 章 数据聚合和分组运算	课时	4
学习目标	1. 了解分组与聚合的原理，能够说出分组与聚合的原理 2. 掌握分组方法的使用，能够通过 <code>groupby()</code> 方法按照不同的拆分标准对数据进行分组 3. 掌握分组信息的查看方式，能够通过多种方式查看分组的信息 4. 熟悉内置统计方法的使用，能够通过统计方法聚合数据 5. 掌握 <code>agg()</code> 方法的使用，能够通过 <code>agg()</code> 方法聚合数据 6. 掌握 <code>transform()</code> 方法的使用，能够通过 <code>transform()</code> 方法转换数据 7. 掌握 <code>apply()</code> 方法的使用，能够通过 <code>apply()</code> 方法聚合数据		
学习内容	知识点	掌握程度	重点
	分组与聚合的原理	了解	
	通过 <code>groupby()</code> 对数据进行分组	掌握	√
	查看分组信息	掌握	
通过统计方法聚合数据	熟悉		

	通过 agg()聚合数据	掌握	√	√
	数据转换	掌握	√	√
	数据应用	掌握		
	案例：篮球运动员信息分析	掌握		

## 第 6 章 数据可视化

学习单元	第 6 章 数据可视化		课时	10
<b>学习目标</b>	1. 了解数据可视化，能够说出可视化的概念以及意义 2. 熟悉常见的图表类型，能够说出图表的特点以及适用场景 3. 熟悉图表的基本组成元素，能够说出每个组成元素的用途 4. 掌握 Matplotlib 库的基本使用，能够使用 Matplotlib 库绘制常见的图表 5. 掌握 Seaborn 库的基本使用，能够使用 Seaborn 库绘制常见的图表 6. 掌握 Pyecharts 库的基本使用，能够使用 Pyecharts 库绘制常见的图表			
<b>学习内容</b>	<b>知识点</b>	<b>掌握程度</b>	<b>重点</b>	<b>难点</b>
	什么是数据可视化	了解		
	常见的图表类型	了解		
	图表的辅助元素	熟悉		
	使用 Matplotlib 绘制折线图	掌握	√	
	使用 Matplotlib 绘制柱形图	掌握	√	
	使用 Matplotlib 绘制直方图	掌握	√	
	使用 Matplotlib 绘制散点图	掌握	√	
	可视化数据的分布	掌握		
	用分类数据绘图	掌握		√
	Pyecharts 简介	熟悉		
	使用 Pyecharts 绘制柱形图	掌握	√	√
	使用 Pyecharts 绘制词云图	掌握	√	√
	使用 Pyecharts 绘制气泡图	掌握		
	使用 Pyecharts 绘制圆环图	掌握		
案例：电影数据分析	掌握			

## 第7章 文本数据分析

学习单元	第7章 文本数据分析			课时	12
<b>学习目标</b>	1. 了解 NLTK 与 jieba 库，能够说明 NLTK 与 jieba 库的用途 2. 掌握 NLTK 语料库的安装，能够在计算机中成功安装 NLTK 语料库 3. 熟悉文本预处理的流程，能够归纳出文本预处理的基本流程 4. 掌握分词的方式，能够通过 NLTK 与 jieba 库对文本进行分词 5. 掌握词性标注的方式，能够通过 pos_tag()函数对英文文本进行分词标注 6. 掌握词形归一化操作，能够通过 nltk.stem 模块实现词形归一化的操作 7. 掌握删除停用词操作，能够通过 stopwords 模块实现删除停用词的操作 8. 熟悉文本情感分析，能够通过多种方式实现简单的文本情感分析 9. 熟悉文本相似度，可以结合 NLTK 与余弦相似度实现简单的文本相似度分析 10. 熟悉文本分类，可以结合 NLTK 与朴素贝叶斯算法实现简单的文本分类分析				
<b>学习内容</b>	<b>知识点</b>	<b>掌握程度</b>	<b>重点</b>	<b>难点</b>	
	认识 NLTK 与 jieba	了解			
	安装 jieba 和 NLTK 语料库	熟悉			
	文本预处理基本流程	熟悉			
	分词	掌握	√		
	词性标注	掌握	√		
	词形归一化	掌握	√		
	删除停用词	掌握	√		
	文本情感分析	掌握		√	
	文本相似度	熟悉		√	
	文本分类	熟悉		√	
	案例：商品评论分析	掌握			

## 第8章 机器学习入门

学习单元	第8章 机器学习入门			课时	4
<b>学习目标</b>	1. 了解机器学习，能够说出什么是机器学习 2. 熟悉机器学习的基本概念，能够归纳机器学习涉及的基本概念 3. 了解机器学习算法的分类，能够区分监督学习、无监督学习和强化学习				

	4. 了解机器学习解决问题的流程，能够说出机器学习解决问题的流程 5. 熟悉机器学习库 scikit-learn，能够列举至少 3 个 scikit-learn 的模块和数据集 6. 掌握 KNN 算法，能够使用 scikit-learn 的 API 实现 KNN 算法			
<b>学习内容</b>	<b>知识点</b>	<b>掌握程度</b>	<b>重点</b>	<b>难点</b>
	什么是机器学习	了解		
	机器学习的基本概念	熟悉		
	机器学习算法的分类	了解		
	机器学习解决问题的流程	了解		
	认识机器学习库 scikit-learn	熟悉		
	KNN 算法的思想	掌握	√	
	使用 sklearn 实现 KNN 算法	掌握	√	
	超参数	掌握	√	√
	网格搜索与交叉验证	掌握		√
	归一化	掌握		√
	使用 sklearn 实现归一化	掌握	√	
	案例：预测签到位置	掌握		

<b>学习单元</b>	<b>第 9 章 期末总结</b>	<b>课时</b>	<b>2</b>
<b>学习目标</b>	期末总结，完成考查任务		
<b>学习内容</b>	<b>知识点</b>	<b>掌握程度</b>	<b>重点</b>
	期末总结，完成考查任务	——	

## 四、课时分配

章目	学时	合计
第1章 数据分析概述	2	2
第2章 科学计算库 NumPy	6	6
第3章 数据分析库 pandas 基础	6	6
第4章 数据预处理	8	8
第5章 数据聚合和分组运算	4	4
第6章 数据可视化	10	10
第7章 文本数据分析	12	12
第8章 机器学习入门	4	4
期末任务讲解、总结、验收	2	2
合计	54	54

## 五、考核模式与成绩评定办法

本课程为考查课程，成绩由平时成绩（50%）和期末考查（50%）组成，其中，平时成绩包括出勤（20%）、上机实训作业（30%）。

## 六、选用教材和主要参考书

本大纲是根据黑马程序员编著的教材《Python 数据分析与应用：从数据获取到可视化（第2版）》所设计。

## 七、大纲说明

本课程的授课模式为：课堂讲授+上机，其中，课堂主要采用多媒体的方式进行授课，并且会通过测试题阶段测试学生的掌握程度；上机主要是编写程序，要求学生动手完成指定的程序设计或验证。