

# 实验七 科学计算与数据分析基础

## 一、实验目的

- 1. 掌握 numpy 库的基本操作
- 2. 掌握 pandas 库的基本操作
- 3. 掌握使用 matplotlib 库绘制图形的方法

## 二、实验内容

- 1. 创建一个 3 行 3 列的 ndarray 数组，数组元素为 1、2、3、...、9 这 9 个数，编写程序，计算输出其所有的元素的和、每行的平均值以及每列的平均值。

1	2	3
4	5	6
7	8	9

- 2. 正则化一个 5 行 5 列的随机矩阵，并输出。正则的概念是假设  $a$  是矩阵中的一个元素， $\max$  和  $\min$  分别是矩阵元素的最大值和最小值，则正则化后  $a = (a - \min) / (\max - \min)$ 。
- 3. 请根据下图创建 Excel 文件 stu.xlsx，编写程序，
  - (1) 输出所有出现不及格科目的同学的信息；
  - (2) 统计输出每个班级的各门课的平均分。

	A	B	C	D	E
1	姓名	班级	语文成绩	数学成绩	英语成绩
2	刘德军	19-1	67	56	76
3	方丽丽	19-2	77	82	87
4	张琳琳	19-1	57	79	88
5	王威	19-2	82	67	65
6	李建军	19-2	86	99	90
7	赵欧	19-1	90	84	98

- 4. 设计程序，绘制函数  $f(x) = 3x^2 + 7x - 9, x \in [0,8]$  的图形，图标题设置为“我是图标题”，x 轴标签设置为“x 的取值”，y 轴标签设置为“y 的值”，并添加图例“我是图例”，添加图文字描述“我是曲线”。
- 5. 设计程序，按下列要求绘制有关函数  $f(x) = x^4 + 3x^3 + x^2 + 4x, x \in [-5,5]$  的图形。
  - (1) 绘制 2 行 2 列的 4 个子图；
  - (2) 第 1 行第 1 列的子图，使用红色实线绘制  $f(x)$  的图形；
  - (3) 第 1 行第 2 列的子图，使用蓝色虚线绘制  $f(x)$  的一阶导数  $f'(x)$  的图形；
  - (4) 第 2 行第 1 列的子图，使用绿色圆点绘制  $f(x)$  的二阶导数  $f''(x)$  的图形；
  - (5) 第 2 行第 2 列的子图，使用黄色三角绘制  $f(x)$  的三阶导数  $f'''(x)$  的图形。

# 实验八 网络爬虫基础

## 一、实验目的

1. 学会使用 requests 库连接 HTML 网页
2. 学会使用 BeautifulSoup 库解析 HTML 网页标签树

## 二、实验内容

1. 这是一个简单的 HTML 页面,请输入以下 HTML 文件内容,在 D 盘根目录下保存为“1.html”,完成后面的计算要求。

```
<html>
  <head>
    <title>Smile</title>
  </head>
  <body>
    <p id="Hi">Python</p>
    <p id="Hello">HTML</p>
  </body>
</html>
```

- (1) 打印 head 标签的内容;
  - (2) 打印 body 标签的内容;
  - (3) 打印 id 为“Hi”的标签对象。
2. 中国天气网,徐州天气网址为 <http://www.weather.com.cn/weather/101190801.shtml>,设计程序完成以下功能:
    - (1) 设计函数 get\_content(url), url 为抓取数据网页地址,函数返回值是网页文本;
    - (2) 设计函数 get\_data(html\_text), html\_text 为 get\_content(url) 的返回值,函数返回值是需要爬取的徐州 7 天的天气的日期、天气情况、最高温度和最低温度;
    - (3) 设计函数 write\_data(data, fname), data 为 get\_data(html\_text) 的返回值, fname 是抓取数据写入的文件路径,函数没有返回值,功能是将 html\_text 的内容写入文件 fname;
    - (4) 设计主函数,调用以上三个函数,从网页爬取徐州 7 天的天气的日期、天气情况、最高温度和最低温度,并写入 D 盘根目录下的文本文件 8-2.txt。提示:需要字段都在属性 id=“7d”的标签“div”的 ul 中。日期在每个 li 中 h1 中,天气状况在每个 li 的第一个 p 标签内,最高温度和最低温度在每个 li 的 span 和 i 标签中。
  3. 设计程序,爬取豆瓣电影中排名前十位的电影。
  4. 设计程序,爬取豆瓣读书中对某本书的前 10 条短评内容并计算星级评定分数的平均值(保留两位小数)。

设计程序,爬取中国矿业大学教务部通知公告 (<http://jwb.cumt.edu.cn/tzgg.htm>) 中,最新的 5 条通知公告的标题。