# 第一章

1、填空题

（1）人自然智能、人工智能、机器学习、深度学习之间的关系，它们是（包含）关系，同时是（部分继承）关系

（2）人自然智能、人工智能、机器学习、深度学习的英文缩写分别是（NI）、（AI）、（ML）、（DL）

（3）本书将要学习的Python六大第三方库是（Jupyter）、（Numpy）、（Scipy）、（Pandas）、（Scikit-learn）、（Matplotlib）

（4）基于Python语言的第三方库支持（Windows）、（Linux）、（Mac）操作系统下的使用

（5）数据处理流程从项目角度，它涉及（产生）、（获取）、（规整）、（分析）、（应用）、（反馈）六大环节。

1. 判断题

（1）人工智能研究的主要学科包括了计算机科学、信息论、控制论、自动化、仿生学、生物学、心理学、数理逻辑、语言学、医学和哲学等（ √）

（2）AlphaGo的核心底层技术是DL（√ ）

（3）Scipy库是基础科学计算包，Numpy库是高级科学计算包（╳）

正确答案为：Numpy库是基础科学计算包，Scipy库是高级科学计算包

（4）Jupyter Notebook和Python自带的IDLE都是代码编辑、调试工具（√）

（5）Anaconda包支持不同操作系统的安装和使用，下载时无需考虑版本（╳）

正确答案为：Anaconda包支持不同操作系统的安装和使用，下载时需要根据不同的操作系统选择不同的安装版本。

3、实验题

（1）实验一，安装Anaconda包

❶根据实验计算机的环境，自行下载并安装Anaconda包

❷查看Anaconda包里所包含的第三方库名和版本号，并记录

❸用Jupyter Notebook输出Numpy、Scipy、Pandas、Scikit-learn、Matplotlib版本号，并记录

❹用Spyder编程循环5次输出“你好，三酷猫！”，分五行输出

❺对上述操作过程，形成完整的实验报告（含步骤、代码、输出结果）

实验一答案

参考书1.2.1节 安装准备工作相关内容

# 第二章

1、填空题

（1）Jupyter Notebook是广受科学家、数据工程师、人工智能工程师喜欢的一款（探索）式的代码调试及数据处理工具。

（2）新建最新版本的Python代码文件，需要先选择（Python3）内核。

（3）在Jupyter Notebook上运行代码可以通过点（运行）快捷按钮或直接在单元格里输入（Shift+Enter）快命令来执行。

（4）若想知道代码开发文件存放的具体位置，可以用（%pwd）魔法命令。

（5）若想直接在Jupyter Notebook里执行shell命令用（!!）魔法命令。

1. 判断题

（1）Jupyter Notebook是一款探索式的代码编辑器，同时具有Word类似的编辑操作风格（√）

（2）修改Jupyter Notebook代码文件名称只能在文件上通过拉下菜单修改（╳）

也可以通过主界面选择未运行的代码文件，然后点“Rename”快捷键修改。

（3）Jupyter Notebook40支持40多种语言的混合代码开发，也支持文本格式的编辑（√）

（4）Jupyter Notebook可以实现在单元格内的40多种语言的混合开发（╳）

只能部分实现，如python和R语言。

（5）Jupyter Notebook单元格里的代码可以通过魔法命令导出到.py文件，也可以通过魔法命令导入.py文件的代码（√）

3、实验题

（1）实验一，代码文件丢失问题的解决

小张在用Jupyter Notebook编程过程，发生系统崩溃问题，重装系统后发现所编写的代码文件全部丢失，请分析造成原因，并提出解决办法，写出解决办法的操作过程，形成实验报告.

1. 问题分析

当系统崩溃重装后发现所编写德代码文件全部丢失，可以确定原先的代码文件存放于C盘目录下，同时也说明了Jupyter Notebook安装完成后，没有把工作路径指向其它盘符存放路径下。

1. 解决办法

为了解决代码文件的安全问题，必须在Jupyter Notebook安装完成后，通过参数配置文件，把工作路径指向其它盘符的路径下。

配置过程详见书2.1.2节 配置Jupyter内容。

配置完成后，启动Jupyter Notebook界面，将产生新的主界面列表，同时下拉选择“python3”核，生成新Notebook页面文件。去指定工作路径下，可以看到对应的扩展名为ipynb的生成文件。

（2）实验二，用魔法命令调试

用python的IDLE（或Spyder）编写一段小程序（任意），然后用魔法命令导入，并执行。记录所有操作过程，形成实验报告。

U2Experiment1.py

print('求100的因数：')

i=1

while(i<=100):

if 100%i==0:

print(i)

i+=1

# 第三章

# 第四章

1、填空题

（1）Numpy库提供了（det()）函数用于计算行列式，行列式数组要求（行）、（列）相等。

（2）Numpy库中矩阵对象由（matrix()）函数建立。

（3）向量与向量积分（内积）、（外积）。内积结果为（数值），外积结果为（矩阵）。内积又叫数量积、点积。

（4）利用Numpy库的（solve()）、（lstsq()）、（tensorsolve()）可以实现线程方程组的求值。

（5）函数eig(a)用于计算方阵的（特征值）和（特征向量）。

1. 判断题

（1）m为行列式，tril(m, k=0)，该函数计算结果主对角线值都为0（╳）

tril(m, k=-1)时，主对角线值都为0。

（2）函数dot()与matmul()函数计算内积方法一样（╳）

不一样，dot()函数支持标量的内积，matmul()仅支持数组之间的内积。

（3）solve(a, b)函数可以求任意方程组的解（╳）。

solve(a, b)只能求行列式的值不为0的解

（4）在Numpy里可以把只有一行或一列的矩阵（或数字数组）看作向量（√）

（5）特征值分解包括了LU分解 、QR分解、SVD分解等（√）。

3、实验题

（1）实验一，求方阵*A*的特征值、特征向量

*A*=

要求先进行是否有值判断，再决定是否求特征值、特征向量。

（2）实验二，三酷猫想绘制围棋格子，已知围棋共有18\*18=324个格子，请用函数计算所有格子的(x,y)坐标。

并以二维数组形式输出元素为（x,y）的棋盘格式坐标。

# 第五章

1、填空题

（1）文本文件主要指以（ASCII）码方式存储的文件，具体内容包括了英文、数字、特殊符号，也包括了（Unicode）码

（2）基本的读写文本文件的函数为（savetxt()）、（loadtxt()），基本的读写二进制文件的函数为（save()）、（load()）

（3）Numpy数据结构包括了（数据结构信息）区和（据存储）区两大部分

（4）广播原理通过（c）语言的快速处理机制，实现不同（形状）的数组之间的算术运算。

（5）（lower()）方法实现字符串里英文字母的小写化，（upper()）方法实现字符串里英文字母的大写化。

2、判断题

（1）文本文件、二进制文件读写函数都可以实现对内存数据对象的读写（√）

可以，通过StringIO() 、BytesIO()对象代替文件名来实现。

（2）memmap()实现对在线数据源资源的读取， DataSource()对象实现对大型磁盘文件内容的读写（╳）

错，DataSource()对象实现对在线数据源资源的读取，memmap()实现对大型磁盘文件内容的读写

（3）Numpy的同一个数组对象可提供不同类型数据对象的存储与处理（╳）

Numpy的同一个数组对象只能提供一种类型的数据对象的存储与处理，如整型的只能处理整数。

（4）副本与视图是使用原数组数据的两种方式。副本产生的数据在内存单独开辟新空间，视图指向的数据与原数组对象属于一个内存空间。（√）

（5）isdecimal()方法仅用于判断字符串中的单个字符是否是十进制数，isdigit()判断十进制、二进制、全角数字是否是数字，isnumeric()判断十进制、全角数字、汉字数字是否是数字。（√）

3、实验题

（1）实验一，广播计算

A= B=

对B进行改造，以符合广播计算规则，然后建立数组，进行A-B运算

（2）实验二，三酷猫文字处理

I am from China. My name is Tom.Alias is “Three cool cat”

I like Programming.

Lift is short ,I select Python.

❶把上述文字保存到文本文件，

❷把文件内容读入数组

❸统计所有的a的个数。

# 第六章

1、填空题

（1）（figure）是绘图框架——画板，（axes）是绘图区域——画纸，（plot()）是画笔。

（2）Matplotlib支持（直角）坐标系、（极）坐标系。

（3）Matplotlib提供了（text()）、（arrow()）、（annotate()）三种常见的标注方法。

（4）（Figure()）可以指定多绘图框架，（subplot()）可以指定多绘图区域。

（5）RGBA格式图片的像素由（红Red）、（绿Green）、（蓝Blue）、（透明度Aplha）四个通道组成，每个通道占（8Bit），一个像素长度为（32Bit）位。

2、判断题

（1）Matplotlib绘制的图都是由二维坐标、三维坐标点构成（√）

（2）默认情况下figure不支持中文显示（√）

（3）Rectangle() 、Circle()、 Ellipse()等函数提供了plot()相似的绘图参数设置功能，如颜色、线形、图标等（√）

（4）plt.imread()的参数format，指定文件扩展名，这意味着Matplotlib支持很多格式的图片读取（╳）

默认情况下只支持PNG格式的图片读取，要支持其它格式图片的读取，需要额外安装PIL库。

（5）带灰度值处理的图片，就是俗称的黑白照片，有利图像的进一步处理（√）

3、实验题

（1）实验一，在猫图片上加一个数字5，采用两种方法添加

❶通过标记文本方法，并且要求数字逆向旋转90度

❷通过数组填补值的方法

（2）实验二，把实验一的图片进行灰度处理，切割数字，保存数字图片

# 第七章

1、填空题

（1）通过Matplotlib库建立三维坐标，可以通过设置plt.axes的参数（projection='3d'）来实现，也可以通过导入（Axes3D）实例对象来实现。

（2）从二维坐标系到三维坐标系最大的区别是增加了（Z）坐标。

（3）在三维空间可以用（scatter()）绘制散点，用（plot3D）绘制线，用（plot\_surface）绘制三维上的一个平面。

（4）Matplotlib库工程化主体通过（Web）项目、（GUI）项目两种方式的结合来实现。

（5）在工作路径下配置文件参数的设置，可以影响工作路径下的（Jupyter Notebook）页内的代码。

2、判断题

（1）理论上可以通过三维坐标构建数字地球（√）

（2）在三维模型上，可以通过LightSource()对象实现不同角度的打光的功能（√）

（3）若要从不同角度观看三维实体，可以通过设置axes对象的参数来实现（╳）

若要从不同角度观看三维实体，可以通过设置subplot()对象的参数来实现。

（4）三维像素体指在三维空间指定三维坐标的情况下，用固定三维体连续叠加绘制组合形状，像素体为小立方体（╳）

像素体为小立方体也可以是其它形状的三维实体。

（5）因Jupyter Notebook本身的缺陷，用代码演示三维动画建议用GUI界面的代码调试工具（√）

3、实验题

（1）实验一，用两种方法各建立一个立方体

（2）实验二，旋转立方体

# 第八章

1、填空题

（1）Scipy是基于(python)、（numpy）基础上进行功能拓展的高级科学计算库。

（2）Special模块专门为求立方根提供了（cbrt(x)）函数，等价于（pow(x,1/3)）函数。

（3）用（linalg.block\_diag()）函数建立块对角矩阵，使对角值为指定值。

（4）quad()是（通用积分）函数用于求面积，dblquad()是（双积分）函数用于求体积，trapz()是（复合梯形积分）函数用于求不同子区间的面积。

（5）对于稀疏矩阵在实际存储时，为了节省（空间），采用不同的（压缩存储）方式。

2、判断题

（1）在Linux操作系统里可以用scipy.io.wavfile.read()打开WAV音频文件。（╳）

WAV文件是只能在Windows操作系统里读写的音频文件。

（2）当需要对一个样本数进行均值、方差等特征值统计时，可以用描述性统计函数一次调用解决问题。（√）

（3）核密度估计模型，可以用于预测地理空间点的数据的分布规律，如热力图 的建立；也可以用于金融等领域的基于密度的预测，应用范围比较广。（√）

（4）常微分方程的形成与发展是和力学、天文学、物理学，以及其他科学技术的发展密切相关的。（√）

（5）ConvexHull()函数适用于任何形状的图形的面积计算（╳）

ConvexHull(()适用于外围线凸出的图形的面积的计算。

3、实验题

（1）实验一，对手写图片5进行描述性统计，并用条形图展现其特征

（2）实验二，对如下数学公式在二维坐标图上产生的曲线封闭空间进行面积计算

# 第九章

1、填空题

（1）Scipy提供的信号滤波函数达到了20几种，所有的（滤波函数）都必须借助（滤波器）才能过滤掉（无用）的信号，保留（有用）的信号。

（2）（interp1d()）实现单变量插值，（griddata()）实现多变量插值，（InterpolatedUnivariateSpline()）实现一维样条插值。

（3）Scipy库的optimize子模块包的最小二乘法包括（非线性）最小二乘法、（线性）最小二乘法计算功能。

（4）在独立安装Scipy、Matplotlib库的情况下，（plt.imread()）、（plt.imsave()）默认只能读写.png扩展名的图像文件，要读写更多的图像文件（如jpg、bmp、gif），必须先安装第三方库（PIL）可以通过pip install pillow在线安装。

（5）Scipy库提供了（scipy.cluster.vq）、（scipy.cluster.hierarchy）两大类聚类计算子模块。

2、判断题

（1）在信号处理中，常见的窗函数有矩形窗、三角窗、汉宁窗（Hann Windows）、高斯窗等。除了矩形窗，其他窗在时域上体现为中间高、两端低的特征。 应用最广泛的是矩形窗。（╳）

应用最广泛的是汉宁窗。

（2）插值要求曲线通过所有样本点，拟合是逼近样本点（√）

（3）优化（Optimize），在计算机算法领域，优化往往是指通过算法得到要求问题的更优解。拟合是优化的方法之一。（√）

（4）平滑滤波采用简单平均法求邻近像素点的平均亮度，领域的大小与平滑的效果相关，领域越大平滑效果越好，但会使边缘信息放大，让图像边缘变得棱角分明。（╳）

但会使边缘信息丢失越大，让图像变得模糊。

（5）类与分类的区别，在于聚类所要求划分的类是未知的，而分类的划分类型是事先确定的。（√）

3、实验题

（1）实验一，根据提供的散点值，样条插值方法连接曲线

提供的散点y值：dots=np.array([9,8,7,5,4,3,3,2,1,6,7,8,9,10,11,12,14,17,19,21])，散点x值围默认顺序号值。

（2）实验二，在图9.25的基础上，切割字母A，并直方图统计A的特征

# 第十章

1、填空题

（1）目前，Pandas库主要提供（Series）、（DataFrame）两类的数据结构对象，用于（数据）的存储及（分析）处理 。

（2）带索引的一维数据对象用（Series()）创建，带索引的二维数据对象用（DataFrame()）创建。

（3）对DataFrame的数据修改可以通过（iloc）属性、用（Series）值修改指定列值、用（where()）方法修改符合条件的数据。

（4）DataFrame可以通过（index）、（columns）参数建立多层级索引，也可以通过（MultiIndex）对象的方法建立多层级索引。

（5）Pandas的数据获取、加工处理、数据分析，工程应用，离不开通过各种（文件）、（数据库）等对数据的读写操作。

2、判断题

（1）Series是一维带标签（Label）的类似数组对象，能够保存Python所支持类型的值，如整数、字符串、浮点数、布尔及Python对象等。（√）

（2）Series、DataFrame对象可以通过迭代对象、字典、常量进行创建，适用于所有python3.X版本。（╳）

迭代对象、字典（仅在Python3.6开始支持）、常量

（3）对DataFrame数据进行排序不会改变数据值，进行排名改变数据值（╳）。

进行排名也不改变原数据，只是生成了临时新的排名数据。

（4）对于加法、减法、乘法、除法、取整、取余、对数、幂等运算，可以把DataFrame看作一个运算数据对象。（√）

（5）利用to\_clipboard()实现指定数据复制后，粘贴时有可能会发生数据丢失的问题，需要第一时间处理数据 （√）

3、实验题

（1）实验一，对MySQL数据库数据进行读写操作

要求：

❶安装MySQL数据库系统，

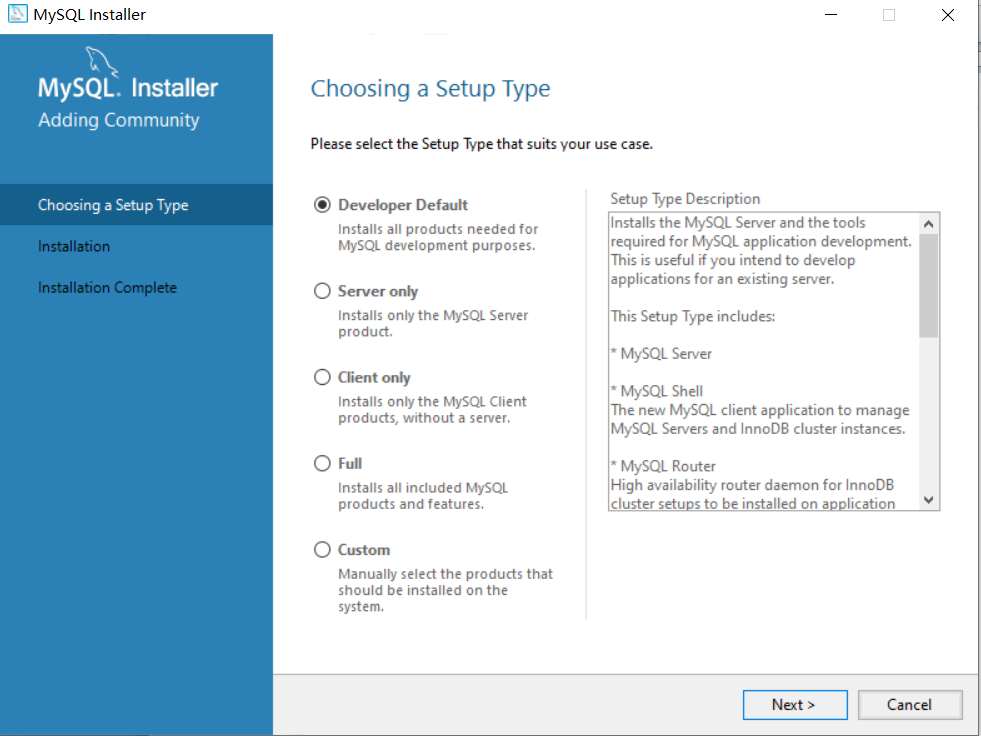
❷安装MySQL数据库系统配套的驱动程序

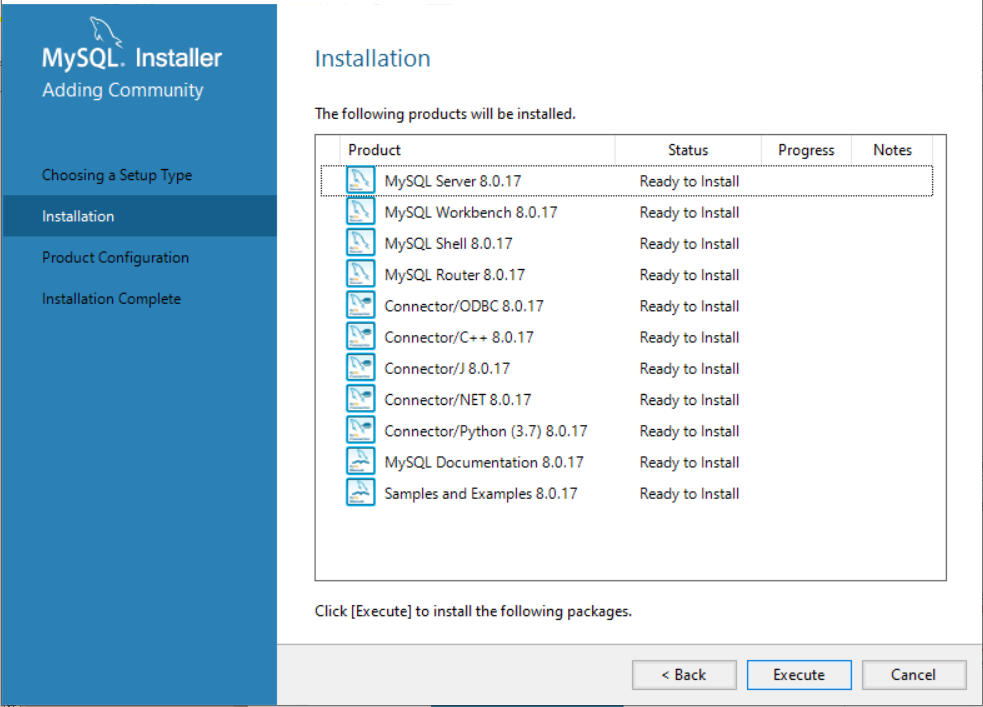
❸建立数据库、表

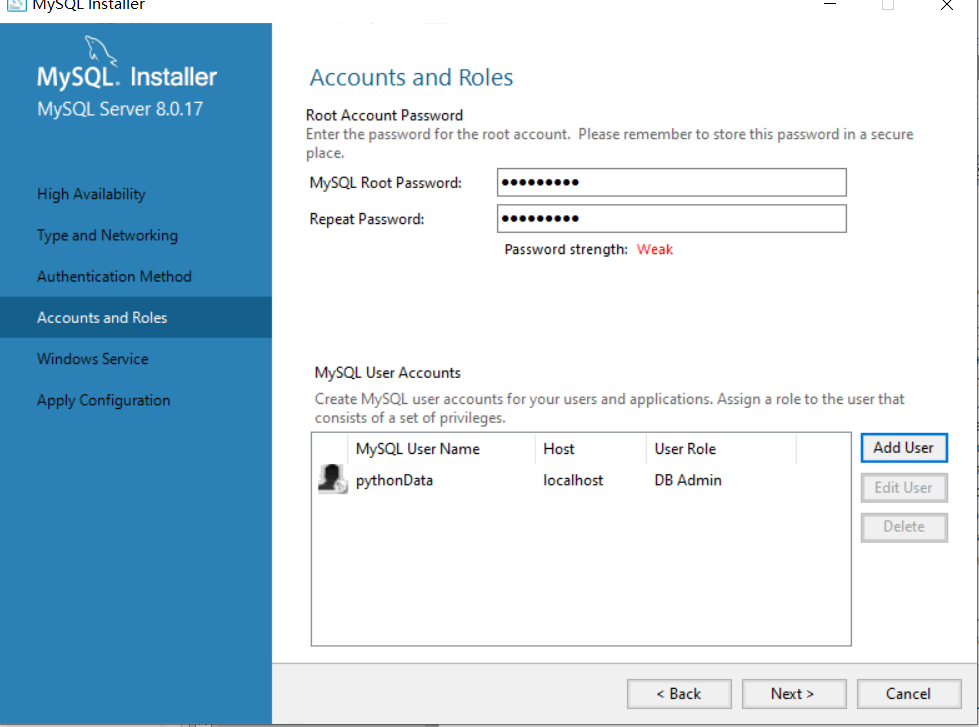
❹把DataFrame数据存入数据库表

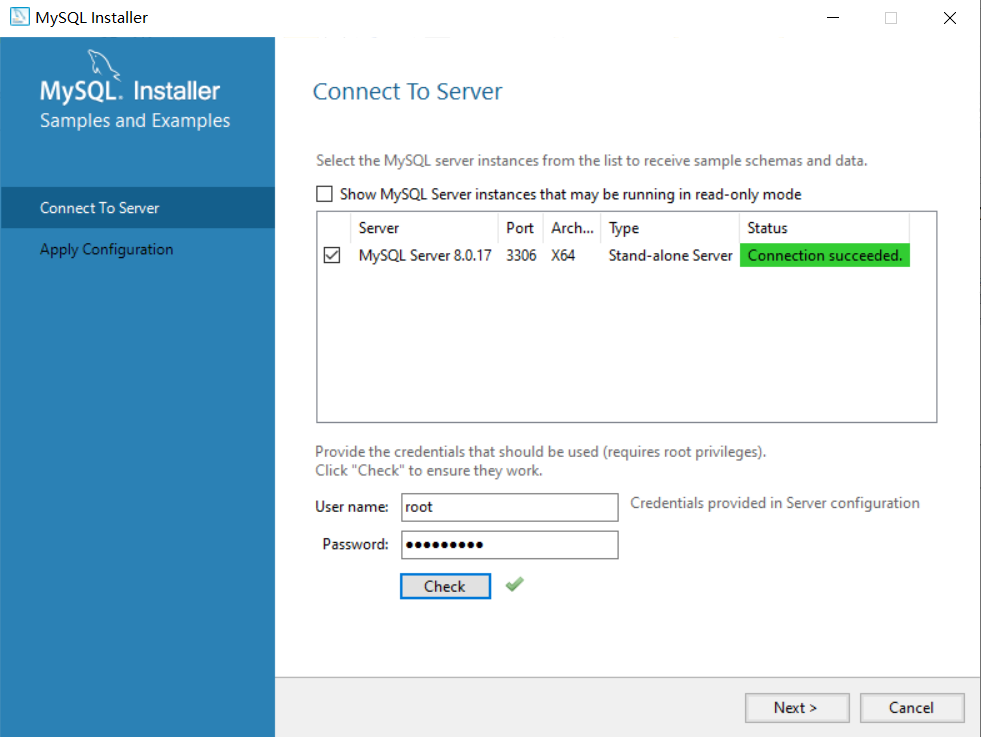
写出上述操作步骤，并截取相应的安装界面。

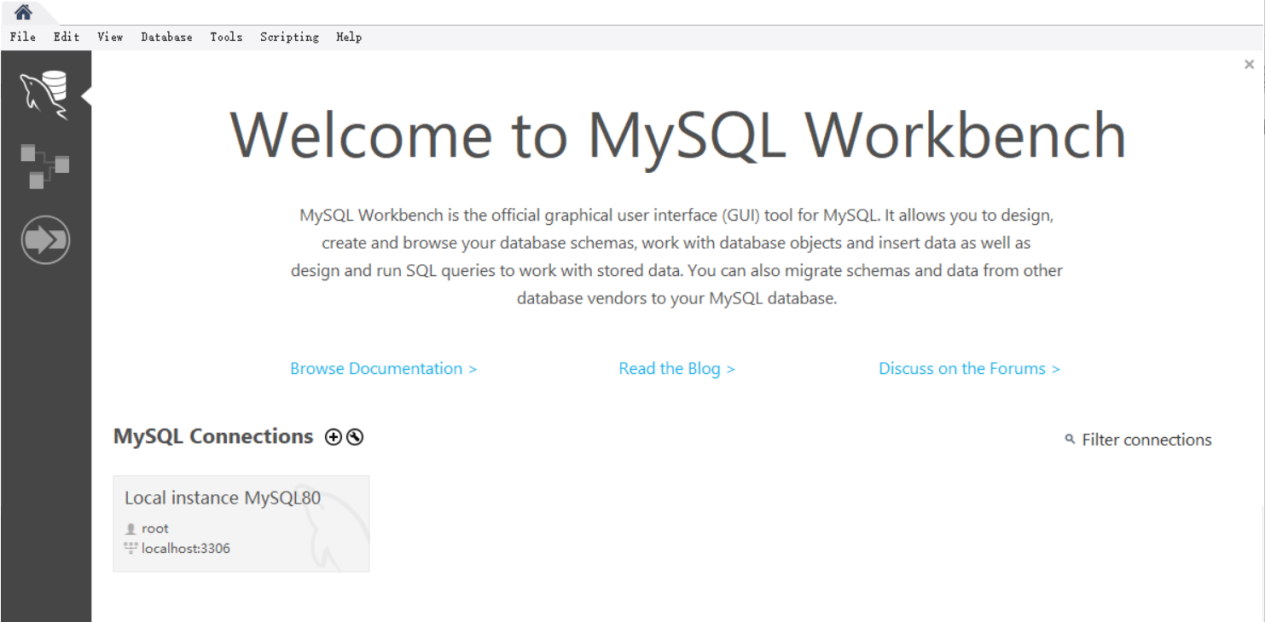
第一步，MySQL数据库系统安装

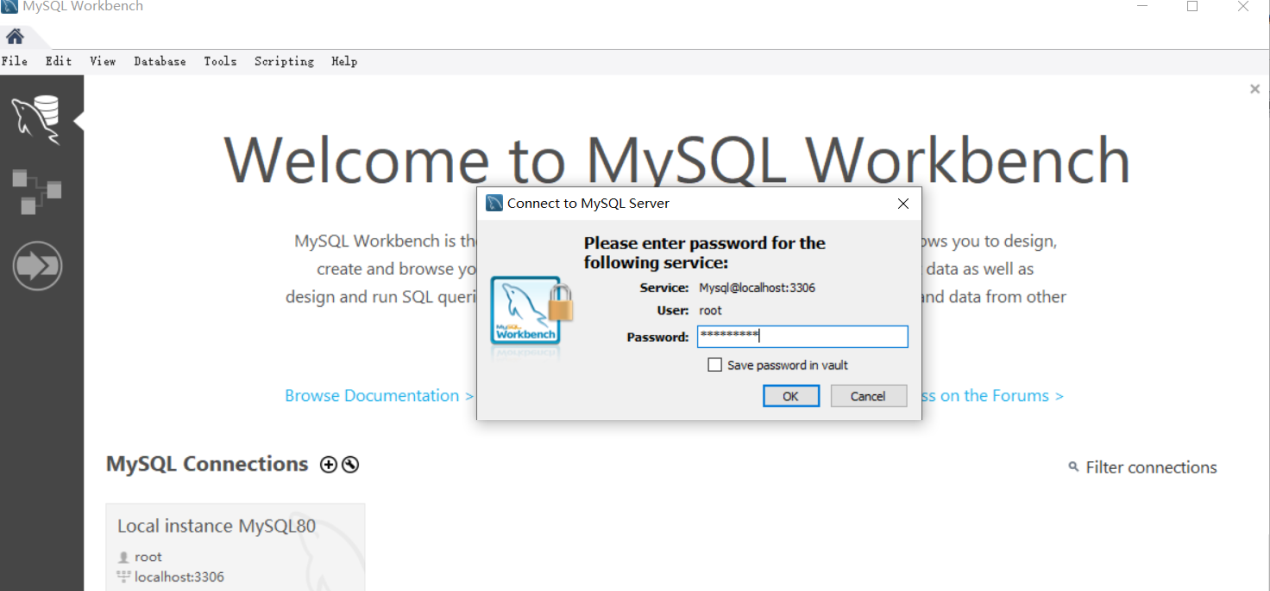




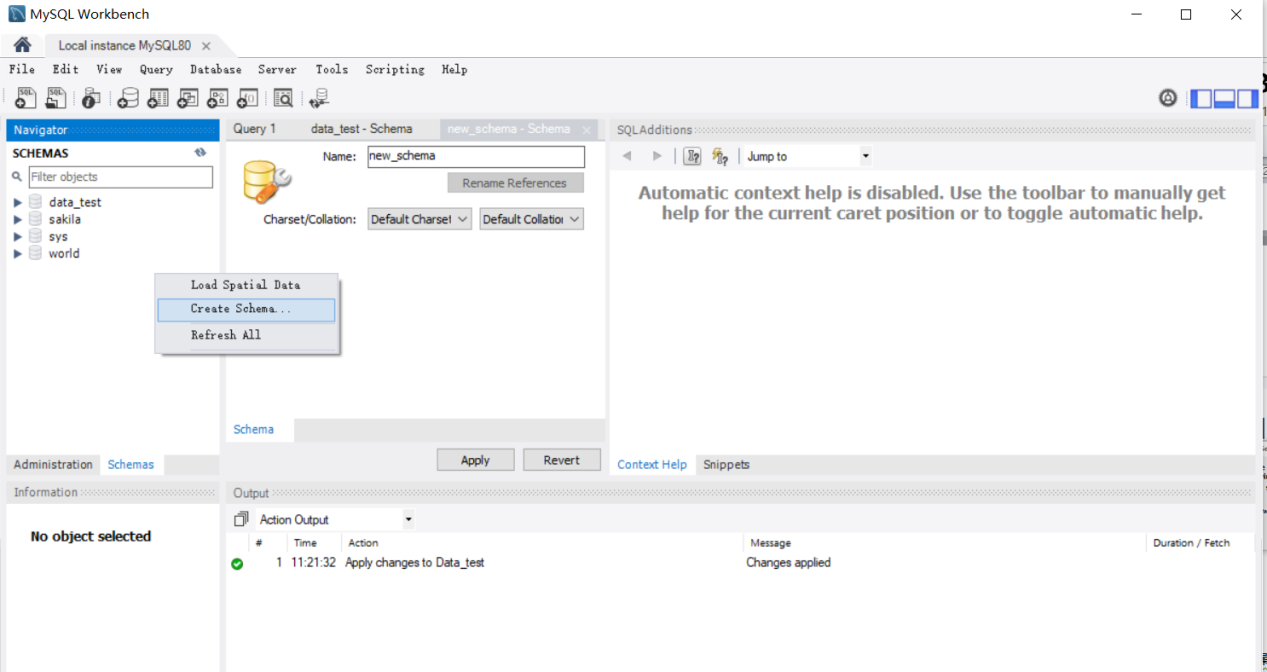




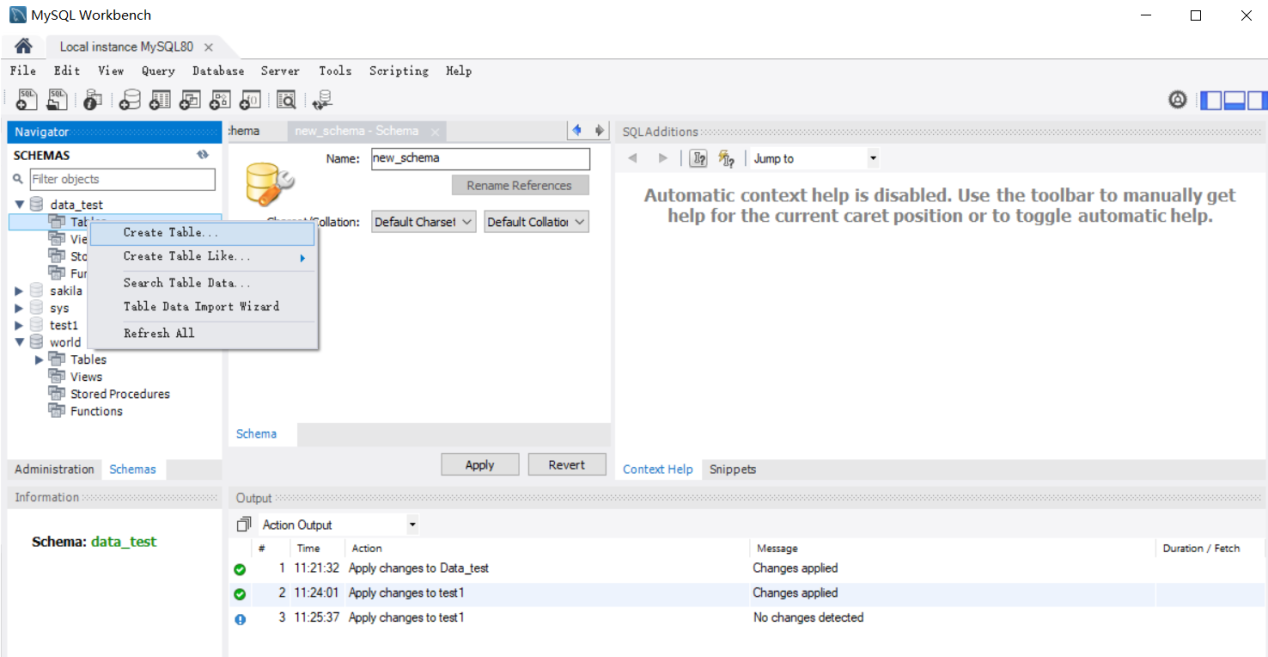




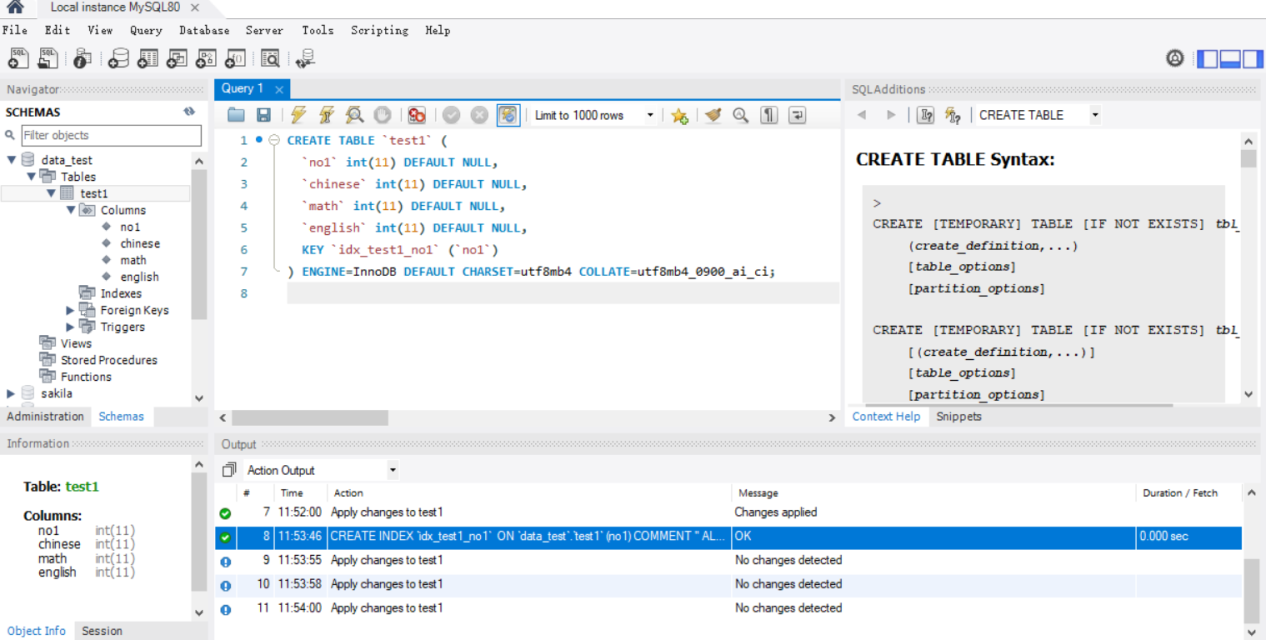
输入安装时设置的第一个密码。进入数据库表操作界面。



建立Data\_test数据库



test1表



CREATE TABLE `test1` (

`no1` int(11) DEFAULT NULL,

`chinese` int(11) DEFAULT NULL,

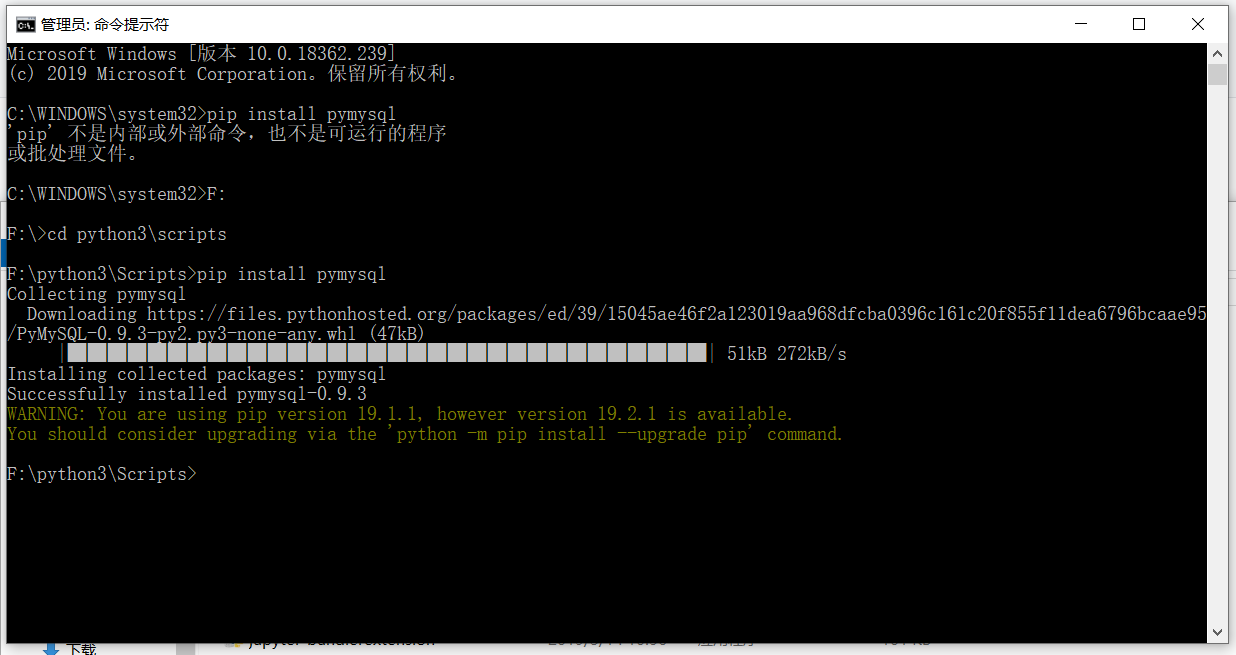
`math` int(11) DEFAULT NULL,

`english` int(11) DEFAULT NULL,

KEY `idx\_test1\_no1` (`no1`)

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

安装pymysql驱动程序



import pymysql

import sys

#============================连接数据库

try:

conn=pymysql.connect(host='localhost', user='root', passwd='mysql123.', db='data\_test', port=3306,

charset='utf8')

except:

print("打开数据库连接出错，请检查！")

conn.close()

sys.exit()

#============================对表进行插入、修改、删除、查找操作

cur=conn.cursor()

insertSQL='''insert into test1 values(1,100,99,99)'''

insertSQL1='''insert into test1 values(2,99,100,99)'''

try:

cur.execute(insertSQL)

cur.execute(insertSQL1)

conn.commit()

print("两条记录插入成功！")

except:

print("两条记录插入失败！")

conn.close()

sys.exit()

selectSQL='Select \* from test1'

cur.execute(selectSQL)

l\_records=[]

for row in cur.fetchall():

l\_records.append(row)

print(l\_records)

（2）实验二，对实验一数据进行统计， 并发布

❶求每门课的和、平均值，三门课的和、平均值

❷把带和、平均值的表格发布到HTML上

# 第十一章

1、填空题

（1）在Pandas里缺失数据用（NaT）、（NaN）、（nan）等来表示。

（2）Pandas为不同DataFrame对象的合并，提供了merge()函数，其合并方法为（left）、（right）、（outer）、（inner）。

（3）把行置换为（列），把列转为（行），可以用pivot()方法、stack()方法、unstack()方法、melt()方法。

（4）对DataFrame的数据绘制线形图，可以直接通过（自带）的plot()对象功能实现，也可以通过（Matplotlib）的plot()对象功能实现。

（5）对于字符串进行深度处理，如大小写的转换、缺失值的处理等，需要借助DataFrame、Series的（str）对象的（方法）或（正则）表达式。

2、判断题

（1）对DataFrame对象进行合并或连接，仅用于数据的横向合并或连接。（√）

（2）concat()通过axis参数的设置可以实现两DataFrame对象的横向或竖向的合并。（√）

（3）sum()、count()、mean()、median()、min()、max()、idxmin()、idxmax()、argmin()、argmax()都具有忽略缺失值的功能。（√）

（4）DataFrame的describe()可以一次性提供count、mean、百分位数、标准差、最小值、最大值统计内容，但是百分位数的百分比是固定的（╳）

可以通过参数percentiles对百分比进行调整。

（5）groupby()方法 、aggregate()方法都可以直接实现DataFrame对象数据不同方向的累加统计（╳）

groupby()方法需要借助DataFrame的sum()方法。aggregate()方法需要借助’sum’函数才能做累加统计。

3、实验题

（1）实验一，对下列二维数据对象，实现数据的四种合并操作

表11.4 学生信息表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 姓名 | 性别 | 国籍 | 年龄 |
| 0 | 三酷猫 | 男 | 中国 | 18 |
| 1 | 加菲猫 | 男 | 美国 | 19 |
| 2 | 凯蒂猫 | 女 | 日本 | 16 |
| 3 | 机器猫 | 男 | 日本 | 17 |

表11.5 学生成绩表

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 姓名 | 班级 | 总分 |
| 0 | 三酷猫 | 1 | 290 |
| 1 | 加菲猫 | 2 | 270 |
| 2 | 凯蒂猫 | 1 | 280 |
| 3 | 土猫 | 1 | 260 |

（2）实验二，对图片进行样本描述性统计

对书附赠的F5\_T.png图片通过DataFrame进行样本描述性统计。

用直方图统计均值。

# 第十二章

1、填空题

（1）Pandas提供了四种类型的生成日期时间的对象：（日期时间）Date Times、（时间增量）Time Deltas、（时间跨度）Time Spans、（日期偏移量）Date Offsets。

（2）时间增量是相对指定时间点上的（绝对）时间差异，用不同的单位表示，例如天、小时、分钟、秒。它们可以是（正数），也可以是（负数）。

（3）周期表示一段范围的时间，如一天、一个周、一个月、一个季度、一个年度等。规则的时间周期由Pandas中的（pd.Period()）对象表示，（pd.period\_range()）产生连续的周期序列对象PeriodIndex。

（4）日期偏移量更加遵循（日历）持续时间规则，如DateOffse()在增加日时，总是增加到指定日的同一时间，（而忽略夏令时等所带来的时间差异）；而Timedelta()在增加日时，每天增加（24小时）。

（5）Pandas的Series、DataFrame对象都提供了（resample()）重采样方法。根据时间采样数的不同，可以分为（降采样）、（升采样）。

2、判断题

（1）pd.Timestamp(year=2019, month=6, day=1, hour=8,minute=1,second=10,microsecond=1000)，用键值对形式指定年、月、日、时、分、秒、毫秒。（╳）

Microsecond是微秒。

（2）Timedelta是datetime的一个子类用于提供时间增量计算功能，其提供增量时间值时，必须在参数里指定增量时间单位。（√）

（3）M= pd.Period('2019-01', freq='M') 表示建立时间在以分钟为周期的一个变量M，然后可以对该变量做周期加减等操作。（╳）

以月为周期的一个变量。

（4）在实际工作当中，为了准确安排日期，在做时间计划安排时，采用日期偏移更加准确科学（√）

（5）利用D1.resample()方法，时间采样频率从1秒变成了5秒，属于升采样。（╳）

采样频率降低了，从1秒 采样一条记录变成了5秒采样一条记录，属于降采样。

3、实验题

（1）实验一，获取2020年12个月的月末日期，并计算当日的星期。

（2）实验二，观察读者入书群时间规律

作者的《Python编程从零基础到项目实战》一书在2018年10月正式上市，并为读者建立了学习交流QQ群。由于作者好奇性很大，想知道一天内读者入群的时间规律，方便作者集中精力给读者解疑答惑。于是把入QQ群的读者信息保存为“QQ入群信息.csv”文件。

要求通过DataFrame的数据读取，数据整理，数据转化，数据统计，数据显示，用图直观地显示一天24小时，以小时为单位统计入群人数，以发现一天哪些时段入群的读者比较多。

# 第十三章

1. 填空题

（1）特征（Feature）刻画一事物异于其他事物的特点，由（属性）和（属性）值组成。

（2）模型（Model），机器学习中的模型，借助（算法）和（数据）形成固定模式，实现对事物性质的准确判断或表达。

（3）机器学习的样本集数据可以分（训练）集、（测试）集和（验证）集，其中前两类使用场景最多。

（4）Scikit-learn库所提供的模型采用的统一的使用方法，（fit()）方法用于训练模型，（predict()）方法用于预测模型、（score()）方法用于预测准确度的评分。该特点简化了学习难度，有利读者更好地掌握模型使用功能。

（5）训练完成的模型在实际使用中表现不佳，可以从（数据集）、模型训练（方法）选择、模型(参数)调优、(模型)选择着手进行完善。

2、判断题

（1）SVC全称C-Support Vector Classification（C-支持向量分类），通过样本数据间的间隔（距离）计算对数据进行分类。（√）

（2）分类模型、聚类模型都应用于事物的分类，都是监督学习模型，都借助训练集、测试集训练模型。（╳）

分类模型属于监督学习模型，聚类模型属于非监督学习模型。

（3）通过降维方法，可以降低数据集处理的特征维度，减少对存储空间的要求，提高运行效率，使数据处理可视化，使机器学习更加实用。属于无损数据处理。（╳）

属于有损失数据处理，只不过要求损失信息可以接受。

（4）欠拟合是指模型拟合程度不高，数据距离拟合曲线较近；或模型没有很好地捕捉到数据特征，不能够很好地拟合数据。（╳）

数据距离拟合曲线较远。

（5）在实际情况下，采集的原始数据需要进行严格检查，根据数据的特点进行各种预处理后，才能被用于机器学习。（√）

3、实验题

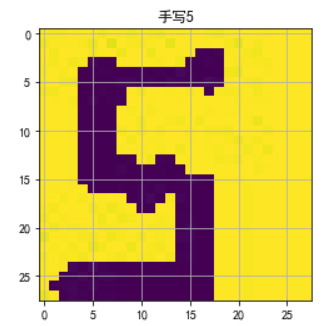
（1）实验一，降维前后的训练时间比较

利用13.6.2节 手写数字图像降维代码，并选择SGDClassifier()模型，测试降维前后模型训练所花时间。

（2）实验二，分析手写数字泛化性不强的原因

把案例13[三酷猫预测手写数字]的测试图片换成F51.png图片，图片资料见书附赠地址。该图片数字5由作者用画图软件通过鼠标绘制。

分析泛化性不强的主要因素，并提出完善学习模型建议（至少两点）。



鼠标绘制的5数字，识别结果为7，显然不正确。

[7.]

训练完成的模型在实际使用中表现不佳，可以从数据集、模型训练方法选择、模型参数调优、模型选择着手进行完善。

问题原因分析及建议：

1、训练集图片，没有类似特征的数字图片，可以在训练集增加鼠标绘制图片的数据，来提高训练效果。

2、书上采用的是tree.DecisionTreeRegressor()模型，存在模型不适应的可能性，可以通过选择其它模型进行对比尝试，并调整模型的算法、参数。

# 书配套赠书代码

|  |  |  |
| --- | --- | --- |
| **章次** | **代码文件名称** | **所属节内容** |
| 第一章 | 无 |  |
| 第二章 | U2\_23节.ipynb | 2.3节Jupyter Magic（魔法）命令 |
| U2\_Experiment1 | 实验一，代码文件丢失问题的解决  实验二，用魔法命令调试 |
| U2Experiment1.py |  |
| 第三章 | U3\_31\_32节.ipynb | 3.1 接触Numpy  3.2建立数组（3.2.1，3.2.2） |
| U3\_32节\_属性方法.ipynb | 3.2.3节数组属性的使用  3.2.4节数组方法的使用 |
| U3\_33\_Slice.ipynb | 3.3 索引与切片 |
| U3\_34\_Mathematical\_calculation | 3.4基本数学计算 |
| U3\_35\_1初等函数 | 3.5.1初等函数 |
| U3\_35\_23随机集合函数 | 3.5.2随机函数  3.5.3数组集合运算 |
| U3\_35\_45基础高级统计 | 3.5.4基础统计函数  3.5.5高级统计函数 |
| U3\_35\_6\_sort\_other | 3.5.6排序  3.5.7将数值替换到数组指定位置  3.5.8增加和删除行（列）  3.5.9数值修约等杂项函数 |
| Un3\_link\_split\_example1 | 3.2.6 案例1[建立学生成绩档案] |
| U3\_Experiment1 | 实验一，三酷猫批发冰淇淋 |
| U3\_Experiment2 | 实验二，三酷猫图像处理 |
| 第四章 | U4\_4142\_行列式矩阵建立1 | 4.1行列式进阿里及计算  4.2.1构建矩阵 |
| U4\_424344 | 4.2.2-4.2.4节  4.3求线性方程组  4.4向量、特征向量、特征值  5.5案例4三酷猫求三维空间面积 |
| U4\_Experiment1 | 实验一，求方阵A的特征值、特征向量  实验二，三酷猫求围棋格子的所有坐标 |
| 第五章 | U5\_51\_File | 5.1处理数据文件 |
| U5\_52\_12\_narray | 5.2.1数组结构  5.2.2副本与视图 |
| U5\_54\_Example\_food | 5.4案例5[三酷猫制订减肥计划] |
| U5\_52\_3broadcasting | 5.2.3广播原理 |
| U5\_Experiment1 | 实验一，广播计算  实验二，三酷猫文字处理 |
| 第六章 | U6\_6162\_Image | 6.1开始绘图  6.2绘制图形 |
| U6\_6364\_photo | 6.3处理图像  6.4案例6[三酷猫戴皇冠] |
| U6\_Experiment1 | 实验一，在猫图片上加一个数字5，采用两种方法添加  实验二，把实验一的图片进行灰度处理，切割数字，保存数字图片 |
| 第七章 | U7\_717273\_3D | 7.1绘制三维图形  7.2动画  7.3工程化 |
| U7\_7475\_Params | 7.4参数配置  7.5案例7[三酷猫设计机械零配件] |
| U7\_Experiment1 | 实验一，用两种方法各建立一个立方体  实验二，旋转立方体 |
| 7\_Experiment1.py | 建议在IDLE、Pycharm上运行三维动画 |
| 第八章 | U8\_1234\_base | 8.1接触Scipy  8.2特殊数字函数  8.3读写数据文件  8.4线性代数 |
| U8\_56\_stats-ODE | 8.5统计  8.6积分 |
| U8\_789\_spatial | 8.7空间算法和数据结构  8.8稀疏矩阵  8.9案例8[三酷猫统计岛屿面积] |
| 8\_Experiment1 | 实验一，对手写图片5进行描述性统计，并用条形图展现其特征  实验二，对如下数学公式在二维坐标图上产生的曲线封闭空间进行面积计算 |
| 第九章 | U9\_123456\_Scipy2 | 9.1信号处理  9.2插值  9.3优化于拟合  9.4多维图像处理  9.5聚类  9.6案例9[三酷猫图像文字切割] |
| 9\_Experiment1 | 实验一，根据提供的散点值，样条插值方法连接曲线  实验二，在图9.25的基础上，切割字母A，并直方图统计A的特征 |
| 第十章 | U10\_12\_Series | 10.1接触Pandas  10.2 Series基本操作 |
| U10\_34\_DataFrame | 10.3DataFrame基本操作  10.4DataFrame数据索引深入 |
| U10\_567\_CRW | 10.5数据计算  10.6读写数据  10.7案例10[三酷猫发布交易公告] |
| 10\_Experiment1 | 实验一，对MySQL数据库数据进行读写操作  实验二，对实验一数据进行统计， 并发布 |
| 第十一章 | U11\_1234\_Pandas | 11.1缺失数据处理  11.2多源数据操作  11.3数据转置和透视表  11.4数据统计 |
| U11\_5678\_groupby | 11.5数据分组和聚合运算  11.6数据可视化  11.7字符串数据处理  11.8案例11[散酷猫分析简历] |
| 11\_Experiment1 | 实验一，对下列二维数据对象，实现数据的四种合并操作  实验二，对图片进行样本描述性统计 |
| resume.csv、MD.csv |  |
| 第十二章 | U12\_12\_DateTime | 12.1时间处理基础  12.2时间增量处理 |
|  | U12\_34567\_Period | 12.3时间周期处理  12.4日期偏移处理  12.5日期重采样  12.6基于时间的绘图处理  12.7案例[三酷猫分析 历年分数线] |
|  | 12\_Experiment1 | 实验一，获取2020年12个月的月末日期，并计算当日的星期。  实验二，观察读者入书群时间规律 |
|  | 中国GDP数据.csv |  |
| 第十三章 | U13\_123\_Skilearn | 13.1机器学习入门  13.2数据准备  13.3分类 |
| U13\_4\_sklearn | 13.4 回归 |
| U13\_5\_聚类 | 13.5聚类 |
| U13\_6\_降维 | 13.6降维 |
| U13\_7\_ModelSelection | 13.7模型选择 |
| U13\_810\_Example | 13.8数据预处理  13.10案例13[三酷猫预测手写数字] |
| 13\_Experiment1 | 实验一，降维前后的训练时间比较  实验二，分析手写数字泛化性不强的原因 |