

零 级 标 题

科技入门一本通之 MATLAB

一、MATLAB 简介

MATLAB 是美国 MathWorks 公司出品的商业数学软件，用于算法开发、数据可视化、数据分析以及数值计算的高级技术计算语言和交互式环境，主要包括 MATLAB 和 Simulink 两大部分。

MATLAB 是由美国 mathworks 公司发布的主要面对科学计算、可视化以及交互式程序设计的高科技计算环境。它将数值分析、矩阵计算、科学数据可视化以及非线性动态系统的建模和仿真等诸多强大功能集成在一个易于使用的视窗环境中，为科学研究、工程设计以及必须进行有效数值计算的众多科学领域提供了一种全

面的解决方案，并在很大程度上摆脱了传统非交互式程序设计语言（如 C、Fortran）的编辑模式，代表了当今国际科学计算软件的先进水平。

MATLAB 的基本数据单位是矩阵，它的指令表达式与数学、工程中常用的形式十分相似，故用 MATLAB 来解算问题要比用 C, FORTRAN 等语言完成相同的事情简捷得多，并且 MATLAB 也吸收了像 Maple 等软件的优点，使 MATLAB 成为一个强大的数学软件。在新的版本中也加入了对 C, FORTRAN, C++, JAVA 的支持。

二、功能特性

1、主要功能：

- » 数值分析数值和符号计算
- » 工程与科学绘图
- » 数字图像处理
- » 数字信号处理
- » 通讯系统设计与仿真
- » 财务与金融工程
- » 控制系统的设计与仿真

2、重要功能：

- (1) MATLAB: MATLAB 语言的单元测试框架。
- (2) Trading Toolbox : 一款用于访问价格并将订单发送到交易系统的新产品。
- (3) Financial Instruments Toolbox: 赫尔-怀特、线性高斯和 LIBOR 市场模型的校准和 Monte Carlo 仿真。

(4) Vehicle Network Toolbox: 为访问 CAN 总线上的 ECU 提供 XCP。

(5) Image Processing Toolbox: 使用有效轮廓进行图像分割、对 10 个函数 实现 C 代码生成，对 11 个函数使用 GPU 加速。

(6) Image Acquisition Toolbox: 提供了用于采集图像、深度图和框架数据的 Kinect for Windows 传感器支持。

(7) Statistics Toolbox: 用于二进制分类的支持向量机 (SVM)、用于缺失数据的 PCA 算法和 Anderson-Darling 拟合优度检验。

(8) Data Acquisition Toolbox: 为 Digilent Analog Discovery Design Kit 提供了支持包。

三、优势特点

1、优势：

- (1) 高效的数值计算及符号计算功能，能使用户从繁杂的数学运算分析中解脱出来；

(2) 具有完备的图形处理功能，实现计算结果和编程的可视化；

(3) 友好的用户界面及接近数学表达式的自

然化语言，使学者易于学习和掌握；

(4) 功能丰富的应用工具箱（如信号处理工具箱、通信工具箱等），为用户提供了大量方便实用的处理工具。

2、特点：

(1) 编译环境：MATLAB 由一系列工具组成。这些工具方便用户使用 MATLAB 的函数和文件，其中许多工具采用的是图形用户界面。简单的编程环境提供了比较完备的调试系统，程序不必经过编译就可以直接运行，而且能够及时地报告出现的错误及进行出错原因分析。

(2) 简单易用：Matlab 是一个高级的矩阵/阵列语言，它包含控制语句、函数、数据结构、输入和输出和面向对象编程特点。用户可以在命令窗口中将输入语句与执行命令同步，也可以先编写好一个较大的复杂的应用程序(M 文件)后再一起运行。

(3) 强大处理：MATLAB 是一个包含大量计算算法的集合。其拥有 600 多个工程中要用到的数学运算函数，可以方便的实现用户所需的各种计算功能。函数中所使用的算法都是科研和工程计算中的最新研究成果，而且经过了各种优化和容错处理。在通常情况下，可以用它来代替底层编程语言，如 C 和 C++。在计算要求相同的情况下，使用 MATLAB 的编程工作量会大大减少。

(4) 图形处理：MATLAB 自产生之日起就具

有方便的数据可视化功能，以将向量和矩阵用图形表现出来，并且可以对图形进行标注和打印。高层次的作图包括二维和三维的可视化、图象处理、动画和表达式作图。可用于科学计算和工程绘图。

(5) 模块工具：MATLAB 对许多专门的领域都开发了功能强大的模块集和工具箱。用户可以直接使用工具箱学习、应用和评估不同的方法而不需要自己编写代码。工具箱中具有诸多强大工具。

(6) 程序接口：新版本的 MATLAB 可以利用 MATLAB 编译器和 C/C++ 数学库和图形库，将自己的 MATLAB 程序自动转换为独立于 MATLAB 运行的 C 和 C++ 代码。允许用户编写可以和 MATLAB 进行交互的 C 或 C++ 语言程序。另外，MATLAB 网页服务程序还容许在 Web 应用中使用自己的 MATLAB 数学和图形程序。

(7) 软件开发：在开发环境中，使用户更方便地控制多个文件和图形窗口；在编程方面支持了函数嵌套，有条件中断等；在图形化方面，有了更强大的图形标注和处理功能，包括对性对起连接注释等；在输入输出方面，可以直接向 Excel 和 HDF5 进行连接。

四、常用工具

MATLAB 包括拥有数百个内部函数的主包和三十几种工具包。工具包又可以分为功能性工具包和学科工具包。功能工具包用来扩充 MATLAB 的符号计算，可视化建模仿真，文字处理及实时控制等功能。学科工具包是专业性比较强的工具包，控制工具包，信号处理工具包，通信工具包等都属于此类。

开放性使 MATLAB 广受用户欢迎。除内部函数外，所有 MATLAB 主包文件和各种工具包都是可读可修改的文件，用户通过对源程序的修改或加入自己编写程序构造新的专用工具包。

```
#include "reg52.h"           //此文件中定义了单片机的一些特殊功能寄存器
#include<intrins.h>          //因为要用到左右移函数，所以加入这个头文件

typedef unsigned int ul6;    //对数据类型进行声明定义
typedef unsigned char u8;

#define led P0                 //将P0口定义为led 后面就可以使用led代替P0口

/********************* 函数名 : delay
 * 函数功能   : 延时函数, i=1时, 大约延时10us
 *****/
void delay(ul6 i)
{
    while(i--);
}

void main()
{
    u8 i;
    led=0x01;
    delay(50000); //大约延时450ms
    while(1)
    {
        for(i=0;i<7;i++) //将led左移一位
        {
            led=_crol_(led,1);
            delay(50000); //大约延时450ms
        }
        for(i=0;i<7;i++) //将led右移一位
        {
            led=_cror_(led,1);
            delay(50000); //大约延时450ms
        }
    }
}
```

五、常用函数

整数
» $\text{fix}(x)$: 无论正负, 舍去小数至相邻整数;
» $\text{floor}(x)$: 下取整, 即舍去正小数至相邻整数;
» $\text{ceil}(x)$: 上取整, 即加入正小数至相邻整数;
» $\text{rat}(x)$: 将实数 x 化为多项分数展开;
» $\text{rats}(x)$: 将实数 x 化为分数表示;
» $\text{sign}(x)$: 符号函数 (Signum function);
当 $x < 0$ 时, $\text{sign}(x) = -1$;
当 $x = 0$ 时, $\text{sign}(x) = 0$;
当 $x > 0$ 时, $\text{sign}(x) = 1$;
» $\text{rem}(x, y)$: 求 x 除以 y 的余数;
» $\text{gcd}(x, y)$: 整数 x 和 y 的最大公因数;
» $\text{lcm}(x, y)$: 整数 x 和 y 的最小公倍数;
» $\text{exp}(x)$: 自然指数;
» $\text{pow2}(x)$: 2 的指数;
» $\text{log}(x)$: 以 e 为底的对数, 即自然对数或;

MATLAB 基本绘图函数

» plot : x 轴和 y 轴均为线性刻度 (Linear scale) ;
» loglog : x 轴和 y 轴均为对数刻度 (Logarithmic scale) ;
» semilogx : x 轴为对数刻度, y 轴为线性刻度;
» semilogy : x 轴为线性刻度, y 轴为对数刻度;
» plot 绘图函数的参数
字元颜色 字元图线型态:
y 黄色 . 点;
k 黑色 o 圆;
w 白色 x x;
b 蓝色 + +;
g 绿色 * *;
r 红色 ;
- 实线;
c 亮青色;

六、推荐书籍

《MATLAB 数学建模经典案例实战》 余胜威
编著 (仅适用于对数学建模有兴趣的同学)
《MATLAB5 手册》 ChinaPub 出版
其余相关书籍可以到学校图书馆找。

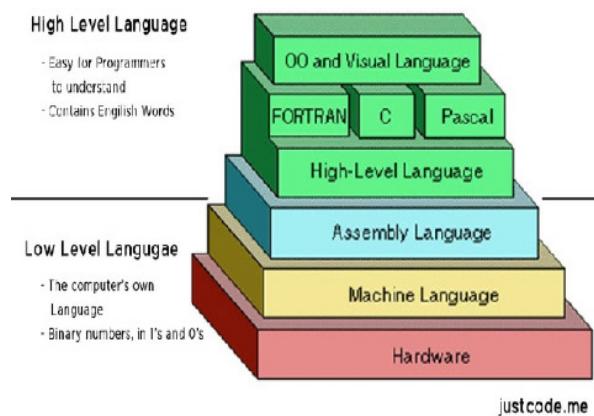
七、建议

- 1、想学 matlab 的同学可以先了解一些相关命令, 再从网上找一些教程 (哔哩哔哩、MOOC、睿思上都有) 进行深入学习;
- 2、可以试着去参加数学建模的比赛, 在比赛中能让你更快的掌握一项技能;
- 3、软件下载 Google、百度、睿思上都有。

科技入门一本通之汇编语言

汇编语言

有些人说“学习编程其实就是学高级语言”。但计算机不理解高级语言，必须通过编译和汇编转成二进制代码，才能运行。由于高级语言将一些操作和接口用函数封装起来，所以学会高级语言，并不等于理解计算机实际的运行步骤。



例如 $2*12$ 这样一个操作，如果拿 C 语言来写，就是直接的 $2*12$ ，而写成汇编，则要：

```
assume cs:code  
code segment  
start: mov ax, 2  
       mov cx, 11  
       p:add, ax, ax  
       loop p
```

这里还没有变量的声明与引用，所以这还只是在寄存器中进行操作，并没有涉及到内存读写的操作。所以笔者觉得在这里有必要告诉读者这么繁琐的汇编语言优点在哪里。

优点

1. 为解决某些特定的问题代码中必须嵌入汇编语言以代替 C 语言完成的单片机的编程。

从左图可以看到，硬件上面的是机器语言，是一大堆二进制码。相传刚刚发明出计算机，人们要敲出这些二进制码来向 CPU 发出指令，比如加法指令是 00000011。终于，前辈们实在是觉得机器语言过于繁琐，于是将机器语言用简单的英文单词或缩写以及标点符号来代替这些繁琐的（根本看不懂的）机器语言，这就是汇编语言。汇编语言是所有语言中（除了机器语言），最为直接的和硬件打交道的语言。如果我们要形成一个软、硬兼备的编程知识体系，更好的了解计算机的工作原理，汇编语言则是首选。

虽然学习汇编语言如此重要，然而读者可能在学过 C 语言或者一些面向对象的语言之后，根本不想学汇编语言，觉得汇编语言对于更高层的语言来说过于繁琐。

也就是说想要写好一个单片机程序，让他可以顺利的进行，往往汇编语言是必不可少的。

2. 它通常被应用在底层，硬件操作和高要求的程序优化的场合。高级语言处理底层操作时使用的是封装后的函数。这些函数为了普适性，往往所有可能性全部编写出来，这就加大了代码的编译量。而在底层操作时用汇编语言，可以直接根据特定情况直接调用硬件进行处理。

3. 在 Debug 过程中可能需要查看代码编译后的汇编程序。而且在高级语言学习过程中不明白原理也可以查看编译后的汇编程序。

所以，还不快快开始学习汇编语言？

科技入门一本通之数据科学

数据科学是一门新兴的学科，最早在 20 实际 60 年代被提出，但当时并未受到学术界的广泛关注和认可。直到近年来，大数据、机器学习等的兴起，使数据科学这一概念很广泛地被提及。

1、数据科学是什么

数据科学（英语：Data Science），又称资料科学，是一门利用数据学习知识的学科，其目标是通过从数据中提取出有价值的部分来生产数据产品 [1]。它结合了诸多领域中的理论和技术，包括应用数学，统计，模式识别，机器学习，数据可视化，数据仓库，以及高性能计算。数据科学通过运用各种相关的数据来帮助非专业人士理解问题。数据科学技术可以帮助我们如何正确的处理数据并协助我们在生物，社会科学，人类学等领域进行研究调研。

2、背景介绍

接下来，简单介绍一下数据科学在如今互联网体系中已经成熟落地的应用。

1) 电商领域。比如淘宝的猜你喜欢。

2) 信息流产品。在这一方面，国内的字节跳动比较成熟，旗下的产品矩阵，比如抖音、今日头条等等，都是数据科学在推荐算法很成熟的应用。

3) 广告营销。如今，“增长黑客”概念逐渐兴起。增长黑客是介于技术和市场之间的新型团队角色，主要依靠技术和数据的力量来达成各种营销目标，而非传统意义上靠砸钱来获取用户的市场推广角色。比如微信裂变营销等等，都是增长黑客的杰作。对于增长黑客来说，数据为王，他们的主要目标都是靠数据量化的，比如 DAU (日活)、MAU (月活)、留存率等等，而他们也是靠数据驱动的。

4) 金融交易。如今，全球贸易至少 50% 的交易是由机器执行的，而算法是基于 quants 理论的，也就是量化交易。

当然数据科学几乎总是能应用到各种行业，在这里就不再加以更多叙述。

数据科学中有许多不同种类的数据类型，主要可分为以下几类：

结构化数据：数据依存与数据模型和记录驻留在一个固定的字段中，主要存储在数据库或 Excel 表格中。

非结构化数据：数据结构不规则或不完整，没有预定义的数据模型，不方便用数据库二维逻辑表来表现的数据。

自然语言数据：一种特殊类型的非结构化数据，也就是语音分析。

计算机数据：没有人工干预的，由计算机、过程、应用程序或其他机器自动创建的信息

图类数据：在这里主要指的是图论，即使用节点、边和属性来存储的。

音频、视频和图像数据

流数据：可以是以上任何一种数据，但是它有一个额外的属性，当事件法发生时，数据流入系统，而不是在批处理中加载到数据存储区。

3、数据科学过程

对于一个数据科学项目来说，通常会涉及如下 6 个步骤

a) 设置研究目标

一个项目始于这个项目是什么，为什么要做什么和怎么去做，如果定义不好这三个问题，那么在后续的步骤中有很大可能在做无用功。对于一个数据科学项目来说，输出的结果有很大一部分是去辅助客户决策的，如果把握不住项目目标和背景，往往会导致南辕北辙。所以数据科学家有很大一部分的工作内容是深入地理解业务需求，并且深入地去解释项目结果怎样去辅助决策。这也就导致了，身为一个数据科学家，不仅仅要有很好的数据分析能力，而且要有某一特定领域的专业知识，比如金融、广告营销等等。

在这个步骤结束后，通常会生成一个项目章程，主要包括以下内容

- 明确的研究目标
- 项目人物和背景
- 如何进行分析
- 期望使用什么数据
- 可交付成果和确保成功的措施
- 时间表

该项目章程也不仅仅适用于数据科学项目，也适用于其他许多项目，诸如游戏制作等。

b) 检索数据

数据来源可以是从公司内部得到的数据，也可以是外部数据，比如专业数据公司购买的数据。虽然如今普遍认为数据是比石油更有价值的资产，但是已经有很多组织开始免费地提供数据了，比如 Data.gov 就是一个美国政府的开放数据库。

在这里有几个概念，数据库的主要目的是存储数据，数据仓库则是用来读取和分析数据的，数据集市是数据仓库的一个子集，为特定的部门提供服务，而数据湖则充满了自然或原始格式的数据。

c) 数据准备

该步骤主要包括数据清洗、数据转换和数据整合。垃圾输入导致垃圾输出，这个步骤的重要性怎么强调都不为过。通常，这个步骤会反复进行，比如在数据建模阶段会发现错误，进而返回该步骤。

数据清洗的主要目的是消除数据中的错误，包括数据输入过程中的错误、冗余的空白、不可能的值、缺失值以及异常值等等。

数据转换的主要目的是使数据的格式满足模型的要求，比如将两点间的距离转换为横纵坐标之间的关系，或者将变量转换为虚拟变量。虚拟变量用于表明观测值的分类结果，如果属于某种类别，就置“1”，否则置“0”。比如将名为“月份”的观测值分为多个列，如果观测值在一月，就将数据的一月这一列置“1”，其他列置“0”。

数据整合主要指的是从不同的数据源整合数据，用到的一些手段有连接表、表添加、使用视图、丰富数据度等等。

d) 数据探索

数据探索阶段主要用到的是数据可视化技术，因为以图片的方式来展示数据和变量之间的交互更容易理解。

经常会用到的图形包括简单的线图或直方图、以及更为复杂的箱线图、桑基图等等。但是有时候将几个简单的图形组成一个复杂的图形更容易理解，或者多个图形关联更新等等。

当然也有其他的数据探索方法，比如制表、聚类或简单的建模。

e) 数据建模

在该阶段主要步骤如下：

- i. 特征工程和模型选择
- ii. 模型训练
- iii. 模型的验证与选择
- iv. 经过训练的模型应用在未知数据上

该阶段用到的技术包括机器学习、数据挖掘和统计领域等等。机器学习在该阶段的应用将在下一节中深入介绍。

f) 展示和自动化

该阶段你将主要将自己的成果展示给其他人并反复解释结果的意义，所以这就要求一定的相关领域专业知识。而且，如今各大公司对人工智能算法工程师的需求大大降低，更多地是在寻找和业务需求相匹配的工程师，所以将相关的专业知识和机器学习相结合愈加重要。

另外，成果模型可能会部署到实际的应用当中，所以可能会需要模型自动更新报告，即自动化。

4、机器学习在数据科学中的应用

机器学习在数据科学中的应用，主要有三个方向：

- 1) 回归：即预测，比如基于现有的数据预测之后的股价从而辅助交易分析员做出决策
- 2) 分类：比如在文本中识别人名和地名、图像识别等等

3) 根本原因分析：当业务的目标是更深入地理解一个现象的基本过程时，常常建立一个模型来完成此任务，比如理解和优化业务流程。

机器学习的类型：

1) 有监督学习

通常应用于有标注的数据集，需要人机交互来对数据进行标注。比如识别图像中的数字。

2) 无监督学习

不依赖于有标注的数据集，尝试在没有人机交互的情况下标注数据。

3) 半监督学习

需要有标注的数据集，也需要人机交互来对数据进行标注，但是通常从标注很少的数据集开始。

i. 特征工程和模型选择

进行特征工程，必须为模型选取并创造出可能的预测因子。模型最终将根据预测因子实现预测功能，所以这是很重要的一步。

在这里，介绍一下交互变量，任何一个单一的变量影响很小，但是如果两个变量同时存在，那么它们的影响就会很大。就像化学上的醋和漂白剂，单独用的话没有什么影响，但两者混合的话，就会产生有毒的氯气。

除此之外，还有易得性偏差：如果选取的特征只是那些容易获得的特征，那么构建的模型就会呈现出片面的倾向。

在这一阶段，通常用到的方法有 PCA 主成分分析法等等。

ii. 模型训练

在这一阶段，将清洗好的数据输入模型。

iii. 模型的验证

机器学习中有两种常见的误差衡量：

分类错误率（针对分类问题）；

均方误差（针对回归问题）；

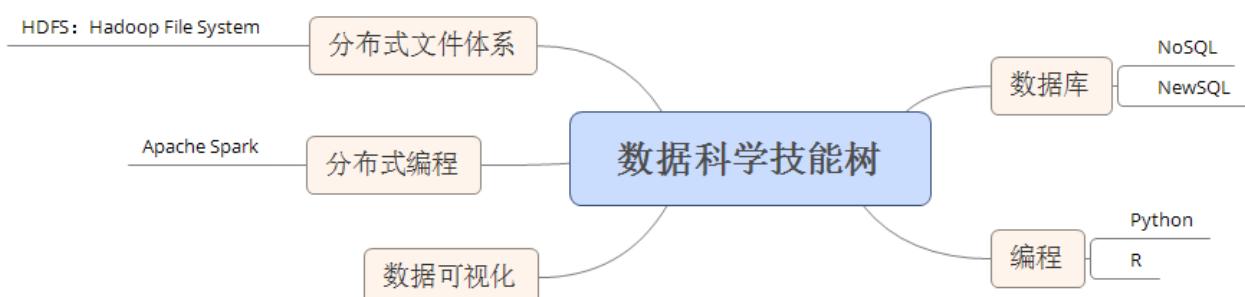
在这一阶段有很多验证策略，比如：

将观测值的 X% 划分为训练数据集，其余的部分作为测试数据。

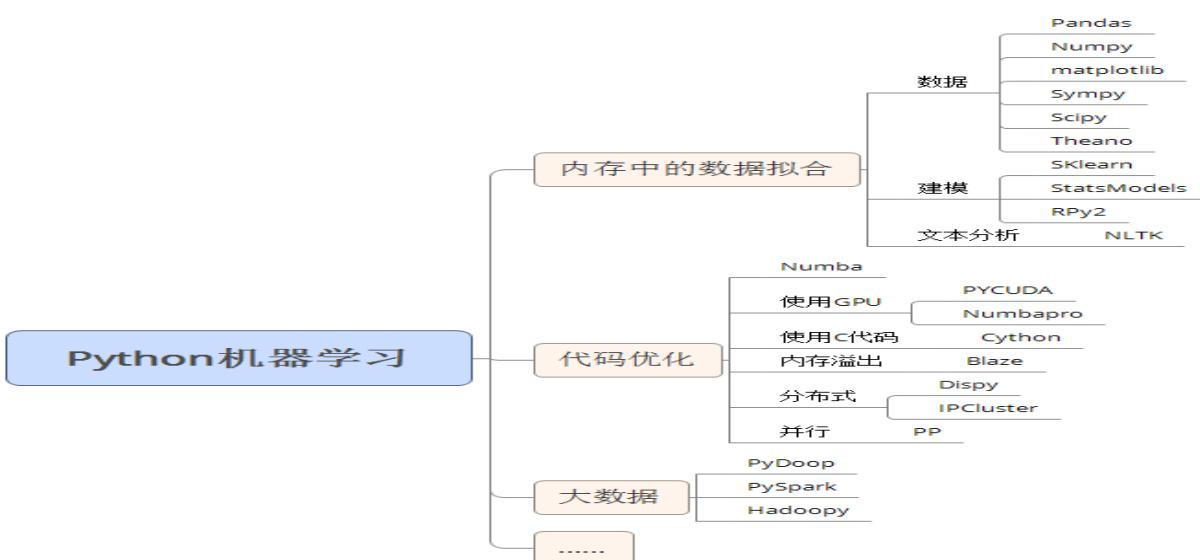
K 交叉验证——将数据集分为 K 个部分，每一部分都作为测试集，同时其他的部分作为训练数据集。这种方法的优势是数据集中所有的数据都是可用的。

iv. 经过训练的模型应用在未知数据上

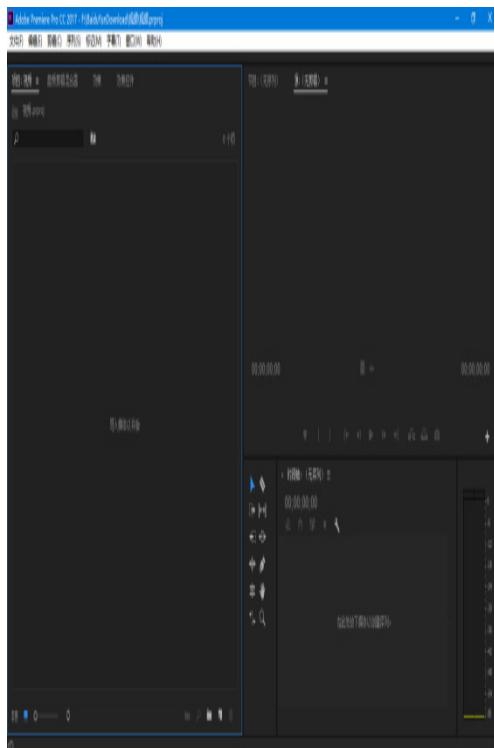
5、数据科学技能树



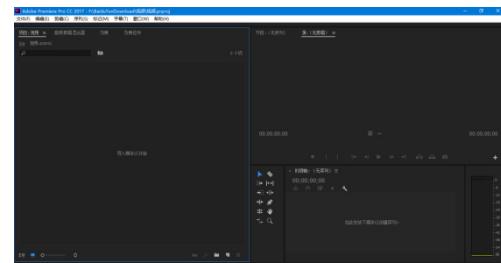
笔者在这里再列举一下 Python 在数据科学常用的库：



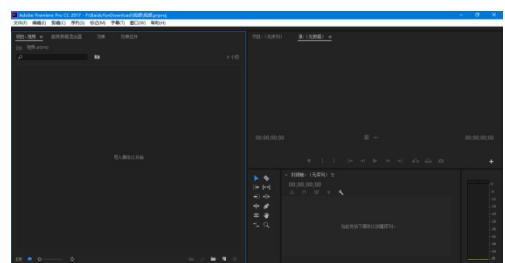
步骤一：本书通篇以一种趣味方式来叙述，大、
步骤二：量引用了各种各样的生活知识来类比，并充分运
步骤三：用图形语言来体现抽象内容，对数据结构所数量即可观看两点距离可达十六大学老师看了
涉及到的一些经典算法做到
步骤四：逐行分析、多算法比较。与市场上的同类数据结构图书相比，本书内容趣味易读，算法讲
解细致深刻，是一本非常适合自学的读物。



步骤一



步骤四



步骤三



步骤二



