# 数据类型

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | C | JAVA | LUA | PYTHON | PHP | R | JS |
| 数字 |  |  | 123 | 123 | 123 | 123 | 123 |
| 整数和浮点小数转换 |  |  |  | Int和float |  |  |  |
| 数字按一定小数位输出 |  |  | print(string.format("%.yf",  xx.xxxx)) | (1)print(round(xx.xxx)  ,y)（会四舍五入）  (2)print(“%.yf”%xx.xxx) |  | Print(round(xx.xxx,y)) |  |
| 字符串 |  |  | “123” or abc | “123” or abc | “123” or abc | “123” or abc | “123” or abc |
| 字符串和数字转换 |  |  | tonumber和tostring | Int和str |  | as.numeric和as.character |  |
| 字符串长度 |  |  |  | len(a)用于求字符串长度 | strlen(a)用于求字符串长度 | nchar(a) |  |
| 删除字符串末尾空白 |  |  |  | a = a.rstrip() |  |  |  |
| 字符串按符号拆分 |  |  | 拆分：Split(a,”xx”)  xx表示拆分符号 | 拆分：a.split(”xx”)  xx表示拆分符号 | 拆分：explode(“xx”,$a) | strsplit(string,”xx”) |  |
| 字符串拼接 |  |  | ..（两点） | +（加号） | .（一点） | Paste(string1,string2,sep=”xx”) | +（加号） |
| 字符串按字节切分 |  |  | a=”123bc”  a:sub(start,end)  a:sub(start,-1) 至末尾 | a=”123bc”  a[start:end+1:length]  a[start: ] 至末尾 | $a=”123bc”  substr($a,start,length) | substr(A,start=xx,stop=yy) |  |
| 判断某一字符是否在字符串中 |  |  | If a:find(xx)~=nil | If a.find(xx)~=-1 |  | grepl("xx",A) |  |
| 常见转义字符串 |  |  | [1]Tab: “\t”  [2]换行: “\n” | [1]Tab: “\t”  [2]换行: “\n” | [1]Tab: “\t”  [2]换行: “<br/>”(浏览器输出)或”\n”(先用<pre>定义) |  | [1]Tab: “\t”  [2]换行: “<br/>”或”<br>”(浏览器输出) |
| 布尔值 |  |  | false, true | False, True | False true | false true | False true |
| 空值 |  |  | nil (==false) | None | NULL (==false) | NA (==false) | NULL (==false) |
| 判断某一对象是否为空值 |  |  | If a==nil then | If a==None:  或者  If not a: |  | exists(“A”) |  |

\*补充：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | C | JAVA | LUA | PYTHON | PHP | R | JS |
| 数据类型是否需要预先声明 |  |  | 不需要 | 不需要 | 不需要 | 不需要 | 不需要 |
| 默认的局部变量 |  |  | 1.For 循环中自增（减）的i,j,k  2.函数内部变量 | 函数内部变量 | 函数内部变量 | 函数内部变量 | 函数内部变量 |
| 默认数字精度 |  |  | 双精度浮点数(14个小数位) | 双精度浮点数(17个小数位) | 双精度浮点数  (14个小数位) | 单精度浮点数(7个小数位) | 双精度浮点数  (17个小数位) |

补充：变量、变量值，地址、指针、索引号的联系和区别：

（1）变量，变量值和地址：变量值是变量的数据具体化，其在内存中的存放位置即为地址。地址是变量在计算机中的物理别名，两者是完全等价的。如下：

import ctypes

a=3 #a为变量，3为变量值

address\_a=id(a) #获取变量的内存地址

print(address\_a)

print(ctypes.cast(address\_a,ctypes.py\_object).value) #由地址来反求出变量值

140733696558816

3

（2）内存地址和指针：指针是用来保存变量内存地址数字的另一变量，简单情况下可理解为上面的address。需注意，由于指针也是变量，故它也有地址！

（3）内存地址和索引号：索引号是向量（数组）中用于索引元素的位置标记，显然它并不是向量（数组）中元素或变量的真实地址，而仅仅是抽象意义上的地址，但某种意义上两者可以等价看待，如下：

import ctypes

c=[1,2,3,4]

cc=[id(c[0]),id(c[1]),id(c[2]),id(c[3])]

for i in range(len(cc)):

    print(ctypes.cast(cc[i],ctypes.py\_object).value)

1

2

3

4

# 基础数据结构

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | C | JAVA | LUA | PYTHON | PHP | R | JS |
| 向量（数组） |  |  | {1,2,3} | [1,2,3] | [1,2,3] | c(1,2,3) | [1,2,3] |
| 向量定义 |  |  | B={1,2,3} | B=[1,2,3] | $B=[1,2,3]  $B=array(1,2,3) | B=c(1,2,3) | B=[1,2,3]  B=array(1,2,3) |
| 向量长度 |  |  | #B | len(B) | count($B) | length(B) | B.length |
| 向量拼接 |  |  |  | A+B |  | c(A,B) |  |
| 向量元素排序 |  |  | 正序：Table.sort(B)  逆序： | 正序：B.sort()  逆序：B.reverse() |  | Sort(B,decreasing=T) |  |
| 向量元素按符号拼接 |  |  | 拼接：table.concat(B,  ”xx”) xx表示拼接符号 | 拼接：concat(B,  ”xx”) xx表示拼接符号 |  |  |  |
| 向量内插入元素，删除元素，修改元素 |  |  | 插入：table.insert(B,sp, xxx)  删除：  table.remove(B,sp)  修改：  B[sp]=xx | 插入：B.insert(sp, xxx)  删除：del A[sp]  或A.pop(sp)（删除值可返回）  修改：A[sp]=xx |  |  |  |
| 向量入栈与出栈 |  |  | 入栈：B[#B+1]=xx  出栈：B[#B]=nil | 入栈：B.append(xx)  出栈：B.pop() | 入栈：B[count($B)+1]=xx  出栈：  B[count($B)]=NULL | 入栈：append(A, xx)  出栈：A[-length(A)] |  |
| 向量按“索引号”切分 |  |  |  | 切分：B[start:end+1]  \*全复制：B[:] | B[1:5] | 切分：B[start:end+1]  \*全复制：B[:] |  |
| 判断元素是否在向量（字典）中 |  |  | If A[i] then | If xx in A: |  | if("xx" %in% A) |  |
| 矩阵 |  |  | {{1,2},{3,4},{5,6}}  B[2][1]=3 | [[1,2],[3,4],[5,6]]  B[2][1]=3 | [[1,2],[3,4],[5,6]]  $B[2][1]=3 | [[1,2],[3,4],[5,6]]  注意：B[2.1]=3 | [[1,2],[3,4],[5,6]]  B[2][1]=3 |
| 矩阵定义 |  |  | B[i]={}  B={{1,2},{3,4},{5,6}} | B[i]=[]  B=[[1,2],[3,4],[5,6]] | $B[i]=[]  $B=[[1,2],[3,4],[5,6]] | C=c(1,2,3,4,5,6)  B=matrix(C,3,2) | B[i]=[]  B=[[1,2],[3,4],[5,6]] |
| 字典（散列表）定义 |  |  | B={[“a”]=1,[“b”]=2} | B={“a”:1,”b”:2} | $B=array(“a”=>1,  “b”=>2) | B=c("a"=1,"b"=2) | B=array(“a”=>1,  “b”=>2) |
| 字典添加，修改，删除键值对 |  |  | 添加：B[“c”]=3  修改：B[“c”]=4  删除：B[“c]=nil | 添加：B[“c”]=3  修改：B[“c”]=4  删除: del B[“c”] |  |  |  |
| 其他 |  |  | 向量索引号从1开始 | A=[1,2,3]，可知A中预先并不存在A[3]，那么A[3]就没有任何意义，则无论对A[3]赋值（A[3]=xx）还是输出A[3]均会报错！  Python中集合定义方式为：B={1,2,3}，特殊在于集合中的重复元素无意义。集合支持列表函数并且具备自身的集合运算函数，如 A |B 表示集合A，B的并集；A & B表示交集； A – B表示差集。  python严格区分列表，集合和字典的定义方式 |  | 还有诸如数据框，列表等独特结构  向量索引号从1开始 |  |

\*补充：数据的物理结构和逻辑结构

物理结构：数据在计算机中的真实存储方式

（1）顺序结构：各数据的在内存中的地址（可简单理解为索引号）是“相邻”（隔4或8个字节）且连续的。例如顺序列表

（2）链式结构：各数据（节点）在内存中的地址是任意的，但对任一节点，其地址仅仅取决于前（后）一个节点的指针。例如链式列表

（3）散列结构：各数据在内存中的地址通过散列函数，实现与其索引号（关键字）的一一对应，但各元素地址间并无直接联系。如散列表

逻辑结构：数据在程序或算法中的抽象关系

（1）集合：各数据（元素）同属于一个集合体，但彼此之间并无关系，简单来说就是各数据之间无先后的序关系。如字典

（2）线性表：除头元素和尾元素之外，各元素有且仅有一个前驱元素（上一位置或节点）和一个后继元素（下一位置或节点）。例如 列表，栈，队列，链表

（3）非线性表：各元素可以有一个或多个前驱元素和后继元素。如二叉树，图

\*线性表和非线性表均可由顺序结构和链式结构构造

特别地，对于线性表来说，当元素为数字时，还可以根据数的序关系分为有序（升序，降序）表和无序表，有序列表对于后续算法有巨大影响。

\*补充：链式结构（以单向链表为例）

# 定义节点（元素）类

class Node:

    def \_\_init\_\_(self,data,next=None):

        self.data=data

        self.next=next

# 定义链表类

class LinkList:

    def \_\_init\_\_(self):

        self.head=None        # 头指针

    def len(self):            # 求链表长度

        cur=self.head; length=0

        while cur!=None:

            length=length+1

            cur=cur.next

        return length

    def printf(self):            # 遍历链表

        node=A.head

        for i in range(A.len()):

            print(node.data)

            node=node.next

    def append(self,data):   # 在链表末尾添加节点

        node=Node(data)

        if self.len()==0:

            self.head=node

        else:

            cur=self.head

            pre=None

            while cur!=None:

                pre=cur

                cur=cur.next

            pre.next=node

    def insert(self,sp,data):  # 在指定位置插入新节点

        node=Node(data)

        if sp<=0:

            node.next=self.head

            self.head=node

        elif sp>=self.len():

            self.append(data)

        else:

            cur=self.head

            for i in range(sp-1):

                cur=cur.next

            node.next=cur.next

            cur.next=node

    def pop(self,sp=None):     # 删除指定位置的节点（若sp不填，则默认删除末尾节点）

        if sp==None:

            if self.len()>1:

                cur=self.head

                pre=None

                length=self.len()

                for i in range(length-1):

                    pre=cur

                    cur=cur.next

                pre.next=None

            else:

                self.head=None

        else:

            if self.len()-1<=sp:

                self.pop()

            else:

                cur=self.head

                pre=None

                for i in range(sp):

                    pre=cur

                    cur=cur.next

                pre.next=cur.next

                cur.next=None

    def insert\_1(node,data):   # 在指点节点node后插入一个新节点

        newnode=Node(data)

        node1=node.next

        node.next=newnode

        newnode.next=node1

    def pop\_1(node):          # 删除指定节点node后的那个节点

        node1=node.next

        node2=node1.next

        node.next=node2

        node1.next=None

    def found(self,data): # 判断链表中是否含有数据data的节点，有就返回True

        cur=self.head

        while cur!=None:

            if cur.data==data:

                return True

            cur=cur.next

        return False

A=LinkList()         # 创建链表

for i in range(5):

    A.append(i)

0 1 2 3 4

A.insert(2,7)        # 在第三个节点的位置插入新节点（数据为7）

0 1 7 2 3 4

A.pop(1)             # 删除第二个节点

0 7 2 3 4

A.printf()           # 遍历并输出链表各节点数据

print(A.len())       # 输出链表长度

5

print(A.found(6))    # 判断链表内是否含有某节点数据为6

False

# 基础语句

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | C | JAVA | LUA | PYTHON | PHP | R | JS |
| 注释 |  |  | 行：“- -“（两横杠）  块：”- - [[  ]]” | 行：“#”  块：”””xxx””” | 行：”//”  块：”/\*  \*/” | “#” | 行：”//”  块：”/\*  \*/” |
| 赋值 |  |  | a=1 (可一次性多变量赋值) | a=1 (可一次性多变量赋值) | $a=1 | a<-1或a=1 | a=1 |
| 赋值运算缩写 |  |  |  | x+=y: x=x+y  x++: 先赋值后运算  ++x: 先运算后赋值 | x+=y: x=x+y  x++: 先赋值后运算  ++x: 先运算后赋值 |  | x+=y: x=x+y  x++: 先赋值后运算  ++x: 先运算后赋值 |
| 输入 |  |  | x=io.read() | x=input()（默认为字符串，可用int()函数变为数字） |  | x=readline() |  |
| 输出 |  |  | print() 默认换行  io.write() 不换行 | print() 默认换行  print(xx,end=” “)不换行 | echo +内容默认输出，不换行);  print() ；sprintf()  向量print\_r() | cat(x) | document.write() 默认不换行 |
| 条件语句 |  |  | If i<=5 then  end  if……then  else  end  if……then  elseif…… then  else  end | If i<=5:  xxx  if……:  xxx  else:  xxx  if……:  xxx  elif……:  xxx  else:  xxx | if(&i<=5){  }  if(……){  }else{  }  if (……){  }else if(……){  }else{  }  switch{  case1……: ;break;  case2……: ;break;  } | if(i<=5){  }  if(……){  }else{  }  if (……){  }else if(……){  }else{  }  switch{  case1……: ;break;  case2……: ;break ;  } | if(i<=5){  }  if(……){  }else{  }  if (……){  }else if(……){  }else{  }  switch{  case1……: ;break;  case2……: ;break ;  } |
| 循环语句 |  |  | 循环句格式：  i=1,100  i=1,100,2  i=100,1,-1  for……do  end  while……do  end  repeat  until…… | i=1  while i<=100:  xxx  i=i+1  for i in range(0,100):  xxx (注意是10而非99)或range(99,-1,-1): | 循环句格式：  $i=1;$i<=100;$i++  $i=0;$i<=100;$i=$i+2  $i=100;$i>=1;$i=$i-1  for(……){  }  while(……){  } | 循环句格式：  i in 1:100  for(……){  }  while(……){  } | 循环句格式：  i=1;i<=100;i++  i=0;i<=100;i=i+2  i=100;i>=1;i=i-1  for(……){  }  while(……){  } |
| 遍历语句（依次逐个遍历所有元素，一般用于字典） |  |  | 遍历句格式：  key, value in pairs(c)  for……do  end | for key,value in A.items():  print(key,value)  或者  for key in A:  print(key,A[key]) | 遍历句格式：  $A as $key=>$value  foreach(……) {  } |  |  |
| 其他 |  |  |  |  |  |  |  |

\*补充：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | C | JAVA | LUA | PYTHON | PHP | R | JS |
| 单行语句结束标志 |  |  | 换行即结束 | 换行即结束 | 分号 ； | 换行即结束 | 换行即结束 |
| 是否严格遵循缩进规则 |  |  | 否 | 是 | 否 | 否 | 否 |

# 基础运算符

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | C | JAVA | LUA | PYTHON | PHP | R | JS |
| 算术运算符 |  |  | +, -, \*, /,^ | +, -, \*, /,\*\* | +, -, \*, /,\*\* | +,-,\*,/,^ | +,-,\*,/,\*\* |
|  |  |  | 模余数: a%b  模商: (a-a%b)/b  取整: Int() | 模余数: a%b  模商: a//b  取整: Int() | 模余数: a%b  模商: (a-a%b)/b  取整: (int) | 模余数: a%%b  模商: (a-a%%b)/b  向上取整:ceiling  向下取整:floor | 模余数: a%b  模商: (a-a%b)/b |
| 关系运算符 |  |  | >,>=,<,<= | >,>=,<,<= | >,>=,<,<= | >,>=,<,<= | >,>=,<,<= |
|  |  |  | ==, ~= | ==,!= | ==,!=  ===,!===（同时判断类型与内容） | ==,!= | ==,!= |
| 逻辑（布尔）运算符 |  |  | and, or, not  x=(x or y)  x为空时取y  x=(A and b) or c  A为假时取c，为真取b | and, or, not | &&, ||, ! | &&, ||, ! | &&, ||, ! |
| 三目运算符  if(b<=c) d=c; else d=b |  |  |  | d = c if b<=c else b | $d=$b<=$c ? $c:$b |  | d=b<=c ? c:b |
| 其他 |  |  | (a or 0)  a为空值时取0，否则取非空值（注意是空值None而不是未定义的值！） | (a or 0)  a为空值时取0，否则取非空值 |  |  |  |

快捷键补充：

Ctrl+Z 取消操作 Tab 多行缩进 Ctrl+Tab 多行退格 Ctrl+/ 多行注释（取消多行注释） Alt+Shift+鼠标左键定位来选中某几列

# 基础函数库

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | C | JAVA | LUA | PYTHON | PHP | R | JS |
| 函数 |  |  | f(a,b) | f(a,b) | f(a,b) | f(a,b) | f(a,b) |
| \*函数定义 |  |  | function f(a,b)  end | def f(a,b): | function f(a,b){  } | f<-function(a,b){  }  注意函数return()加括号 | function f(a,b){  } |
| 返回值格式 |  |  | return() | return |  |  |  |
| 数学函数格式 |  |  | math.xx() | math.xx() | xx() | xx() | Math.xx() |

\*补充

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | C | C++ | LUA | PYTHON | PHP | R | JS |
| 安装函数包 |  |  |  | Cmd: pip install xx  (删除则为uninstall)  更新pip：pip install --user --upgrade pip |  | 建议用Rstudio下载 |  |
| 调用并查看函数包 |  |  |  | 调用包: import xx  调用包中指定函数：form 包 import 函数  查看包: help(dir(xx)) |  | 建议用Rstudio函数库查看 |  |
| 调用其他文件  （注意主程序和包位于同一父目录下！） |  |  | require(“路径/包”) | Import 路径.包（模块）  或 from 路径.包 import 函数 |  | Library（路径/包） |  |

# 文件（数据集）I/O操作：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | C | JAVA | LUA | PYTHON | PHP | R | JS |
| 读取文件 |  |  | file1=assert(io.open(“xxx”,”r”))  file1:read()  file1:close() | file1= open(“xxx”,”r”,encoding=”utf-8”)  file1.read() 读取全部  file1.readline() 只读取一行  file1.close() | $file1=fopen(“xxx”,”r”);  $a=fgets($file1,size);  （按行读取字符串）  $a=fread($file1,size);（读取定长字符串）  fclose($file); | data1=read.csv(“xxx”) |  |
| 写入文件 |  |  | file1=assert(io.open(“xxx”,”w”))  file1:write()  file1:close() | file1= open(“xxx”,”w”,encoding=”utf-8”)  file1.write()  file1.close() | $file1=fopen(“xxx”,”r”);  fwrite($file1,内容);  fclose($file); | write.csv(xx, file=”yy”) |  |
| 其他 |  |  |  |  | $file1=fopen(“xxx”,”a”);用于追加文件内容 |  |  |

# 面向对象编程简介

“面向过程”与“面向对象”是编程的两种最基本思想。

不妨将其看作两名程序员，在解决问题之前，“面向过程”会分析出解决问题的大致思路，并将其拆解为一个个步骤，为实现前后步骤的目的，则只需按步骤之间的逻辑关系依次调用函数即可，因此可以说“面向过程”是解决一般问题的最最基本的编程思想；而“面向对象”则是把问题的主体分解为各个对象（也可以说问题的不同考虑方面），对象不是需要实现的步骤，而是相互联系的事物，它们用于担负起在问题解决过程中自身的某一方面的职责。

关于“面向过程”我们无需多作介绍，下面我们先来引入“面向对象”中最重要的类和对象的概念，在这之后，我们会通过举例子来对比“面向过程”和“面向对象”两种编程思想，从而得出各自的优缺点。

1 类和对象

先定义一个关于狗狗的类，其中包括狗狗的属性，以及一个关于自我介绍的方法或函数：

class Dog:                                 # 类：狗狗

    def \_\_init\_\_(self,name,sex,age):       # 属性（变量）： 名字、性别、年龄

        self.name=name

        self.sex=sex

        self.age=age

    def Introduction(self):                # 方法（函数）：作自我介绍

        print("The dog"+self.name+"is a"+self.sex+"dog,and"+str(self.age)+

        "years old now")

# 每个类均由属性（变量）和方法（函数）两部分组成，其中特殊如方法\_init\_是用来初始化或定义类属性

# 注意，参数self仅用来封装和传递类属性，自身并无任何含义

将类实例化为对象，进一步访问其属性值和调用其方法：

dog1=Dog("Bob","male",3)                    # 类实例化为不同的对象

dog2=Dog("Mary","female",2)

print(dog1.name,dog1.sex,dog1.age)          # 访问不同对象的属性值

print(dog2.name,dog2.sex,dog2.age)

dog1.Introduction()                           # 调用不同对象的方法

dog2.Introduction()

# 类的实例化本质上是调用方法\_init\_并对其中的参数赋值，python中方法\_init\_不作显示

# 由于类已经实例化为对象了，那么当调用某一对象的方法时，参数self（name、sex、age）已知。python中参数self不作显示

Bob male 3

Mary female 2

The dog Bob is a male dog,and 3 years old now

The dog Mary is a female dog,and 2 years old now

允许类中存在公共属性，实例化之后，所有对象公共属性值默认相同。同时允许某一对象修改该属性值

class Dog:

    def \_\_init\_\_(self,name,age):

        self.name=name

        self.sex="male"                         # 公共属性 sex=“male”

        self.age=age

    def Introduction(self):

        print("The dog "+self.name+" is a "+self.sex+" dog,and "+str(self.age)+

        " years old now")

dog1=Dog("Bob",3); print(dog1.sex)

dog2=Dog("Mary",2); print(dog2.sex)    # 类的实例化过程中，公共属性无需赋值（即使赋值为female也不作用）

dog2.sex="female"; print(dog2.sex,dog1.sex)     # 对某一对象修改属性sex的值，但不影响其他对象

male

male

female male

更新某一对象中公共属性值也可采用如下方法：

class Dog:

    def \_\_init\_\_(self,name,age):

        self.name=name

        self.sex="male"

        self.age=age

    def Introduction(self):

        print("The dog "+self.name+" is a "+self.sex+" dog,and "+str(self.age)+

        " years old now")

    def Update\_sex(self,new\_sex):

        self.sex=new\_sex

dog1=Dog("Bob",3); print(dog1.sex)

dog2=Dog("Mary",2); print(dog2.sex)

dog2.Update\_sex("female"); print(dog2.sex,dog1.sex)

male

male

female male

class Dog:

    def \_\_init\_\_(self,name,age,sex="male"):          # 属性sex若未赋值，则实例化后默认为male

        self.name=name

        self.sex=sex

        self.age=age

    def Introduction(self):

        print("The dog "+self.name+" is a "+self.sex+" dog,and "+str(self.age)+

        " years old now")

dog1=Dog("Bob",3); print(dog1.sex)

dog2=Dog("Mary",2); print(dog2.sex)

dog2=Dog("Mary",2,"female"); print(dog2.sex,dog1.sex)

male

male

female male

最后我们简单来介绍类的继承和调用这一概念：

class Dog:

    def \_\_init\_\_(self,name,age):

        self.name=name

        self.sex=None

        self.age=age

    def Introduction(self):

        print("The dog "+self.name+" is a "+self.sex+" dog,and "+str(self.age)+

        " years old now")

class Movement:

    def \_\_init\_\_(self,sex,movement="running"):

        self.sex=sex

        self.movement=movement

    def describe(self):

        if self.sex=="male":

            print("Look, he is "+self.movement)

        else:

            print("Look, she is "+self.movement)

class Male\_Dog(Dog):                             # 子类Male\_Dog继承父类Dog中所有方法(\_init\_和Introduction)

    def \_\_init\_\_(self, name, age,action):

        super().\_\_init\_\_(name, age)              # 方法super().\_\_init\_\_使子类继承父类中所有非公共属性

        self.sex="male"                          # 在子类中增加属性sex="male"

        self.action=action

        self.move=Movement(self.sex,action)       # 将类Movement实例化，并作为Male\_Dog的一个属性来调用

class Female\_Dog(Dog):                             # 同理

    def \_\_init\_\_(self, name, age,action):

        super().\_\_init\_\_(name, age)

        self.sex="female"

        self.action=action

        self.move=Movement(self.sex,action)

dog=Male\_Dog("Tom",4,"sitting")

dog.Introduction()

dog.move.describe()

dog=Female\_Dog("Judy",3,"sleeping")

dog.Introduction()

dog.move.describe()

The dog Tom is a male dog,and 4 years old now

Look, he is sitting

The dog Judy is a female dog,and 3 years old now

Look, she is sleeping

2 面向过程与面向对象编程的特点：

例子：求e的1000位小数并打印出来

def concat(Table,s1,start,end):                                # 列表元素拼接

    s=str(Table[0])

    for i in range(start+1,end+1):

        s=s+s1+str(Table[i])

    return s

#面向过程编程

M=[1]; N=[1]; sp=0; dividend=1

while sp<=999:                                                 # 若小数位小于1000位，则继续运算；反之停止运算

    for j in range(sp,1010):

        if len(N)-1>=j+1: N[j+1]=N[j+1]+(N[j]%dividend)\*10     # 等价于N[j+1]=(N[j+1] or 0)+(N[j]%dividend)\*10

        else: N.append((N[j]%dividend)\*10)

        N[j]=int((N[j]-N[j]%dividend)/dividend)

        if len(M)-1>=j: M[j]=M[j]+N[j]

        else: M.append(N[j])

    while N[sp]==0:

        sp=sp+1

    dividend=dividend+1

for j in range(1009,-1,-1):                                   # 十进制调整

    M[j-1]=M[j-1]+int((M[j]-M[j]%10)/10)

    M[j]=M[j]%10

print(concat(M,"",0,999))

优点：遵循自顶向下的串联式逻辑，程序简洁，效率高，对于复杂问题的处理有一定优势

缺点：对于类似问题较难直接推广，有时需要作较大改动

def concat(Table,s1,start,end):                            # 列表元素拼接

    s=str(Table[0])

    for i in range(start+1,end+1):

        s=s+s1+str(Table[i])

    return s

#面向对象编程

class Division:                                            # 数组除法运算

    def \_\_init\_\_(self,Arr,dividend=1):

        self.Arr=Arr

        self.dividend=dividend

    def division(self):

        for j in range(len(self.Arr)-1):

            self.Arr[j+1]=self.Arr[j+1]+(self.Arr[j]%self.dividend)\*10

            self.Arr[j]=int((self.Arr[j]-self.Arr[j]%self.dividend)/self.dividend)

        return self.Arr

class Product:                                            # 数组乘法运算

    def \_\_init\_\_(self,Arr,exponent):

        self.Arr=Arr

        self.exponent=exponent

    def product(self):

        for j in range(len(self.Arr)):

            self.Arr[j]=self.Arr[j]\*self.exponent

        for j in range(len(self.Arr)-1,-1,-1):

            self.Arr[j-1]=self.Arr[j-1]+int((self.Arr[j]-self.Arr[j]%10)/10)

            self.Arr[j]=self.Arr[j]%10

        return self.Arr

class Sum:                                              # 数组之间加和运算

    def \_\_init\_\_(self,Arr,Arr1):

        self.Arr=Arr

        self.Arr1=Arr1

    def sum(self):

        for j in range(len(self.Arr1)):

            self.Arr1[j]=self.Arr[j]+self.Arr1[j]

        for j in range(len(self.Arr1)-1,-1,-1):

            self.Arr1[j-1]=self.Arr1[j-1]+int((self.Arr1[j]-self.Arr1[j]%10)/10)

            self.Arr1[j]=self.Arr1[j]%10

        return self.Arr1

class Count:                                            # 找出数组元素的非零起始位置

    def \_\_init\_\_(self, Arr):

        self.Arr=Arr

        self.sp=0

    def update\_sp(self):

        while self.Arr[self.sp]==0 and self.sp<1009:

            self.sp=self.sp+1

        return self.sp

class Compute:                                          # 计算e的任一指数的1000位小数

    def \_\_init\_\_(self,Arr,Arr1,exponent):

        self.Arr=Arr

        self.Arr1=Arr1

        self.exponent=exponent

        self.dividend=1

        self.sp=Count(self.Arr).update\_sp()

    def compute(self):

        while self.sp<=999:

            self.Arr=Product(self.Arr, self.exponent).product()[:]

            self.Arr=Division(self.Arr, self.dividend).division()[:]

            self.Arr1=Sum(self.Arr, self.Arr1).sum()[:]

            self.sp=Count(self.Arr).update\_sp()

            self.dividend=self.dividend+1

        print(concat(self.Arr1,"",0,999))

        return self.sp

Arr=[1]; Arr1=[1]; exponent=1

for j in range(1,1010): Arr.append(0); Arr1.append(0)

A=Compute(Arr,Arr1,exponent)

A.compute()

优点：遵循分布式逻辑，可扩展性较强，处理类似问题仅需要对部分程序做修改

缺点：程序较为复杂，对于某些复杂问题不易处理

# 递归编程及其他编程思想简介

递归编程基本思想：

在面向对象编程中，我们学会了将待解决的问题主体分为各个对象，每个对象担负起部分功能，然后根据各个对象之间的联系来相互调用，并最终解决问题。现在，我们要学习一种新的编程思想—递归编程。通过定义递归函数，递归编程可以将最初的大问题拆解为许许多多性质相同的中间问题，所有中间问题的关系类似于“树”的结构，把大问题比作“树顶”，大问题逐渐向下拆解，在抵达“树底”之后，再逐步递归，返回“树顶”。

递归编程基本条件：

（1）必须有递归终止条件，不然会永远执行下去！

（2）必须使得某一状态逐渐接近终止条件，类比于由“树顶”向“树底”靠拢。

（3）在执行递归函数并使上述状态接近终止条件过程中，必须不断调用函数自身

我们举个简单的例子来展示递归编程的思想：

# 用“递归”代替“循环”计算 1+2+3+……+9

# 循环方法

sum=0

for i in range(1,10):

sum=sum+i

# 递归方法：通过执行sum(n-1)使得n不断减小靠近1，在n=1之后，开始朝着n增大的方向返回 sum(n-1)+n的值：1、1+2、1+2+3………

def sum(n):

    if n<=1: # 满足终止条件，返回sum(1) = 1的值至sum(2)

        return 1

return sum(n-1)+n # 满足终止条件前，一直执行sum(n-1)；在满足条件后，不断返回sum(n) = sum(n-1) + n的值至上一层

n:  10 -> 9 -> 8 -> 7 -> 6 -> 5 -> 4 -> 3 -> 2 -> 1（抵达终止条件）

   (1+2+3+……+10) <- …… <- (1+2+3) <- (1+2) <- (1) :sum （开始返回sum(n-1)+n的值）

可以看到相比于一般方法，递归编程并未使用循环语句，但同样计算出相同的结果！在后续的应用中，我们会看到，递归编程不仅能够化简代码，往往还能解决常规编程方法所解决不了的问题！

# （1）十进制换二进制

def Ten\_to\_Two(n):

    if n<=1:

        print(int(n),end="")

    else:

        Ten\_to\_Two((n-n%2)/2)

        print(int(n%2),end="")

Ten\_to\_Two(18)

10010

# （2）汉诺塔问题:共有n个盘子，求把所有盘子正确地从柱子A移到柱子C的操作？

# 思路：先把前n-1个盘子正确移到柱子B，然后把第n个盘子移到柱子C，再把前n-1个盘子从柱子B移到柱子C

def MoveTower(n,start="A",middle="B",end="C"):

    if n==1:

        print(start+"->"+end)  # 终止条件：只有一个盘子时，停止调用函数，直接将盘子移到目标柱子

    else:

        MoveTower(n-1,start,end,middle)

        MoveTower(1,start,middle,end)

        MoveTower(n-1,middle,start,end)

MoveTower(5)

A->C

A->B

C->B

A->C

B->A

B->C

A->C

A->B

C->B

C->A

B->A

C->B

A->C

A->B

C->B

A->C

B->A

B->C

A->C

B->A

C->B

C->A

B->A

B->C

A->C

A->B

C->B

A->C

B->A

B->C0

A->C

最后我们简单介绍一下其他常见的编程思想：

（1）贪婪算法：贪婪算法旨在获取小问题的局部最优解，进而获取大问题的全局最优解或近似全局最优解，但获取最优解的优劣要看贪婪策略的选择，如下：

#背包问题：

"""

一小偷携带负荷20的背包来珠宝店偷东西，现列出物品重量和价值以供考虑:

(1,3),(2,4),(3,5),(5,6),(6,8),(8,10),(12,15),(13,20)，问小偷采取如下何种贪婪策略，可偷取得最大价值:

【1】在不超过负荷情况下，每次价值最大【2】在不超过负荷情况下，每次平均价值最大

"""

max=20

Goods\_Value=[[1,3],[2,4],[3,5],[5,6],[6,8],[8,10],[12,15],[13,20]]

def f1(max, Goods\_Value):  # 策略1：每次偷取价值最大

    sum\_value=0; sum\_goods=0

    for i in range(len(Goods\_Value)-1,-1,-1):

        sum\_value=sum\_value+Goods\_Value[i][1]

        sum\_goods=sum\_goods+Goods\_Value[i][0]

        if sum\_goods>max:

            sum\_value=sum\_value-Goods\_Value[i][1]

            sum\_goods=sum\_goods-Goods\_Value[i][0]

        else:

            print("("+str(Goods\_Value[i][0])+","+str(Goods\_Value[i][1])+")")

    return sum\_goods,sum\_value

def f2(max, Goods\_Value):  # 策略2：每次偷取平均价值最大

    sum\_value=0; sum\_goods=0

    for i in range(len(Goods\_Value)):

        Goods\_Value[i].append(Goods\_Value[i][1]/Goods\_Value[i][0])

    Goods\_Value.sort(key=(lambda x:x[2]))  # 对二维列表按照关键字x[2]倒序排序

    for i in range(len(Goods\_Value)-1,-1,-1):

        sum\_value=sum\_value+Goods\_Value[i][1]

        sum\_goods=sum\_goods+Goods\_Value[i][0]

        if sum\_goods>max:

            sum\_value=sum\_value-Goods\_Value[i][1]

            sum\_goods=sum\_goods-Goods\_Value[i][0]

        else:

            print("("+str(Goods\_Value[i][0])+","+str(Goods\_Value[i][1])+")")

    return sum\_goods,sum\_value

print(f1(max,Goods\_Value))

(13,20)

(6,8)

(1,3)

(sum\_goods=20, sum\_value=31)

print(f2(max,Goods\_Value))

(1,3)

(2,4)

(3,5)

(13,20)

(sum\_goods=19, sum\_value=32)

（2）动态规划算法：相比于贪婪算法中将大问题分为小问题，并求每个小问题的最优解的方法，动态规划算法将某个小问题逐步扩展为大问题，并在这一过程始终求其最优解，那么可想而知，最后求得的最优解一定是全局最优解！

"""

同样是背包问题，在上题中我们还可以用动态规划算法来求解，定义动态规划矩阵，横行代表所考虑的物品个数，从物品(1,3)直到全部物品

；纵列代表背包负荷的变化，从负荷为1到负荷为20。以i行j列为例，此时我们需要的问题是:在考虑前i个物品且负荷为j的情况下，可偷取

的最大价值。不难得知当G[i][0]>=j时，A[i][j]=max(A[i-1][j], G[i][1]+A[i-1][j-G[i][0]])

"""

max=20

Goods\_Value=[[1,3],[2,4],[3,5],[5,6],[6,8],[8,10],[12,15],[13,20]]

def f(max,Goods\_Value):

    A=[]; length=len(Goods\_Value)

    for i in range(length+1):

        A.append([])

        for j in range(max+1):

            A[i].append(0)

    for i in range(1,length+1):

        for j in range(1,max+1):

            A[i][j]=A[i-1][j]

            if Goods\_Value[i-1][0]<=j and A[i-1][j]<Goods\_Value[i-1][1]+A[i-1][j-Goods\_Value[i-1][0]]:

                A[i][j]=Goods\_Value[i-1][1]+A[i-1][j-Goods\_Value[i-1][0]]

    for i in range(length,-1,-1): # 根据动态规划矩阵A的结果，我们从A[length][max]一直回溯至A[1][1]，判断G[i-1]是否被偷取

        if A[i][max]==Goods\_Value[i-1][1]+A[i-1][j-Goods\_Value[i-1][0]]:

            max=max-Goods\_Value[i-1][0]

            print(Goods\_Value[i-1])

f(max,Goods\_Value)

[13, 20]

[3, 5]

[2, 4]

[1, 3]

（3）分治算法：可以说是最常用的一种编程思想，将大问题分为小问题，小问题又继续分为更小的问题。解决每一个小问题，最终即解决了大问题。

# 二分查找问题

A2=[1,2,3,5,5,8,9,12,23,43,45,64]; item=13

def find\_21(List,item): # 非递归

    flag=False; first=0; last=len(List)-1

    while first<=last and flag==False:

        mid=(first+last)//2  # 模运算

        if List[mid]==item:

            flag=True

        else:

            if List[mid]<item:

                first=first+1

            else:

                last=last-1

    return flag

print(find\_21(A2,item))

（4）回溯算法：在处理非线性的树和图问题时，我们往往需要沿着一条路径向下走，如果能走通，则问题解决，反之需要回到上一步并继续尝试，通过大量的穷举试验，我们最终总能找到解决问题的一条路径。回溯算法的应用与递归编程思想有十分紧密的关系，如下：

# 马的遍历问题:nxm的棋盘，选定棋盘上一点(x,y)并按照“马”的规则走，规划出该点不重复地走过棋盘上任一位置的路线？

class Check: # 用于判断走位的正确性

    def \_\_init\_\_(self,n,m,x,y):

        self.n=n

        self.m=m

        self.x=x

        self.y=y

        self.A=[]

        for i in range(self.n):

            self.A.append([])

            for j in range(self.m):

                self.A[i].append(0)

    def check(self):

        if (self.x<1 or self.x>self.m) or (self.y<1 or self.y>self.n) or self.A[self.x-1][self.y-1]!=0:

            return False

        return True

class Run(Check): # 用于递归走位

    def \_\_init\_\_(self, n, m, x, y):

        super().\_\_init\_\_(n, m, x, y)

        self.step=None

        self.D\_x=[-2,-2,2,2,1,-1,1,-1]

        self.D\_y=[1,-1,1,-1,-2,-2,2,2]

    def run(self):

        for i in range(8):

            if self.check():

                self.A[self.x-1][self.y-1]=self.step

                if self.step==self.n\*self.m:

                    return self.A

                else:

                    self.step=self.step+1

                    self.run()

                    self.step=self.step-1

                    self.A[self.x-1][self.y-1]=0

            else:

                self.x=self.x-self.D\_x[i]; self.y=self.y-self.D\_y[i]

class Game(Run): # 用于初始化棋盘和起始点，并输出路线

    def \_\_init\_\_(self, n, m, x, y):

        super().\_\_init\_\_(n, m, x, y)

        self.A[self.x-1][self.y-1]=1

        self.step=2

    def game(self):

        if self.run()!=None:

            for i in range(len(self.A)):

                print(concat(self.A[i],"\t"))

        else:

            print("try again!")

p=Game(5,5,1,1)

p.run()

1       14      3       8       21

4       9       20      13      16

19      2       15      22      7

10      5       24      17      12

25      18      11      6       23

为了避免重复遍历，通过某种“剪枝方法”，回溯算法可达到以空间换时间，从而实现优化，此处不做过多介绍。

# 程序异常与调试

Python中程序异常可分为两种，第一种是语法错误（通常编辑器会提示），第二种是逻辑错误（需要自己调试发现），如下：

#语法错误

print 3

File "c:\Users\熊高庆\Desktop\项目文件\Python\Python程序\Learning\Datebase Learning\training.py", line 1

    print 3

          ^

SyntaxError: Missing parentheses in call to 'print'. Did you mean print(3)?

File指代程序异常的文件

line1表示错误出现在第一行，并给该行代码

SyntaxError表示出现语法错误，错误信息Missing parentheses表示缺少圆括号

Did you mean print(3)? 给出了修改建议

#逻辑错误

print(5/0)

Traceback (most recent call last):

  File "c:\Users\熊高庆\Desktop\项目文件\Python\Python程序\Learning\Datebase Learning\training.py", line 3, in <module>

    print(5/0)

ZeroDivisionError: division by zero

Traceback指异常信息追溯

File 和 line同理

ZeroDivisionError：division by zero 表示犯了“零除”错误

在一个程序中，对于可能出现异常的代码，我们可用try……except语句来处理，如下：

#处理前

x=int(input()); y=int(input())

print(x/y)

print(x\*y)

4

0

Traceback (most recent call last):

  File "c:\Users\熊高庆\Desktop\项目文件\Python\Python程序\Learning\Datebase Learning\tempCodeRunnerFile.py", line 2, in <module>

    print(x/y)

ZeroDivisionError: division by zero

当y取0时，x/y则犯了零除错误，从而给出异常信息并终止程序

#处理后

x=int(input()); y=int(input())

try:

    print(x/y)

except ZeroDivisionError:                 # ZeroDivisionError也可换成一般的Exception

    print("divide by 0")

    print(0)

print(x\*y)

4

0

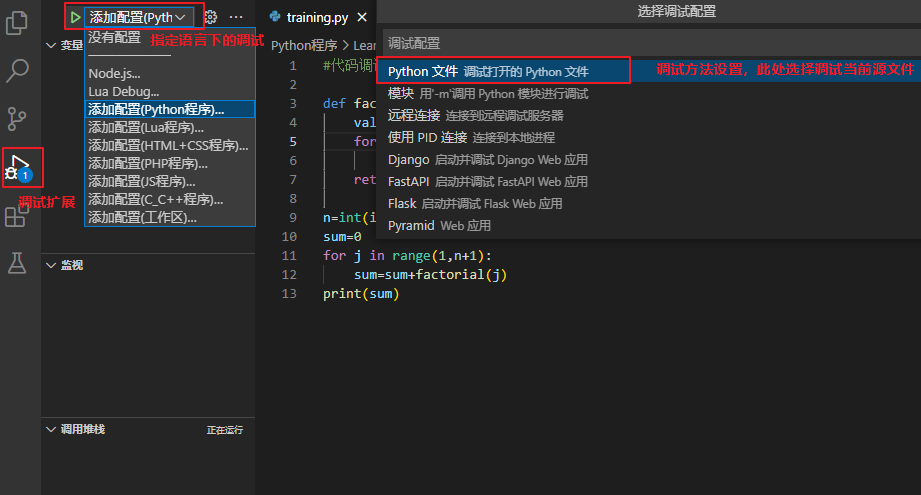
divide by 0

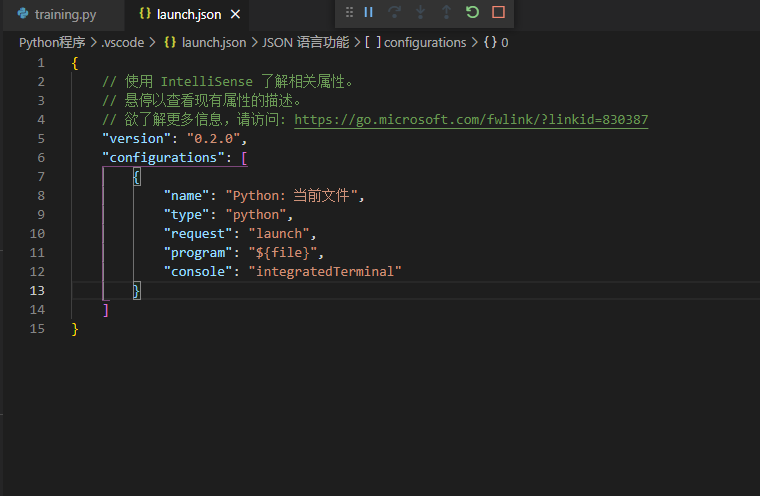
0

0

我们将可能出现异常的代码块用 try语句包含，然后用except语句给出异常信息提示，以及异常时的默认值0，这样做的好处是不会因为出现异常就终止程序，后面的print(x\*y)同样可以执行。关于具体的异常信息以及try……except语句的更多用法，我们不作更多介绍。

当出现程序异常时，最好的处理方法当然是调试代码，并将异常修正。下面我们给出Vscode中代码调试的方法：





此处会生成一个关于调试信息的文件，其中：

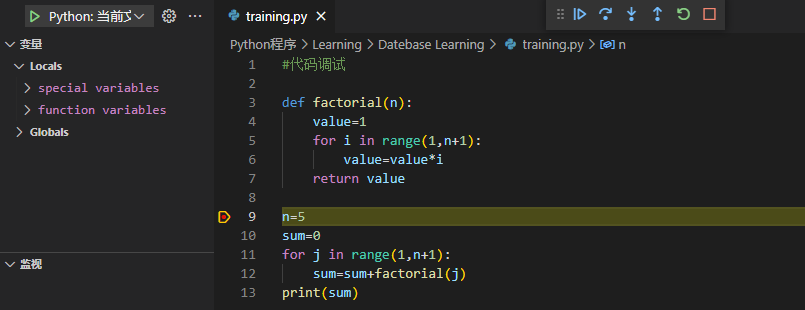
Name：表示当前调试配置的名称

Type：表示对哪种语言进行调试

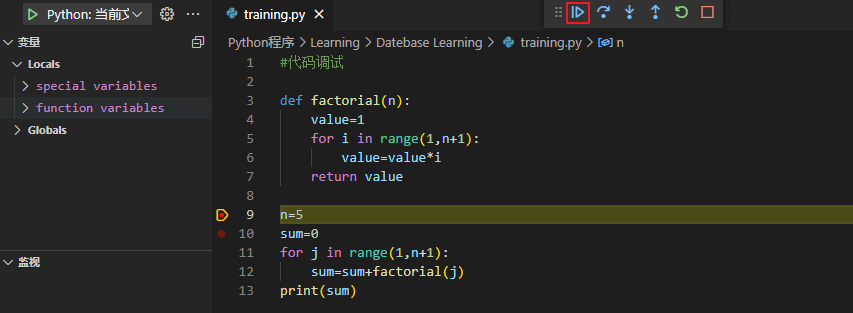
Request：调试方式，通常为launch

Program：调试文件的绝对路径

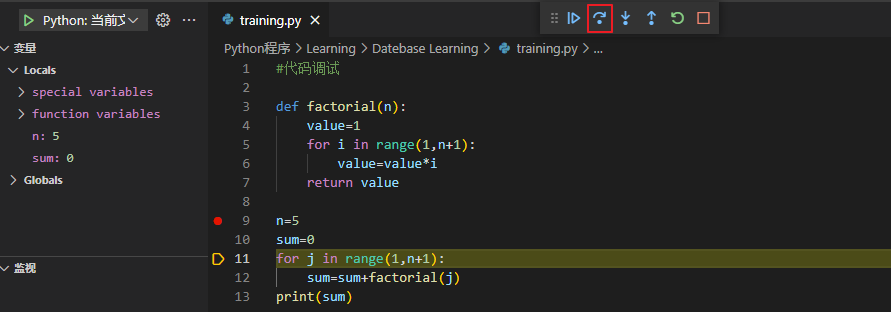
Console：终端类型，通常为IntegratedTerminal

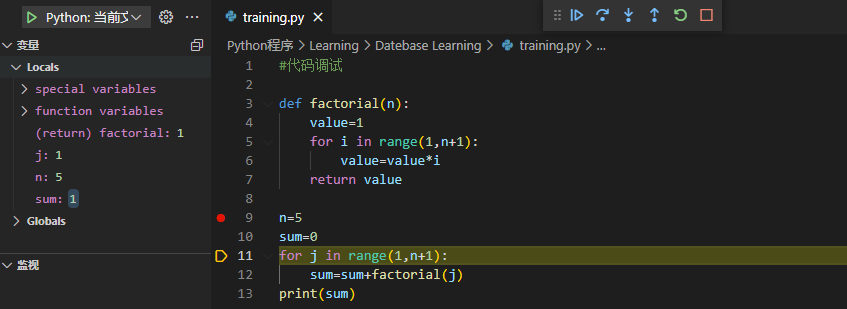


在第九行设置一个断点，则进入调试后，代码会直接运行到断点的前一行（第八行），此时变量n的信息在左侧还未显示

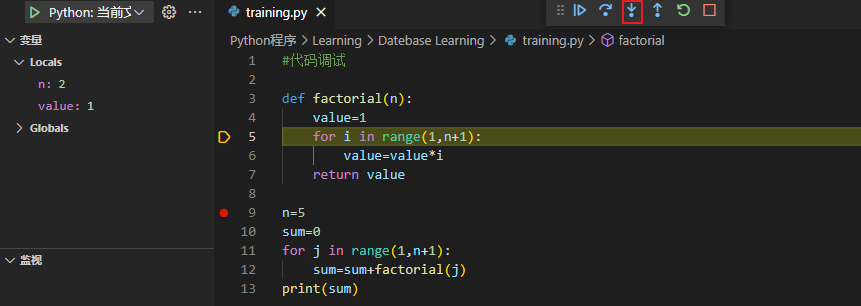


若点击“继续”，则直接运行断点及以后的全部程序，调试结束并在终端打印出sum的结果，此时左侧并不显示调试过程中变量值的信息，因此在调试中一般不直接点击“继续”。

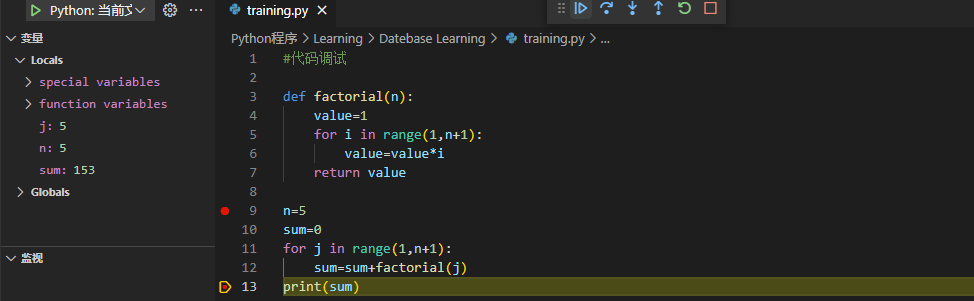




若点击“单步跳过”，则当程序运行至第十一行时（实际运行了前十行），变量n和sum的信息给出。进一步执行进入循环，除变量n和sum外，左侧还给出了变量j以及j的函数factorial的信息。



若点击“单步调试”，则程序会由执行逻辑依次运行，此时调试可进入函数factorial内部，并在左侧给出相应变量的信息。



除此之外，还可以设置多个断点来进行调试！

# 算法分析

算法是用于解决问题的一种方法或思路，而程序是算法的具体化，不仅包括了算法，还关系到用哪种语言，程序如何设计，甚至是用哪种编译器，哪台计算机等等。因此一个程序的好坏不仅与算法有关，还涉及到其他各各方面。

在实际应用中，我们更关心程序的好坏评价，一般是看它的实际运行时间和所占用的内存，CPU等等，如下

import time

start = time.time()    # 查看当前时间，以秒为单位

#---------------------------

def factorial(n):

    value=1

    for i in range(1,n+1):

        value=value\*i

    return value

n=5

sum=0

for j in range(1,n+1):

    sum=sum+factorial(j)

print(sum)

#--------------------------

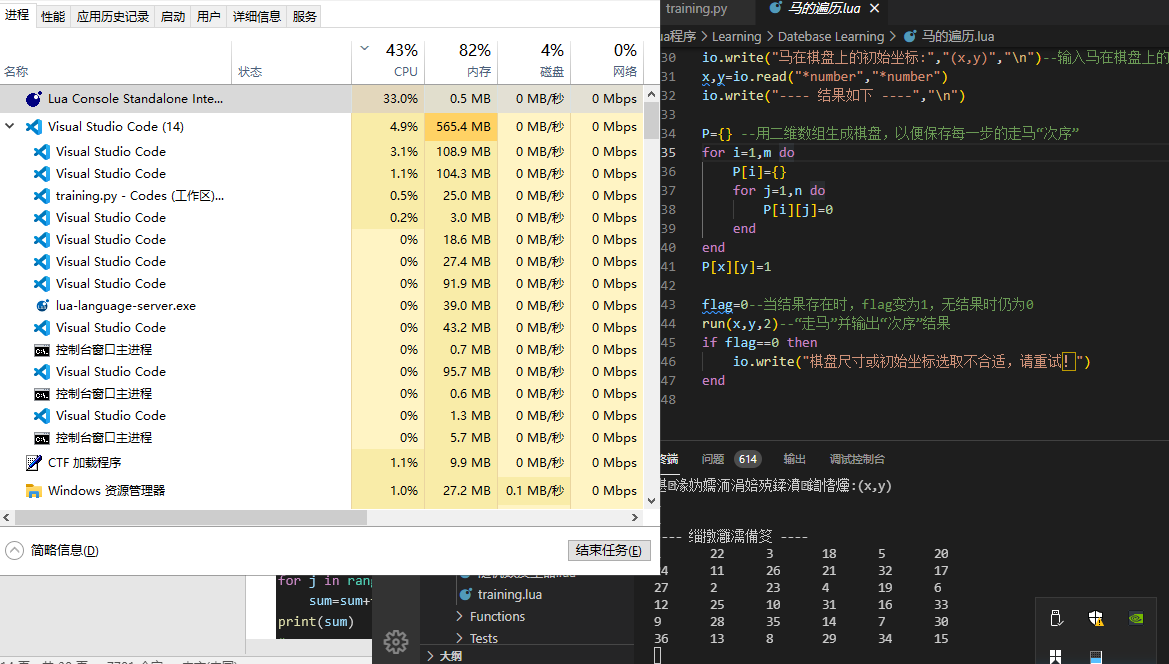
end = time.time()

print(end-start)      # 输出起止时间之差

153

0.0009839534759521484

可见程序运行时间大约为0.00098秒，但需要注意这不一定是程序运行的真实时间，有可能会受到当前正在运行的其他应用的影响。

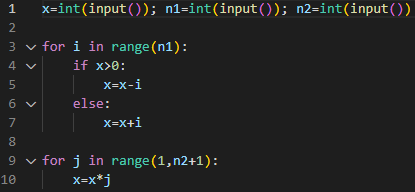


针对某些运行时长较长的程序，可用ctrl+shift+esc打开任务资源管理器，动态监控在当前计算机下，程序运行在当前计算机下所占内存和CPU。

而在算法研究中，排除其他外在因素，为了对比程序运行前算法的优劣，我们需要关注两个指标，时间复杂度和空间复杂度。

1. 时间复杂度：

对于某个程序，根据它每一行代码是否涉及到cpu运算，我们可提取出所有的有效代码行。然后再计算每一行的执行次数（考虑到循环，条件语句等特殊结构），最终加和可得该程序的总执行次数T(n)，示例如下：



可知其有效代码行为3，5，7，9，10，则总执行次数为n1 + n1 + n2 + n2 = 2n1 + 2n2（在条件语句中，5和7只能选其中一条执行）。

考虑n1,n2趋于无穷，则总执行次数T(n)为4n(n趋于无穷)，可知T(n)是关于n的正比函数，即总执行次数T(n)随着n单位增长（以4个单位），由此我们定义该程序的时间复杂度为O(n)

下面我们列出几种常见的时间复杂度O()：

# 常数阶

x=1

print(x+1)

print(x-1)

# T(n)=2, O=O(1)

# 一次线性阶

x=1

for i in range(n):

    x=x+i

    x=x\*i

print(x)

# T(n)=2n, O=O(n)

# 对数阶

x=1; i=1

while i<=n:

    x=x+i

    i=i\*2

print(x)

# T(n)=2log2(n), O=O(log2(n))

# 二次线性阶

x=1

for i in range(n):

    for j in range(n):

        x=x+i

        x=x\*j

print(x)

# T(n)=2n^2, O=O(n^2)

# 二次线性阶（特殊）

x=1

for i in range(n):

    for j in range(i):

        x=x+i

        x=x\*j

print(x)

# T(n)=1+2+3+……+n = (n^2)/2+n/2 （考虑n为偶数）

# O=O(n^2) (考虑n的最高次项)

已知，各种时间复杂度之间的关系为O(1)<O(log(n))<O(n)<O(n^2)

1. 空间复杂度：

我们定义某个程序在运行过程中为辅助变量（自定义变量）所分配的最大储存空间的函数为空间复杂度O（类比于时间复杂度），举例如下：

# 单变量

m=1

# O=O(1)

# 向量（数组、列表、字典）

M=[]

for i in range(n):

    M.append(i)

# O=O(n)

# 二维向量 （数组、列表、字典）

M=[]

for i in range(n):

    M[i]=[]

    for j in range(n):

        M[i].append(i\*j)

# O=O(n^2)

特殊地，当程序中含有递归结构时，我们规定如下：

#单变量递归

def f(n):

    if n<=1:

        return 1

    return f(n-1)

print(f(n))

# 总空间复杂度O = O(递归层数)= O(n)

#一维向量递归

N=[]

def f(n):

    if len(N)>=n:

        return N

    N.append(1)

    return f(n)

print(f(n))

# 总空间复杂度O = O(递归层数）= O(最大向量长度) = O(n)

#二维向量递归

N=[]

def f(n):

    if len(N)>=n:

        return N

    N.append([1,2,……,m])

    return f(n)

print(f(n))

# 总空间复杂度O = O(各层空间复杂度之和）= O(列维数\*最大行维数) = O(mn)

以时间复杂度为例，不难看出，上述例子中其时间复杂度大小总是固定的，并不存在最好与最坏的情况，然而更多的，当我们处理某一数据结构时，往往需要考虑其最佳情况，最糟糕情况和平均情况，如下：

# 最佳，最糟糕和平均情况

N=[x1,x2,x3,……,xn]

for i in range(len(N)):

    if N[i]==y:              # 考虑N中有且仅有一个等于y的元素

        break

# 最佳情况 O = O(1) 数组第一个元素即为y

# 最糟糕情况 O = O(n) 数组最后一个元素为y

# 平均情况 O = O( 1/n + 2/n + 3/n + …… + 1) = O(n/2 + 1/2) = O(n) 考虑第j个元素为y，以及该事件的概率，从而求O的期望值

当然多数情况下，我们首先考虑最糟糕的情况，只有当该三种同情况下时间复杂度存在数量级差别时，才可能首要考虑平均情况下的时间复杂度。

最后，当给定不同类型数据结构时，若用同一算法处理该数据结构，其时间复杂度也有可能不同，如下：

# 列表和字典性能比较：判断其是否含有某一元素

#列表

List=[x1,x2,x3,……,xn]

for i in range(len(List)):

    if List[i]==y:

        print(1)

# O=O(n)

Dict={"x1":1,"x2":1,……,"xn":1}

print(Dict["y"])

# O=O(1)

\*补充：python中列表，链表和字典常见操作的时间复杂度

（1）列表：

index[]            O(1)

in                            O(n)

append                  O(1)

insert(i,item)          O(n)

pop                        O(1)

pop(i)   / del(i)         O(n)

sort                        O(nlogn)

reverse(逆序)         O(n)

get slice[x:y]           O(k)(k是X到Y的长度)

copy              O(n)

（2）链表：

index[]            O(n)

in                            O(n)

append                  O(n)

insert(i,data)          O(n) \*头部插入，删除时间复杂度为O(1)

insert\_1(node,data) O(1)

pop                        O(n)

pop(i)        O(n)

pop\_1(node) O(1)

copy              O(n)

（3）字典

index[]        O(1)

in                   O(1)

set k-v          O(1)

del k-v     O(1)

copy              O(n)

# 数据可视化

1. Python

本文将主要介绍Python中应用最为广泛的绘图模块matplotlib的基本用法，其余绘图模块如serborn, plotnine会作少量介绍：

1 . matplotlib基本绘图属性

（1）创建画布

# (1) 自定义或导入数据

import matplotlib.pyplot as plt

import numpy as np

X=np.array(range(1,11)) # 数据X和Y1，Y2，Y3，Y4

Y1=X

Y2=X\*\*2

Y3=np.sqrt(X)

Y4=np.sin(X)+np.sin(X)

# （2）创建画布

fig=plt.figure(figsize=[8,8]) # 创造一个主画布，并调整主画布的尺寸[width, height]，可不填

axes1=fig.add\_subplot(2,2,1) # 从主画布中分划出一个子画布（坐标轴）1，位于左上角

axes2=fig.add\_subplot(2,2,2) # 子画布2，位于右上角

axes3=fig.add\_subplot(2,2,3) # 子画布3，位于左下角

axes4=fig.add\_subplot(2,2,4) # 子画布4，位于右下角

plt.subplots\_adjust(wspace=0.2,hspace=0.2) # 调整子画布之间的相对宽，高间距，间距均为0时表示各幅子图贴在一起

axes1.plot(X,Y1)

axes2.plot(X,Y2)

axes3.plot(X,Y3)

axes4.plot(X,Y4)

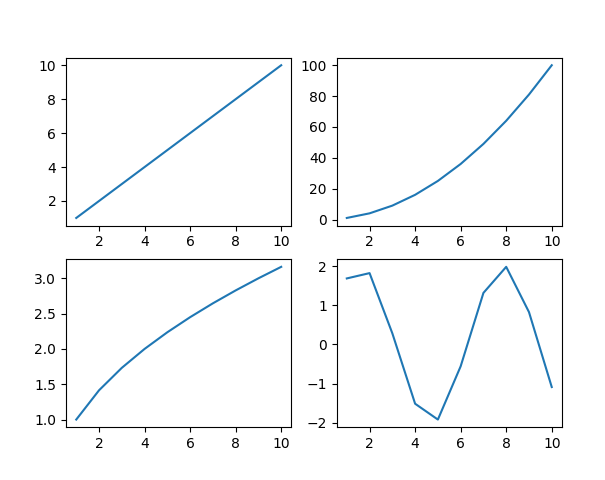
"""

也可以： fig,axes=plt.subplots(nrows=2, ncols=2) 将返回一个包含两行两列子画布的数组axes。

"""

plt.savefig("figure1.png") # 用于下载绘制的图表，不过一般直接在figure窗口交互下载

plt.show()



以上是最最基本的绘图操作，下面我们来逐一介绍绘图属性：

（2）画布操作

# (3) 画布操作：下面的演示以单图为例

import matplotlib.pyplot as plt

import numpy as np

X=np.array(range(1,11)); Y=np.sqrt(X)+np.sin(X) # 自定义数据X，Y

plt.rcParams["font.sans-serif"]=["SimHei"] # 解决中文标题，坐标轴名，刻度名等问题

plt.rcParams["axes.unicode\_minus"]=False

fig=plt.figure() # 创建画布

plt.title("图1") # 为当前图表添加标题

plt.xlabel("X轴") # 为当前图标添加横纵坐标轴名

plt.ylabel("Y轴")

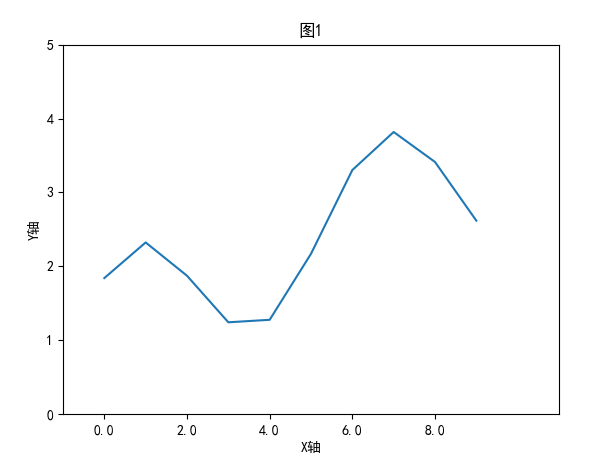
plt.xticks(range(1,11,2),labels=[0.0, 2.0, 4.0, 6.0, 8.0]) # 重新定义X轴的刻度（一般不重新定义Y轴的刻度）

plt.xlim(xmin=0, xmax=12) # 调整X轴和Y轴的范围（使用较少）

plt.ylim(ymin=0, ymax=5)

plt.plot(X,Y)

plt.show()



（3）图表操作

# (4) 图表操作

import matplotlib.pyplot as plt

import numpy as np

X=np.array(range(1,11)); Y1=np.sqrt(X)+np.sin(X); Y2=X\*\*2-X

fig=plt.figure()

plt.plot(X, Y1, linestyle="--", linewidth="2", color="blue",

     marker="o", markerfacecolor="red",  label="x-y1")

plt.plot(X, Y2, linestyle="-.", linewidth="2.5", color="red",

     marker="D", markerfacecolor="blue",  label="x-y2")

plt.fill\_between(X, Y1, Y2, facecolor="green",alpha=0.3)

"""

linestyle表示线条样式, linewidth表示线条粗细, color表示线条颜色

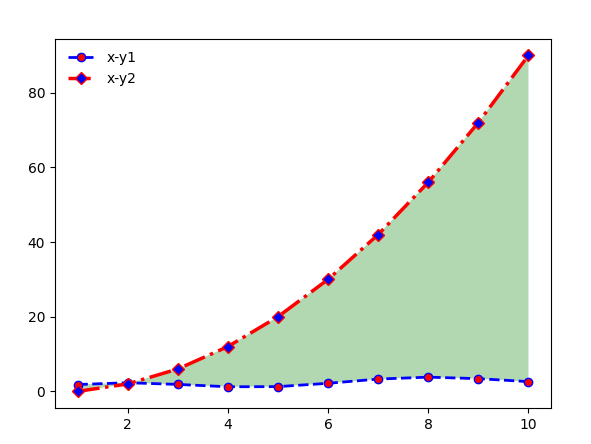
marker表示坐标点样式, markerfacecolor表示坐标点颜色, label用于给该线图起名, 用于绘制图例

fill\_between函数用于曲线间颜色填充, Y1, Y2用于给填充区域划界, 其中facecolor表示填充颜色, alpha指填充透明度

"""

plt.legend(frameon=False, facecolor="blue") # 前者指去除图例边框，后者指图例背景颜色

plt.show()



若想要在同一画布内绘制两个Y坐标轴，可采用如下方法：

import matplotlib.pyplot as plt

import numpy as np

X=np.array(range(1,11)); Y1=np.sqrt(X)+np.sin(X); Y2=X\*\*2-X

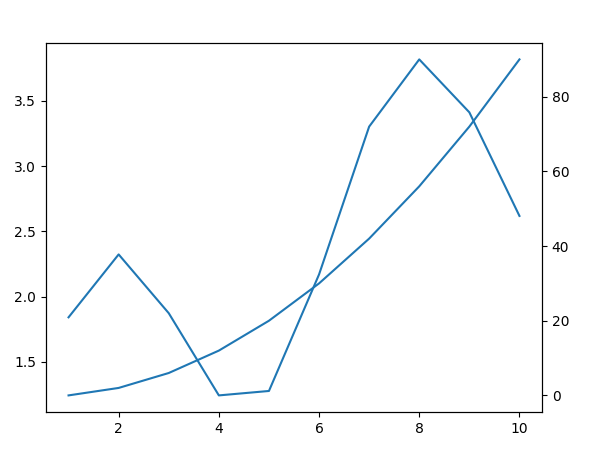
fig,axes1=plt.subplots()

axes2=axes1.twinx()

axes1.plot(X,Y1)

axes2.plot(X,Y2)

plt.show()



2. 基本绘图类型

（1）散点图

# (1) 散点图

import matplotlib.pyplot as plt

import numpy as np

X=np.array(range(1,11)); Y=np.random.rand(10)

fig=plt.figure()

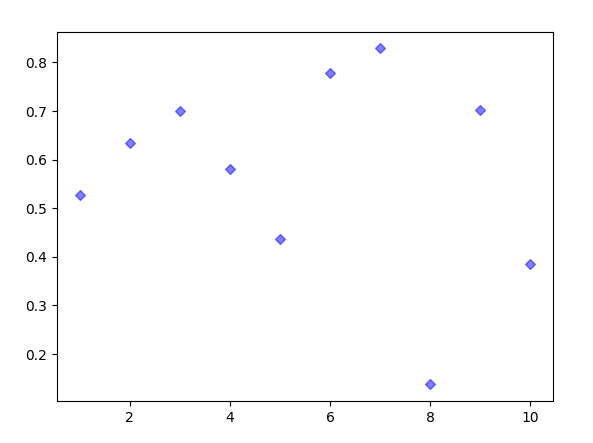
plt.scatter(X, Y, marker="D", s=25, color="blue",alpha=0.5)

"""

marker表示散点样式，s表示散点大小，color表示散点颜色，aplha表示透明度

"""

plt.show()



（2）柱状图

# (2) 柱状图

import matplotlib.pyplot as plt

import numpy as np

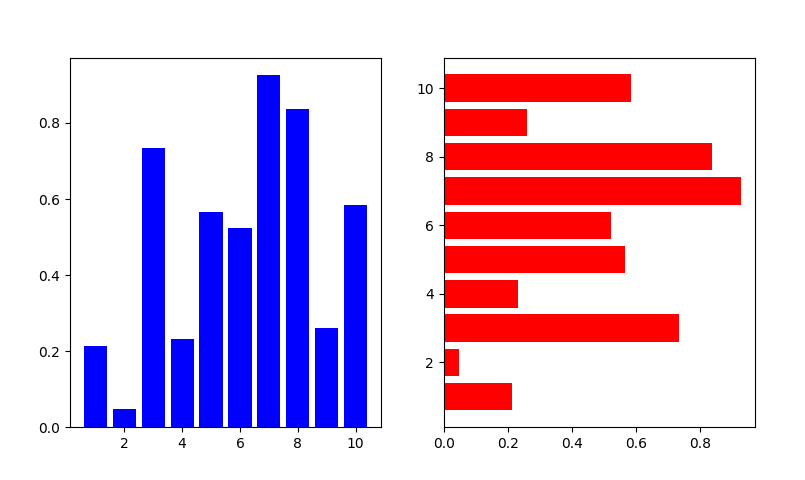
X=np.array(range(1,11)); Y=np.random.rand(10)

fig,axes=plt.subplots(1,2)

axes[0].bar(X, Y, color="blue") # 柱形图

axes[1].barh(X, Y, color="red") # 条形图

plt.show()



（3）直方图

# (3) 直方图

import matplotlib.pyplot as plt

from scipy import stats

Y1=stats.binom.rvs(n=100, p=0.6, size=1000)

Y2=stats.binom.rvs(n=100, p=0.4, size=1000)

fig=plt.figure()

plt.hist(Y1, bins=50, range=[1,60], density=True, color="blue")

plt.hist(Y2, bins=50, range=[41,100], density=True, color="red")

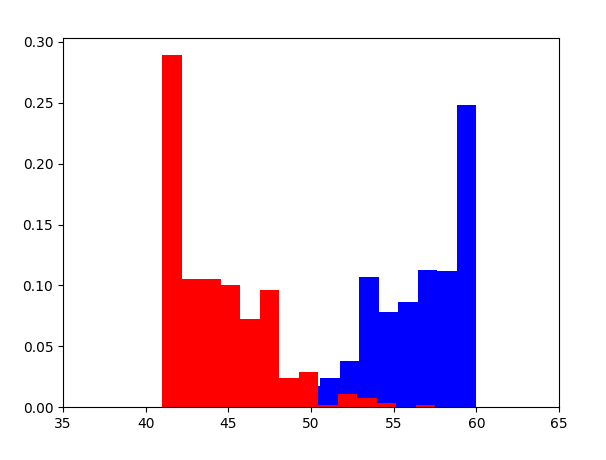
"""

bins表示直方图箱的总数，range=[min,max]表示直方图的范围，density=True表示为频率直方图，False为频数直方图

"""

plt.xlim(xmin=35, xmax=65)

plte.show()



（4）箱线图

# (4) 箱线图

import matplotlib.pyplot as plt

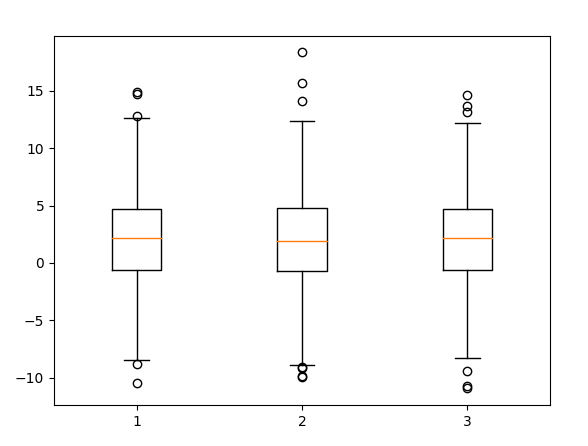
from scipy import stats

Y=stats.norm.rvs(loc=2, scale=4, size=3000).reshape(1000,3)

fig=plt.figure()

plt.boxplot(Y, vert=True) # vert=True表示箱线图垂直，FaLse表示水平

plt.show()



（5）饼图

# (5) 饼图

import matplotlib.pyplot as plt

import numpy as np

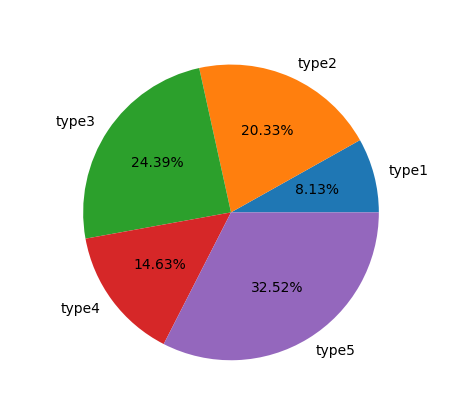
Y=np.array([10,25,30,18,40])

label=np.array(["type1", "type2", "type3", "type4", "type5"])

fig=plt.figure()

plt.pie(Y,autopct="%.2f%%", labels=label) # autopct="%.2f%%"表示各项比率大小按两位小数展示

plt.show()



（6）三维图

# (6) 三维图：包括三维曲线图，曲面图，柱状图和散点图

import matplotlib.pyplot as plt

import numpy as np

X=np.array(range(1,11))

Y=np.random.normal(0,1,size=10)

Z=np.sin(Y)+np.sqrt(X)

XX,YY=np.meshgrid(X,Y) # 由坐标向量构建坐标矩阵

ZZ=np.sin(YY)+np.sqrt(XX)

height=np.zeros\_like(Z) # 3D柱状图中柱的底部位置

width=depth=0.3 # 3D柱状图中柱的长度和宽度

fig=plt.figure()

axes1=fig.add\_subplot(2,2,1,projection="3d")

axes2=fig.add\_subplot(2,2,2,projection="3d")

axes3=fig.add\_subplot(2,2,3,projection="3d")

axes4=fig.add\_subplot(2,2,4,projection="3d")

plt.subplots\_adjust(wspace=0.2,hspace=0.2)

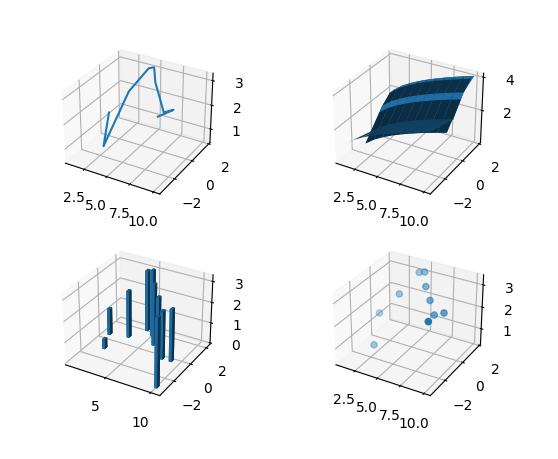
axes1.plot3D(X,Y,Z) # 三维曲线图

axes2.plot\_surface(XX,YY,ZZ) # 三维曲面图

axes3.bar3d(X,Y,height,width,depth,Z) # 三维柱状图

axes4.scatter3D(X,Y,Z) # 三维散点图

plt.show()



3. Python绘图补充

# 3 高级绘图种类：seaborn模块的部分绘图知识

"""

seaborn是对matplotlib的进一步封装，除了兼容matplotlib大部分功能外，还含有更为高级的绘图技巧

"""

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

X=np.random.binomial(10,0.6,size=10000).reshape(10,1000)

sns.set\_style("whitegrid") # 图表风格，默认为darkgrid（同matplotlib），此外有whitegrid, dark, white, ticks

sns.set\_context("paper") # 绘图比例，默认为notebook，还有paper, talk, poster

pal1=sns.color\_palette("hls",8)

pal2=sns.color\_palette("Paired",8)

"""

颜色主题,pal用于保存预设的颜色主题, "hls"(分类色), "Paired"(对比色), n表示颜色种类

"""

# 直方图与核密度图

sns.displot(X[0], kde=True, rug=True, palette=pal1) # displot是一个独立的画布，只能以单图存在

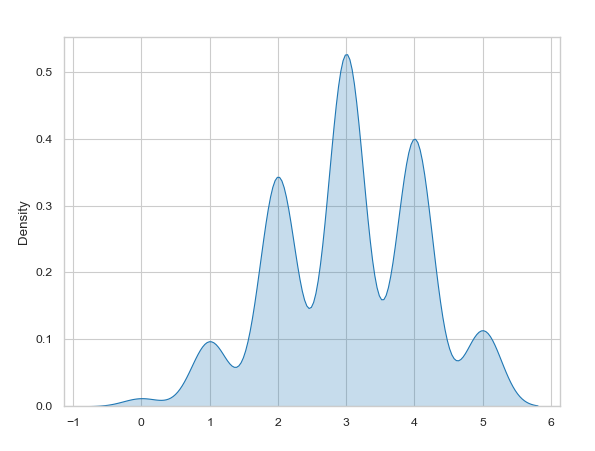
"""

kde(是否添加密度曲线),rug(是否在底部添加观测细条),palette（颜色主题）

"""

sns.kdeplot(X[0], shade=True, palette=pal2)

plt.show()



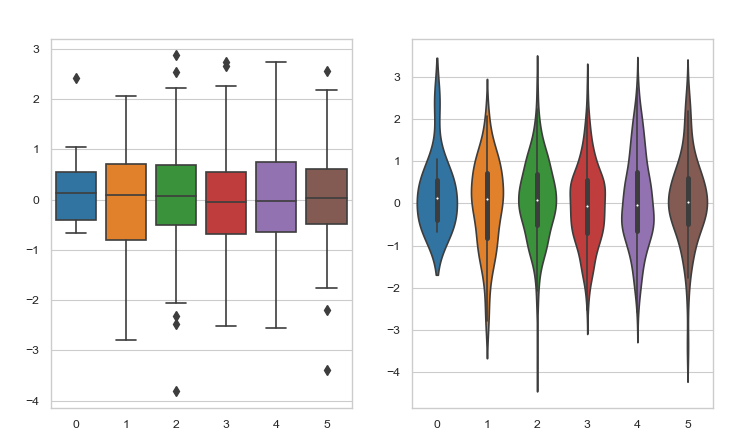
# 箱型图与小提琴图

fig,axes=plt.subplots(1,2)

sns.boxplot(X[0], Y[0], ax=axes[0])

sns.violinplot(X[0], Y[0], ax=axes[1])

plt.show()



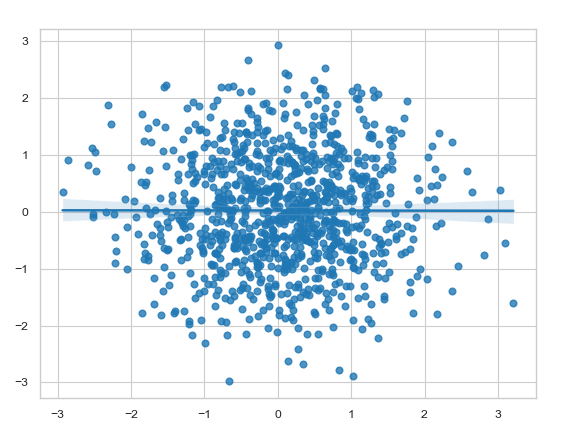
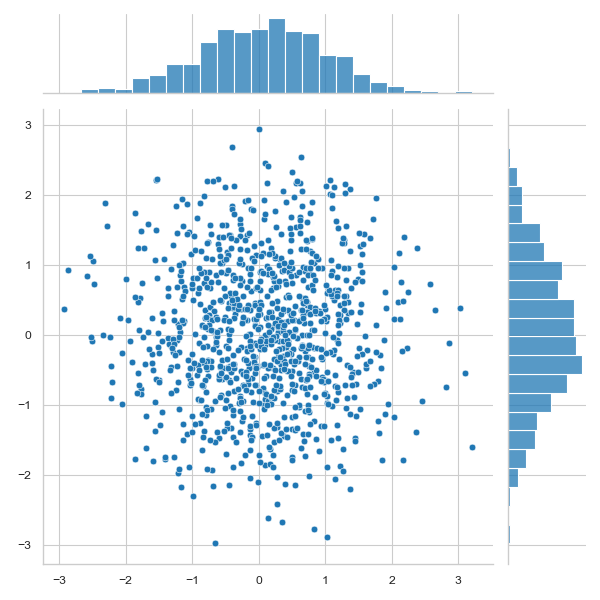
# 拟合散点图与联合分布图

sns.jointplot(Y[0], Y[1]) # jointplot是一个独立的画布，只能以单图存在

fig=plt.figure()

sns.regplot(Y[0], Y[1])

plt.show()



# 散点矩阵图与相关系数热力图

data=pd.DataFrame(

    {"num1":np.random.normal(0,1,size=1000),

    "num2":np.random.normal(1,2,size=1000),

    "num3":np.random.exponential(0.2,size=1000),

    "num4":np.random.poisson(5,size=1000)

    },

    columns=["num1","num2","num3","num4"])

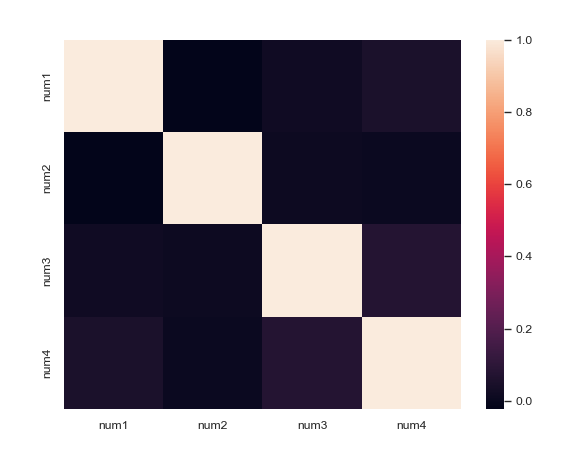
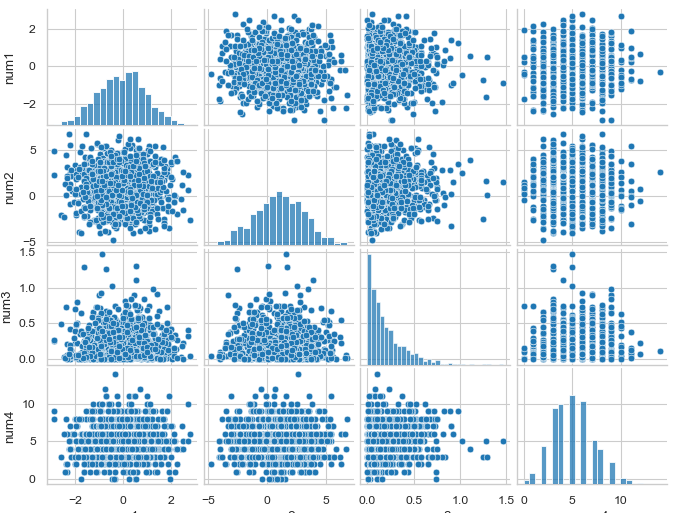
sns.pairplot(data)  # 只用于对数据集绘图

fig=plt.figure()

cor=data.corr() # 相关系数矩阵

sns.heatmap(cor)

plt.show()



# Python数据爬虫

要想真正熟练并精通爬虫技术并不容易，本文只简单介绍以python为语言基础的常用的爬虫手段，如下：

我们先介绍一种最基础的基于静态网页html源码爬取的技术，来帮助我们了解爬虫的底层原理

# 爬虫基础-爬取文字

import requests    # 导入爬虫模块

from bs4 import BeautifulSoup

"""

假定我们的目标是爬取小白盘网站首页的推荐书籍书名，并将其导入至EXCEL中

目标网址:https://www.xiaobaipan.com/

查看robot协议: 点击https://www.xiaobaipan.com/robots.txt可访问网址源文件根目录中robots声明，一般不允许商业爬虫，个人

用于学习研究影响不大。

打开网址html源码:点击f12 -> elements或元素，利用左上角导航箭头，可用于定位

"""

res = requests.get("https://www.xiaobaipan.com")  # 以国家统计局2021国内生产总值为例，向服务器发出请求并捕捉响应

# print(res.status\_code)  返回结果为200表示请求成功

# print(res.text)  返回html标记语言

html=BeautifulSoup(res.text,"html.parser") # 解析html标记语言

# print(html) 内容与上一条输出信息相同，只是变成了解析后的html

"""

对于最常见的爬虫技术来说，其基本原理十分简单，就是从解析后的html中寻找，定位，抓取我们想要的内容，以文本内容为例，

鉴于html语言的分层特点，可以利用find()【寻找满足条件的第一个】和find\_all【寻找满足条件的全部】层层定位，或者快速

定位文本内容所在位置，其中两个函数的参数均为"("标签","对应标签的属性")

"""

tag1=html.find("div",class\_="main-container")

tag2=tag1.find\_all("h6")

file= open("book\_name.txt","w",encoding="utf-8")  # 将爬取的内容写入txt文件

for tag in tag2:

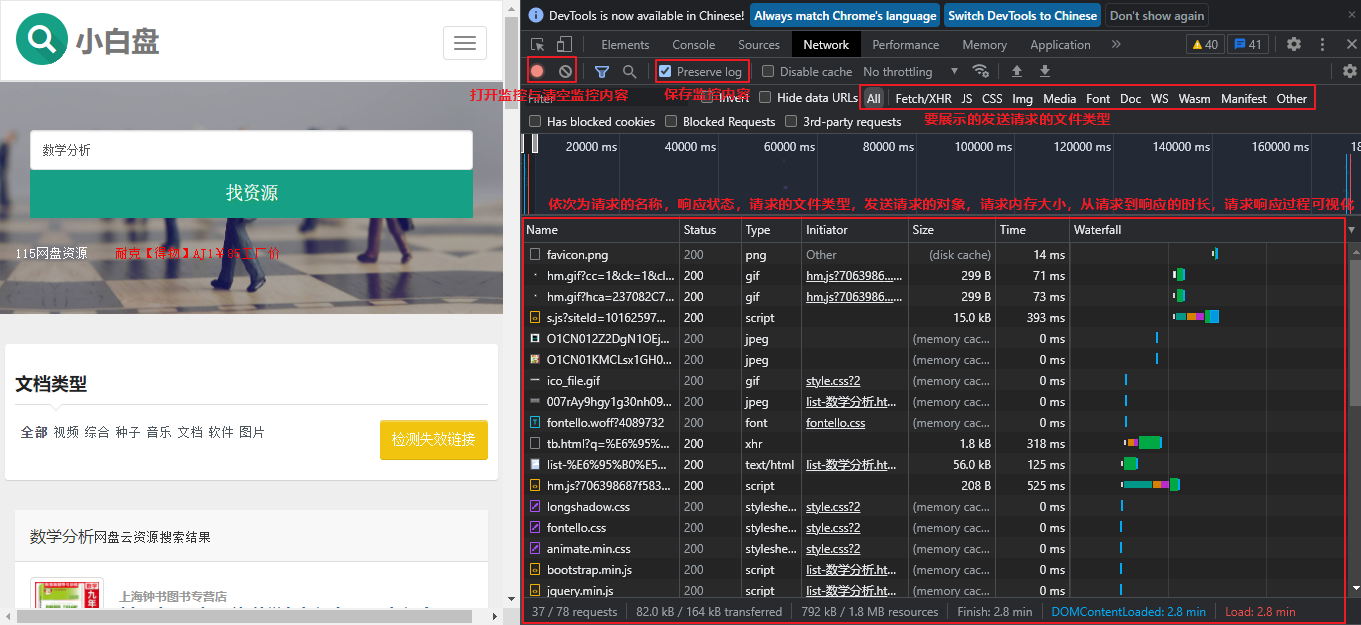
    tag\_root=tag.find("a")  # tag\_root指不可再展开的html标签，即根标签

    # file.write(tag\_root["title"][:-5]+"\n") # tag\_root["title"]用来提取标签的"title"属性值

    file.write(tag\_root.text.strip()+"\n") # tag\_root.text用来提取标签内的text文本内容，效果与上述方法一致

file.close()

借用以上的爬虫方法，一般来说我们可以爬取诸如小说网上的小说，博客网上的博客及评论等等，然而，在实际应用中，我们需要爬取的内容并不来自html源码文件，而是该网站在服务器的文件，如图片，视频，pdf等等，因此显而易见的是，我们并不能从html源码文件中爬取任何关于非文本内容的资源，为什么呢?





同样以小白盘为例，在我们搜索“数学分析”之后，会出现一系列关乎“数学分析”的资源，需注意的是，其中包含了大量的封面图片及头像图片，然而在定位html源码时，我们只能找到打印图片的html指令，其图片由于储存在服务器文件夹中而以链接的形式被调用。因此我们的目的是从服务器文件夹中爬取图片，视频等资源！

上面显示了该页面各种文件（html源码文件一般在DOC中）的请求及响应动态，现在我们来尝试爬取该页面的全部图片

# 爬取图片

import requests    # 导入爬虫模块

import os  # 导入OS文件处理模块

from bs4 import BeautifulSoup

path="C:/Users/熊高庆/Desktop/项目文件/Python/Python程序/Learning/Functions/images"  # 判断该路径下文件夹images是否存在

if not os.path.exists(path):  # 不存在则创建一个新文件images

    os.mkdir(path)

res=requests.get("https://www.xiaobaipan.com/list-SPSS.html?from=1")

html=BeautifulSoup(res.text,"html.parser")

tags=html.find\_all("img")  # 直接暴力搜索标签"img"

Url=[]

for tag in tags:

    Url.append(tag["src"])  # 以上操作类比文字爬虫，仅仅从爬取文字转而为url链接,然后将图片URL链接保存在列表Url中

for i in range(len(Url)):

    try:   # try……except语句用来避免程序因某个子操作异常而终止

        res=requests.get(Url[i])  # 针对图片网络链接发出新的请求

        new\_url=path+"/"+Url[i].split("/")[-1]  # 将网络链接改为本地链接

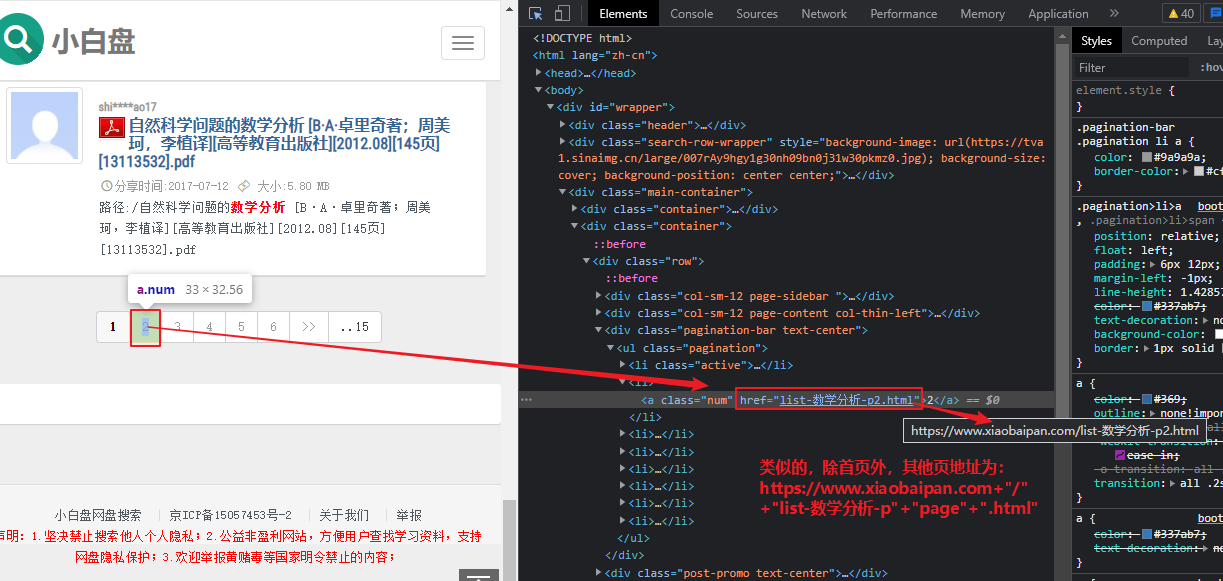
        file=open(new\_url,"wb")  # 创建本地链接，用二进制形式保存图片信息

        file.write(res.content)

        print("爬取成功")

    except Exception:

        print("爬取失败")

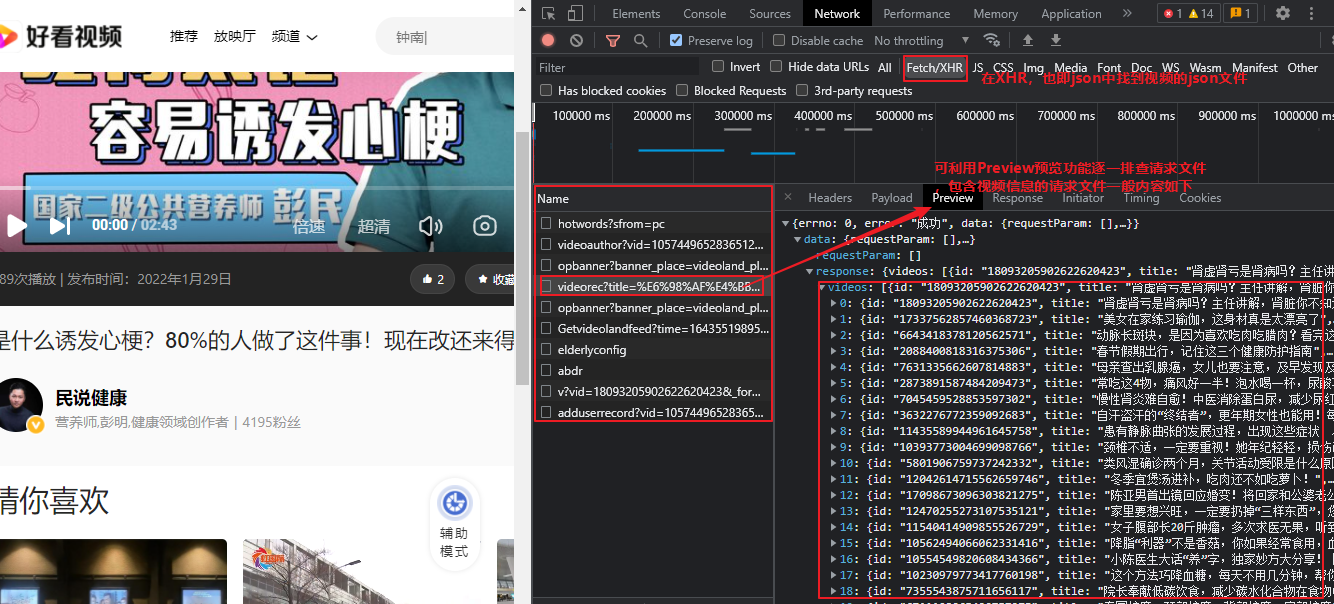


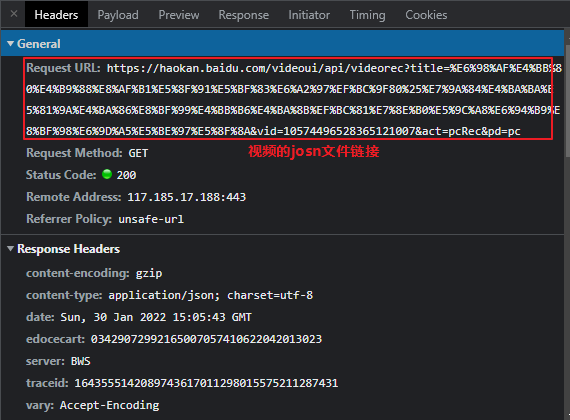
如果我们要爬取其他页的图片，可以根据各页url链接的特点自行构造其他页的链接，如下：

各页链接:

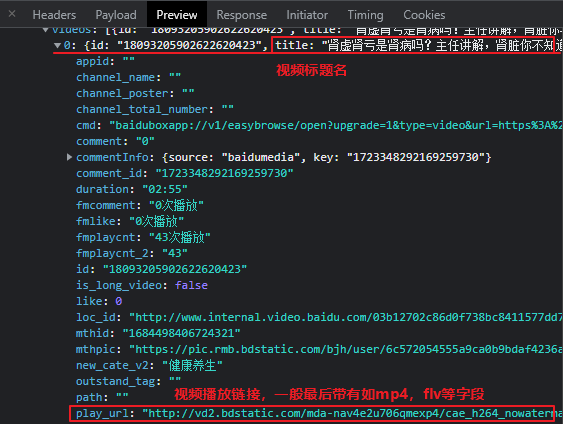
url\_j="https://www.xiaobaipan.com/"+"list-数学分析-p"+"j"+".html"

同样的，我们还可以爬取视频网站的视频，但通常情况下我们会用到json解析模块这一工具，我们以好看视频网站为例，具体如下：





点击XHR，右侧可弹出一个关于该请求的信息框，我们重点关注Headers和Preview两个。利用Preview找到我们需要的那个请求，然后在Headers的Request URL中找到视频的json文件链接，链接可用ctrl+C复制。



在解析完json文件后，我们会得到一个列表和字典交互的数据结构，我们找到最底层的那个列表，里面包含了该页面所有视频的信息，从中找到每一个视频的网络链接，然后依照图片那样即可完成爬虫，完整代码如下：

# 爬取视频

"""

随着网页制作中Ajax技术的发展，制作者往往会利用json文件来封装部分html源码以加快网页的刷新，因此我们直接利用requests获取

的html源码并不完整，自然也就无法找到对应资源（如文字，图片，视频）的内容或链接。一般来说，有两种常用的方法来解决这一问题:

（1）利用selenium库模拟Chrome浏览器来访问网页（功能十分强大）。（2）将json文件中解析为字典，并从中找到资源内容或链接。本文主要介绍较简单的第二种

"""

import requests # 导入爬虫模块

import os  # 导入OS文件处理模块

import json # 导入json模块

from bs4 import BeautifulSoup

path="C:/Users/熊高庆/Desktop/项目文件/Python/Python程序/Learning/Functions/videos"  # 判断该路径下文件夹videos是否存在

if not os.path.exists(path):  # 不存在则创建一个新文件videos

    os.mkdir(path)

Request\_URL="https://haokan.baidu.com/videoui/api/videorec?title=%E6%98%AF%E4%BB%80%E4%B9%88%E8%AF%B1%E5%8F%91%E5%BF%83%E6%A2%97%EF%BC%9F80%25%E7%9A%84%E4%BA%BA%E5%81%9A%E4%BA%86%E8%BF%99%E4%BB%B6%E4%BA%8B%EF%BC%81%E7%8E%B0%E5%9C%A8%E6%94%B9%E8%BF%98%E6%9D%A5%E5%BE%97%E5%8F%8A&vid=10574496528365121007&act=pcRec&pd=pc"

res=requests.get(Request\_URL) # 找到XHR中视频的json链接（可以逐个preview预览排查）

# print(res.text) 打印出来，我们会发现这是一个json文件

json=json.loads(res.text) # 将上述json文件解析为字典

video\_list=json["data"]["response"]["videos"] # 根据preview预览的字典结构进行拆分，找到最底层视频所在的列表

for i in range(5):

    # print(video\_list[i],"\n") 打印列表内各个元素，我们会看到每一个视频的信息，它们是以字典的形式存在

    video\_url=video\_list[i]["play\_url"]  # 找到视频的网络链接

    video\_name=video\_list[i]["title"]+".mp4"  # 将视频重命名

    try:

        res=requests.get(video\_url)

        new\_url=path+"/"+video\_name # 将网络链接改为本地链接

        file=open(new\_url,"wb")

        file.write(res.content) # 用二进制形式保存视频信息

        print("爬取成功")

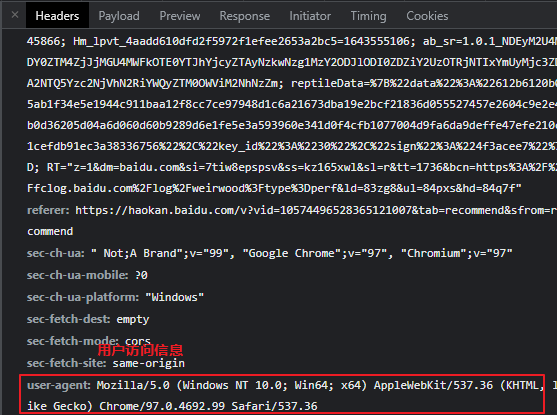
    except Exception:

        print("爬取失败")

通过上述方法，我们可以较轻易地爬取小网站视频以及短视频，但对于较大的视频网站以及爬取电影等长视频时，处理方法会更为复杂，此处不作过多介绍

最后，我们来介绍一些常见的反爬虫操作：

（1）添加headers请求头



在上例中，我们在Headers底部可找到用户访问信息user-agent ，将其封装到headers中，并打包给requests.get函数，可有效将爬虫伪装为正常的访问，代码如下：

header={

    "user-agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/97.0.4692.99 Safari/537.36"

} # 类似的，还可以添加cookie等信息，不过一般user-agent够用

Request\_URL="https://haokan.baidu.com/videoui/api/videorec?title=%E6%98%AF%E4%BB%80%E4%B9%88%E8%AF%B1%E5%8F%91%E5%BF%83%E6%A2%97%EF%BC%9F80%25%E7%9A%84%E4%BA%BA%E5%81%9A%E4%BA%86%E8%BF%99%E4%BB%B6%E4%BA%8B%EF%BC%81%E7%8E%B0%E5%9C%A8%E6%94%B9%E8%BF%98%E6%9D%A5%E5%BE%97%E5%8F%8A&vid=10574496528365121007&act=pcRec&pd=pc"

res=requests.get(Request\_URL,headers=header)

（2）降低爬虫频率

有时候，网页的服务器会因为我们爬虫行为过于频繁而封杀计算机的Ip访问地址，为了避免这种情况发生，有时候我们不得不适当降低爬虫频率，最简单的方法即是time.sleep函数，如下：

import time # 导入time系统时间模块

for i in range(5):

    video\_url=video\_list[i]["play\_url"]

    video\_name=video\_list[i]["title"]+".mp4"

    try:

        res=requests.get(video\_url)

        new\_url=path+"/"+video\_name # 将网络链接改为本地链接

        file=open(new\_url,"wb")

        file.write(res.content) # 用二进制形式保存视频信息

        print("爬取成功")

    except Exception:

        print("爬取失败")

    time.sleep(2) # 每爬取完一次，程序暂停两秒再执行

# Latex文本编辑 +Markdown文本

工具：Overleaf

1 基本架构

%以下为导言区

\documentclass{article} %文章样式

\usepackage[utf8]{inputenc} %引用宏包（中文）

\usepackage{ctex} %引用宏包（ctex编译器）

\usepackage{graphicx} %引用宏包（图形）

\usepackage{float} %引用宏包（表格）

\usepackage{cite} %引用宏包（参考文献引用）

\title{勾股定理} %标题

\author{熊高庆} %作者

\date{2021年7月21日} %写作日期

\newtheorem{thm}{定理} %定理

% 以下为正文区

\begin{document} %正文

\maketitle %标题，作者，日期生成

\begin{abstract} %摘要

\centering %居中

这是一篇关于勾股定理的小短文

\end{abstract}

\tableofcontents %目录生成

\section{勾股定理在古代} %章节一

\part{...} %第一部分

\chapter{...} %第一章

\section{...} %第一节

\subsection{...} %第一小节

西方称勾股定理为毕达哥拉斯定理，将勾股定理的发现归功于公元前6世纪的毕达哥拉斯学派\cite{Kline}。%参考文献引用

该学派得到了一个法则，可以求出可排成直角三角形三条边的三元数组。毕达哥拉斯学派没有

书面著作，该定理的严格表述和证明则见于欧几里得\footnote{欧几里得，约公元前330-275年。} %脚注

《几何原本》的命题47：“直角三角形斜边上的正方形等于两直角边上的两个正方形之和”。证明是用

面积做的。

我国《周髀算经》载商高（约公元前12世纪）答周公问：

\begin{quote} %引用

\zihao{-5}\kaishu %引用内字体变更

勾广三，股修四，径隅五。

\end{quote}

又载陈子(约公元前7--6 世纪)答荣方问：

\begin{quote}

\zihao{-5}\kaishu

若求邪主自者， 以日下为勾， 日高为股，勾股各自乘， 并而开方除之，得邪至日。

\end{quote}

都较古希腊更早，后者已经明确道出勾股定理的一般形式。图\ref{fig:xiaotu} %标签引用

是我国古代对勾股定理的一种证明\cite{Quanjing}。%参考文献引用

\begin{figure}[ht]

\centering

\includegraphics[width=5cm]{xiaotu.pdf} %引用图片

\caption{宋赵爽在《周髀算经》注中作的弦图（仿制），该图给出了勾股定理一个极具对称美的证明。} %图注

\label{fig:xiaotu}

\end{figure}

\section{勾股定理在近代} %章节二

勾股定理可以用现代语言表述如下：

\begin{thm}{勾股定理}

直角三角形斜边的平方等于两腰的平方和。

%空行用于分段

可以用符号语言表述为：设直角三角形ABC，其中$\angle ACB=90^\circ$，则有 %行内公式

\end{thm}

\begin{equation} %行间公式

\centering

AB^2=BC^2+AC^2

\label{eq:勾股定理}

\end{equation}

满足式(\ref{eq:勾股定理})的整数称为勾股数，第一节所说毕达哥拉斯学派得到的三元数组就是勾股数。

下表列出一些较小的勾股数:

\begin{table}[H] %表格环境

\begin{tabular}{|rrr|} %表格生成

\hline %表内横线

直角边$a$ & 直角边$b$ & 斜边$c$\\

\hline

3 & 4 & 5\\

\hline

5 & 12 & 13\\

\hline

\end{tabular}

\qquad

($a^2+b^2=c^2$)

\end{table}

% 参考文献引用放正文末尾

\bibliographystyle{plain} %参考文献风格

\bibliography{勾股定理.bib} %参考文献引用

\end{document}

\documentclass[a4paper]{article}

\usepackage[utf8]{inputenc}

\usepackage{ctex}

\newcommand{\song}{\CJKfamily{song}} % 正文五号字体, 单倍行距，中文默认宋体

\newcommand{\wuhao}{\fontsize{10.5pt}{15.75pt}\selectfont}

\usepackage[margin=1in]{geometry} % 设置页边距为A4样式

\usepackage{setspace}

\renewcommand{\baselinestretch}{2.0}

\usepackage{ragged2e}

\usepackage{indentfirst}

\setlength{\parindent}{2em}

\usepackage{enumerate}

\usepackage{amsmath}

\newtheorem{theorem}{Theorem}

\newtheorem{lemma}{Lemma}

\newtheorem{proof}{Proof}

\usepackage{graphicx}

\usepackage{fontspec}

\setmainfont{Times New Roman} %全文字体设置（英文）

\setsansfont{Verdana} % 常用于显眼的大标题

\setmonofont{Courier New} % 常用于排版代码程序

\usepackage{ulem}

\usepackage{lettrine}

\usepackage{float}

\usepackage{cite}

\title{笔记}

\author{XXX}

\begin{document}

\maketitle

\tableofcontents % 目录生成

\section{标点符号和特殊符号} %------------------------------------------------------%

中文标点符号：。 . ， 、 ； ： ？ ！ · ‘’ ”“ 《》 <> （） 【】 {} …… ——

英文标准标点符号：, . ; : ! ? ' ' () [] - / \* @

"\,'A'" % ”’A’”

A-B % A-B

A--B % A--B

A---B % A---B

$\sim$ % ~

$\dots $ % ……

$\backslash $ % \

$\prime$

\# % #

\$ % $

\% %

\& % &

\{ \} % {}

\\_ % \_

A\ B % 空格

A

B % 换行分段

A\\B % 换行不分段（不缩进）

A \quad B % 行内间隔

希腊文字符： % 大写希腊字母请首字母大写

$\alpha$

$\beta$

$\gamma$

$\delta$

$\epsilon$

$\zeta$

$\eta$

$\theta$ \quad $\Theta$

$\lambda$

$\mu$

$\xi$

$\pi$

$\rho$

$\sigma$

$\tau$

$\upsilon$

$\phi$

$\varphi$

$\chi$

$\psi$

$\omega$

罗马数字

$\uppercase\expandafter{\romannumeral1}$ % 大写

$\romannumeral1$ % 小写

\section{字体和强调} %------------------------------------------------------------%

英文正文默认罗马字体，直立，中等。

字体：

\textrm{A} % 罗马字体 或写{\textrm xxx}

\texttt{A} % 打印字体

字形：

\textup{A} % 直立体

\textit{A} % 意大利体

\textsl{A} % 倾斜体

字宽：

\textmd{A} % 中等体

\textbf{A} % 加粗体

组合：

\texttt{\textit{\textbf{A}}} % 组合

中文如下（中文默认宋体）

\songti{国} % 宋体

\heiti{国} % 黑体

\fangsong{国} % 仿宋

\kaishu{国} % 楷书

现代用法：引用fontspec（英文）或xeCJK（中文）宏包，在导言区预设定字体，如下：

$\backslash$setmainfont\{Tmines New Roman\}

$\backslash$setCJKmainfont\{kaishu\}

文字强调： % 需引用ulem宏包

\uline{ABCD}

\uuline{ABCD}

\uwave{ABCD}

\dotuline{ABCD}

\section{排版} %-----------------------------------------------------------------%

纸张大小：A4（见导言区设置）

（1）字号：

{\tiny A} % 小六号（注意不采用\tiny{A}这种写法，会改变后面所有字体）

{\scriptsize A} % 六号

{\small A} % 五号

{\normalsize A} % 小四号

{\large A} % 小三号

{\Large A} % 小二号

{\LARGE A} % 二号

{\huge A} % 小一号

（2）对齐：

\begin{flushleft} % 左对齐

A

\end{flushleft}

\begin{center} % 居中

A

\end{center}

\begin{flushright} % 右对齐

A

\end{flushright}

环境里面常使用居中命令，如下：\\

$\backslash$begin\{xxx\}\\

$\backslash$centering\\

xxxxx\\

$\backslash$end\{xxx\}

\justifying{123} % 需引用宏包ragged2e

（3）下沉效果： % 引用宏包lettrine

\lettrine{西}方称勾股定理为毕达哥拉斯定理，将勾股定理的发现归功于公元前6世纪的毕达哥拉斯学派

（4）行间距

全局变化： %\renewcommand{\baselinestretch}{1.1} 在导言区设置

局部修改：

123\\

123

123 \vspace{2cm} \\

123

（5）手动设置首行缩进 %需引用宏包indentfirst

\indent{123} % 缩进

\noindent{123} % 不缩进

\hspace{2em}{123} % 不得已的缩进办法

（6）页码设置

本页无页码，后续页码从下一页开始！！！

\thispagestyle{empty} % 当前页无页码

\newpage

\setcounter{page}{5}

\section{公式} %-----------------------------------------------------------------%

（1）行内公式：

$a+b=c$ % 行内公式（内嵌公式）

（2）行间公式：

\begin{equation}

a+b=c \quad \text{当}a,b>=0 % 公式中插入文字

\end{equation}

或者:

\begin{subequations} %需引用宏包amsmath

\begin{equation}

a+b=c

\end{equation}

\begin{equation}

a-b=d

\end{equation}

\end{subequations}

（3）常用公式符号：

常用数学符号：

$\le$ % 小于等于

$\ge$ % 大于等于

$\neq$ % 不等于

$\in$ % 属于

$\notin$ % 不属于

$\rightarrow$ % 右单箭头

$\Rightarrow$ % 右双箭头

$\leftrightarrow$ % 双向单箭头

$\Leftrightarrow$ % 双向双箭头

$\mathrm{d} y$ % 导数

$\partial y$ % 偏导数

$\ln x$\quad$\log x$\quad$\log\_{a}{b}$ % 对数

$\int\_{a}^{b} x^2 \mathrm{d} x$ % 积分

$\sum\limits\_{i=1}^{n}{a\_{i}}$ % 求和

$\lim\limits\_{i\to +\infty}{a\_{i}}$ % 极限

$\prod\limits\_{i\to n}{a\_{i}}$ % 连乘

上标与下标：

$A\_{ij}=2^{i+j}$ %上标与下标，\_表示下标，^表示上标，之后引用的内容用{}括起来

$\sum\limits\_{i=1}^{n}{i^2+3i+4}$

$\lim\limits\_{i\to +\infty}{i^2+3i+4}$

$\prod\limits\_{ii\to +\infty}{i^2+3i+4}$

$\overline{X}$

$\hat{X}$

$\widetilde{X}$

$\dot{X}$

分式：

$\frac{a+b}{\frac{a}{b}+c}$

根式：

$\sqrt{x^2+2x+1}$

矩阵：

$A=\begin{matrix}

a\_{11} & a\_{12} & a\_{13}\\

a\_{21} & a\_{22} & a\_{23}\\

a\_{31} & a\_{32} & a\_{33}

\end{matrix}$

定界符：

$\lim\limits\_{x \to 0}\left(\frac{a^x+b^x+c^x}{3}\right)^{\frac1x}$ %left和right用于定界

多行公式排版1：

\begin{align}

&\lim\limits\_{x\to 1}\left(\frac{1}{1-x}-\frac{3}{1-x^3}\right) \notag\\

= &\lim\limits\_{x\to 1}\left(\frac{x^2+x-2}{1-x^3}\right) \notag\\

=&\lim\limits\_{x\to 1}\frac{(x+2)(x-1)}{(1-x)(x^2+x+1)} \notag\\

=&\lim\limits\_{x\to 1}\frac{-(x+2)}{x^2+x+1} \notag\\

=& -1

\end{align}

多行公式排版2：

\begin{equation}

\begin{cases}

a+b=c\\

a-b=d\\

\end{cases}

\end{equation}

\section{其他元素} %---------------------------------------------------------%

（1）列表：

\begin{itemize}

\item This is the first row.

\item This is the first row.

\item This is the first row.

\end{itemize}

或者：

\begin{enumerate}[A.] %需加载宏包enumerate，自定义过程需要将内容加在后面的[]中。

\item hello

\item hello

\item hello

\end{enumerate}

（2）表格：

\newpage % 避免表格图片等插在文字中间，干脆直接跳页

\begin{table}[h!]

\centering

\begin{tabular}{|c|c|c|} % 竖线“|”表示表格对应列内加竖线

\hline %头行线

x & y & z \\

1 & 123 & 23 \\

\cline{2-3} % 表内行线，也可用\hline

34 & 57 & 789 \\

\hline %尾行线

\end{tabular}

\caption{my first table}

\end{table}

（3）图片：

\begin{figure}[htb] %需引用宏包graphicx，若引用宏包float，则要是写[H]，图片位置会固定

\centering

\includegraphics[scale=0.3]{chapter2-1.jpeg}

\caption{my first figure}

\end{figure}

%width=\* 宽度

%height=\* 高度 高度和宽度必须标明明确的单位，比如厘米（cm）或者英寸（in）

%scale=\* 倍数

%angle=\* 顺时针旋转角度

（4）摘要：

% 英文摘要

\begin{abstract}

This is a demo\\

\textbf{keywords：} Latex

\end{abstract}

还可以通过在导言区加用下列命令来修改摘要：

% \renewcommand{\abstractname}{\large Abstract}

% 中文摘要

\flushleft{\textbf{摘要：} 这是一个demo}

\flushleft{\textbf{关键词：} Latex}

（5）定理与证明：% 需引用宏包amsmath

\begin{theorem}

This is a theorem.

\end{theorem}

\begin{lemma}

This is a lemma.

\end{lemma}

\begin{proof}

This is proof.

\end{proof}

（6）参考文献引用

step1：编辑xxx.bib文件，如myref.bib

step2：在需要引用参考文献的文字后加上这段代码\cite{valentin2016learning} % 所加内容即为article标题（不可为中文）

step3：引用宏包cite，并在文件末尾加上：

\bibliographystyle{unsrt} % 按引用次序编排参考文献

\bibliography{myref}

\end{document}

markdown文本标记

不同于latex，markdown仅仅作为一种简单的文本标记语言来使用（类似html），在Github中为了便于在自己的项目库中展示笔记，说明等，我们需要使用markdown。本文只简单介绍Markdown的最基本功能：标题大纲，加粗斜体，分割线，符号标记，插入图片，插入表格，插入超链接，插入代码块。我们举例如下：

markdown常用命令如下：

\*\*\* \*\*\*用作分割线

【标题大纲】

# 一级标题 #+标题名，下同

## 二级标题

### 三级标题

\*\*\*

【加粗斜体】

\*倾斜\* \*文字\*，下同

\*\*加粗\*\*

\*\*\*倾斜加粗\*\*\*

\*\*\*

【符号标记】

符号前面加转义符"\"即可，如下：

\\*

\-

\\_

\#

\()

\*\*\*

【插入图片】

插入项目库内图片

![](/timg5.png) ! [图片描述，可不加](相对路径+ /+ 图片名)

插入网络图片

![](<http://img0.iplant.cn/image61/b/1633320.jpg>) ![](图片网络路径)

用html语言插入图片，可调整大小

<img src="https://github.com/Eksi123/Codebase/blob/main/timg5.png" width="50%">

\*\*\*

【插入表格】

插入列表

- 第一行

- 第二行

- 第三行

插入表格

学号|姓名|成绩

-|-|-

01|小明|90

02|小红|88

03|小王|85

\*\*\*

【插入超链接】

链接到网站

<https://github.com/Eksi123/Codebase> <网址>

链接到项目库内文件

[file](/README.md) [文件描述，一定要加](相对路径+ / +文件名)

\*\*\*

【插入代码块】

``` 英文半角符号```xxx```，xxx为代码块

for i in range(5)

print(i)

```

# Git+Github项目库

**Github（图形界面项目管理）**

\* 鉴于Github是国外网站，我们需要借助第三方工具“dev-sidecar”来托管访问，访问时切勿关闭该软件。此外该软件还可托管访问诸如Stack Overflow，dns等

\* 现阶段我们暂时不需要使用到Git的历史代码记录功能，而仅仅是用Github来托管保存自己的项目代码，以及搜寻他人优秀项目这一功能，故不介绍Git

1 Github界面介绍



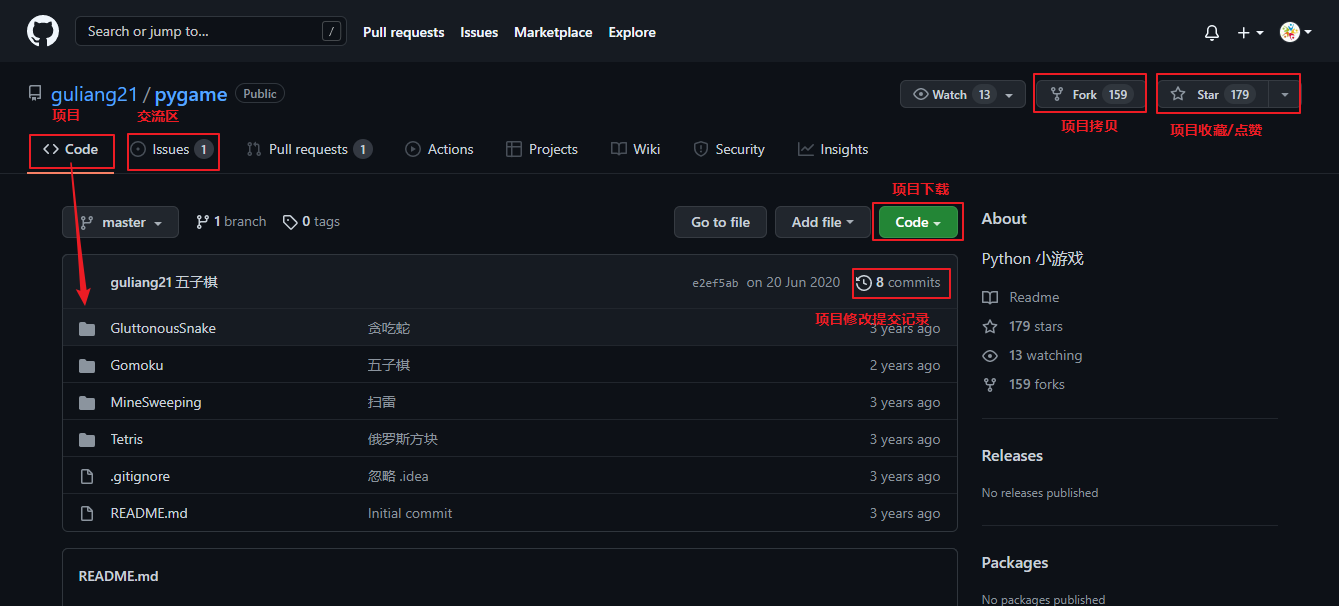
其中我们用得较多的也就是搜索框和个人项目库，分别对应了我们的两个目的：寻找，收藏并下载他人项目，以及管理自己的项目，下面我们来逐一介绍：

2 寻找，收藏并下载他人项目



如上所示，输入Python小游戏，我们可左半框的结果，其中包括repositories项目库，code代码，以及commits提交结果，issues提问，discussion讨论等等，

以项目库为例，我们点击第一个

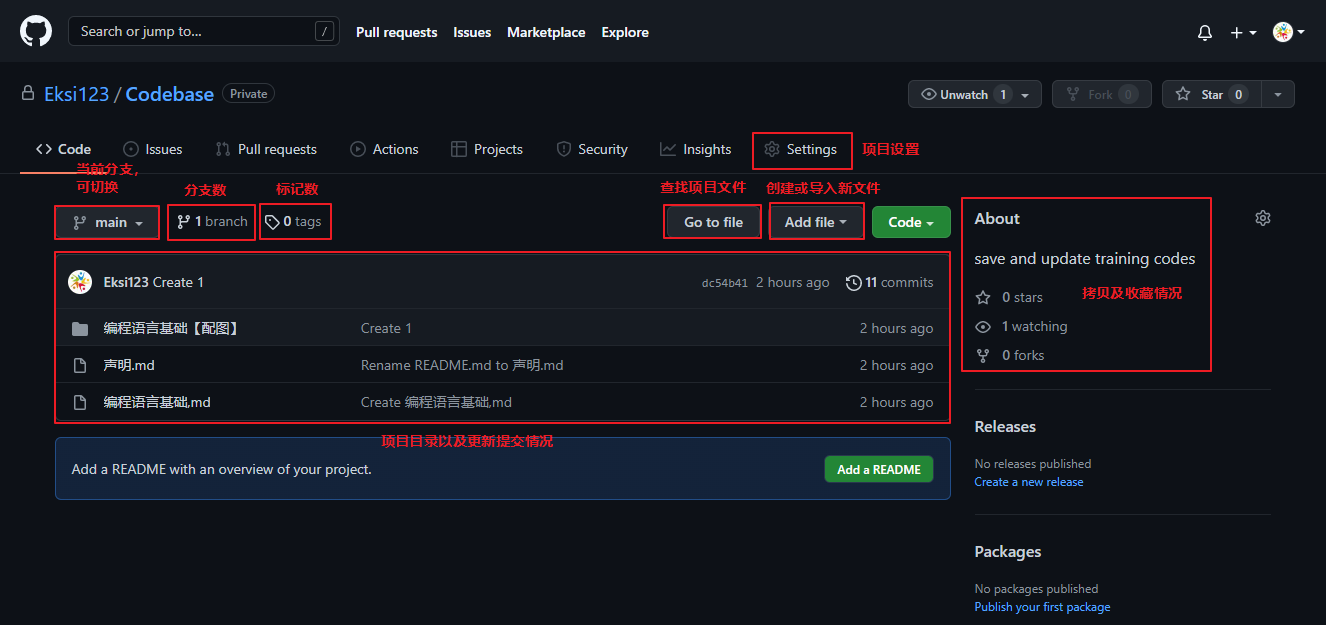


其中Code为项目区，可点击任一项目进行查看，Issue为问题交流区，fork用于拷贝该项目库为己用，star用于收藏该项目，commits用于查看项目修改记录

Code用于下载该项目至本地，如下：

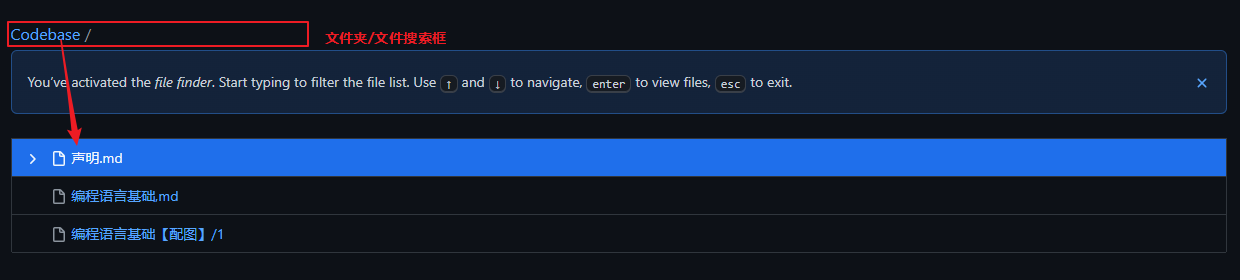


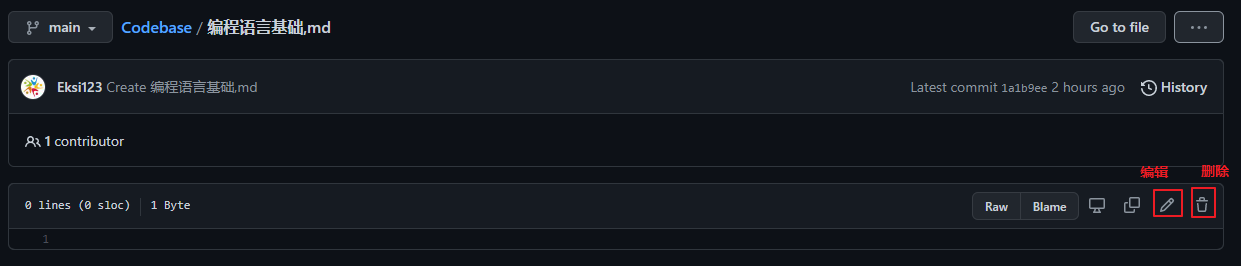
3 管理个人项目

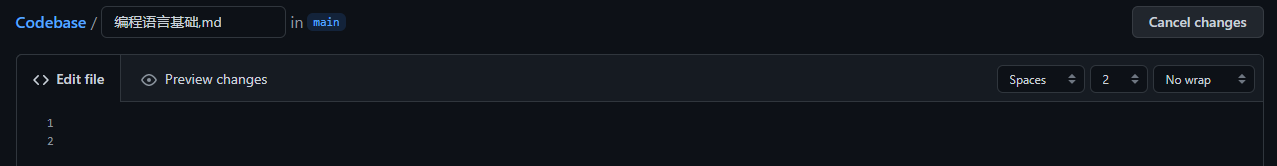


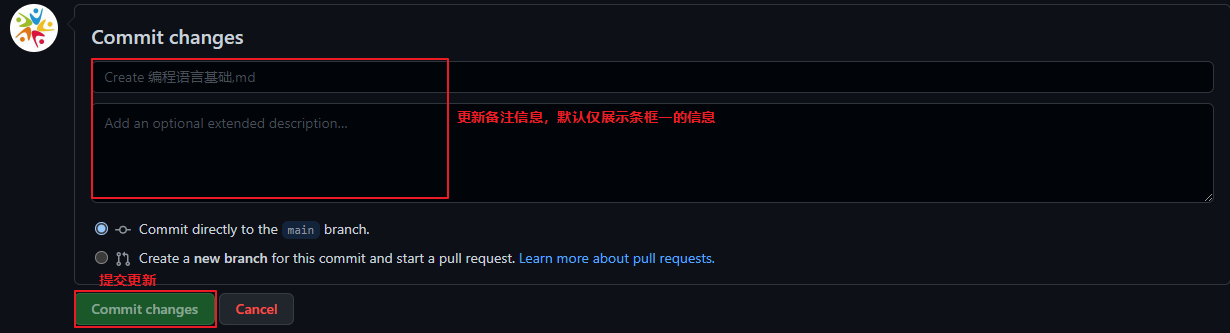
Github项目库中的项目结构为branch -> directory -> file对于我们来说，主分支main就足够使用了，我们可以在该分支下创建，导入文件夹以及文件。在项目目录区域我们可查看各文件夹/文件最近的更新情况，点击右上角的11 commitis我们可查看所有文件夹/文件的历史更新情况。右侧的stars和forks即为别人的收藏，拷贝数。下面我们着重来介绍创建，导入，修改文件夹/文件以及更新备注功能

3.1 修改文件夹/文件及更新备注信息



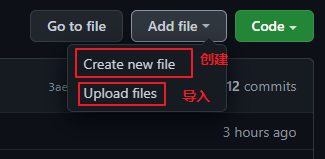


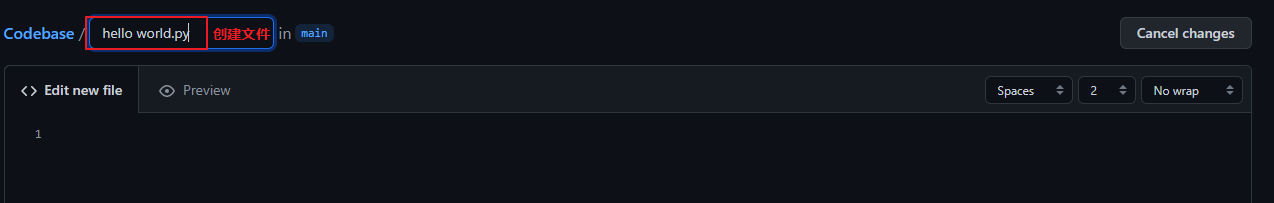


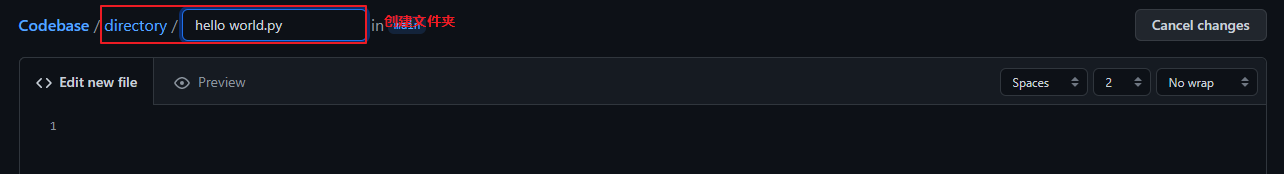


点击 go to file，我们以编程语言基础.md文件为例，在点击编辑之后，即可在Github自带的编辑器中进行内容编辑，最后在底部添加更新备注信息并提交更新，更新备注信息可以是关于内容修改的细节信息，也可以是方法，思路，功能的变更等等，目的在于便于后期项目回溯。

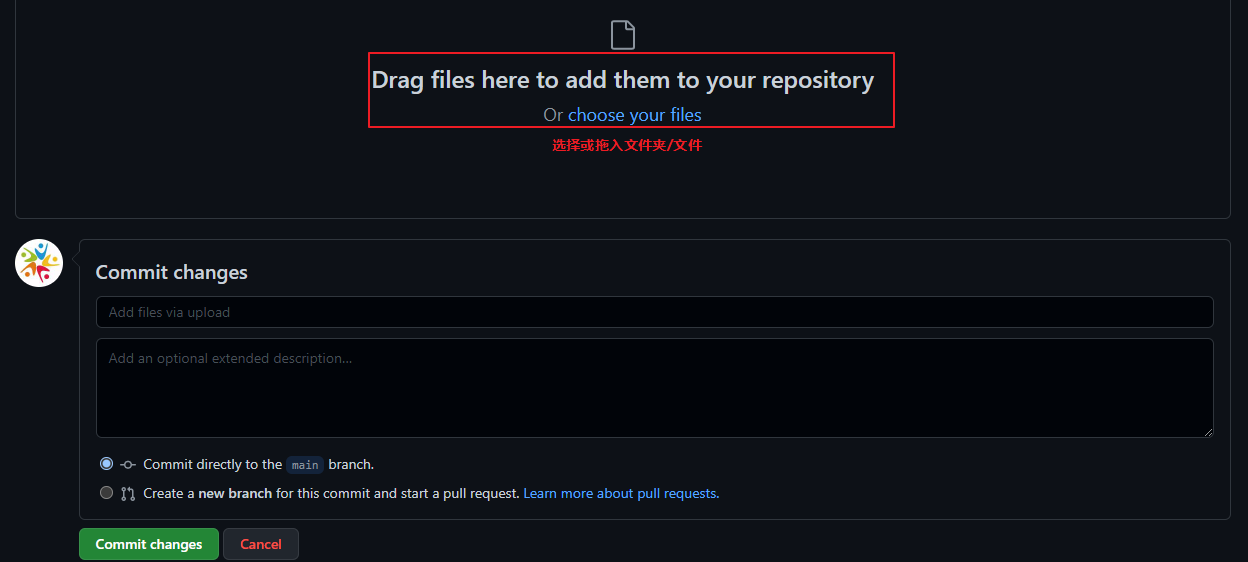
3.2 创建及导入文件夹/文件







点击create new file便可创建文件夹/文件，其中输入“文件名”为创建文件，输入“文件夹名/”为创建文件夹（注意不可创建空文件夹）。创建之后在下方便可以编辑（当然除markdown文件外一般我们不在这里编辑），最后在底部可更新提交。



点击 upload file便可如上直接导入文件。

**Git（命令行项目管理）**

Git是目前世界上最先进的项目分布式版本控制系统！在实际工作中，尽管可以使用Github这一图形界面网站，但要想快速实现远程项目的上传下载、版本记录与恢复、还得看Git。Git的使用分本地和远程（线上）两种，实际工作中主要接触后者，但前者是后者的基础。

1.Git基本概念

（1）Git的使用涉及到4个基本概念：工作文件、工作区、缓存区、版本仓库。

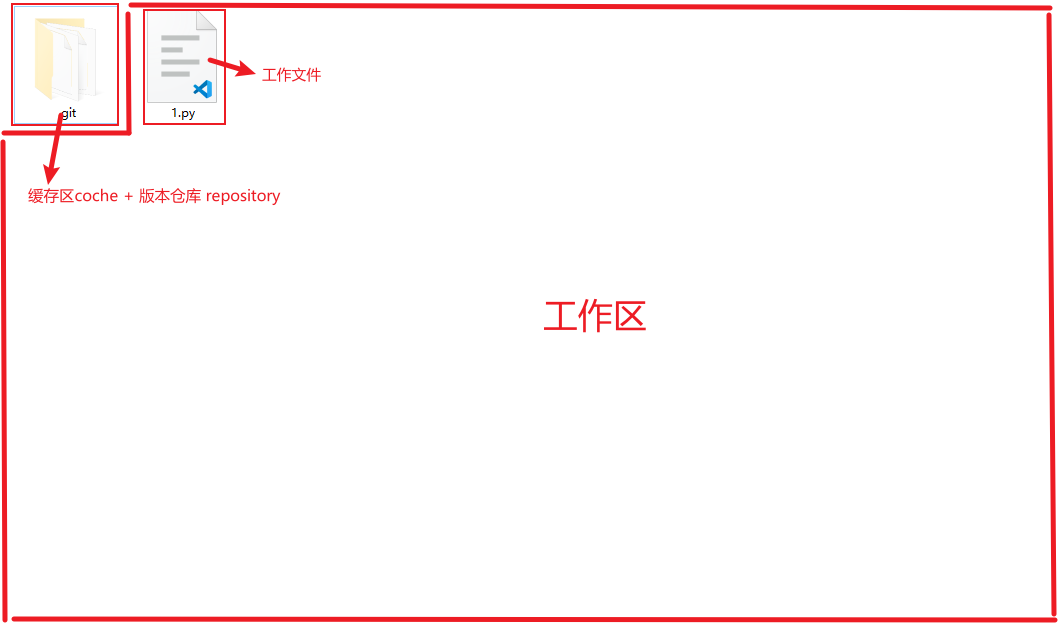
工作文件：待管理的项目或项目文件

工作区：工作文件所在目录（路径）

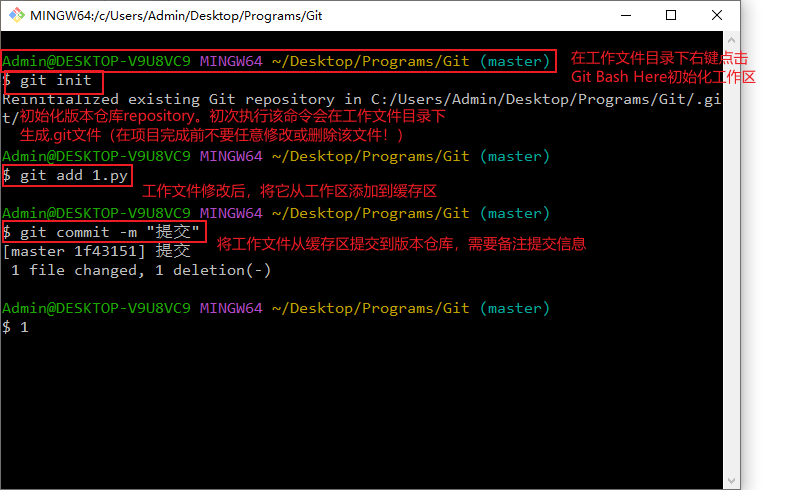
缓存区：Git中工作文件修改后，不能直接提交到版本仓库，需要先添加到缓存区，作用类似于购物时的购物车

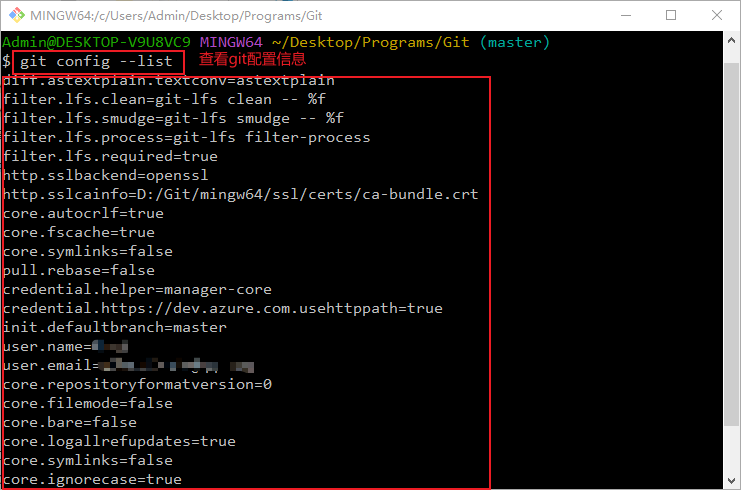
版本仓库：以版本记录的形式存储工作文件的修改信息，版本信息存储在版本仓库的HEAD文件里

Git的使用整体遵循如下流程：修改工作文件（除第一次外）-> 跟踪工作文件 -> 提交工作文件

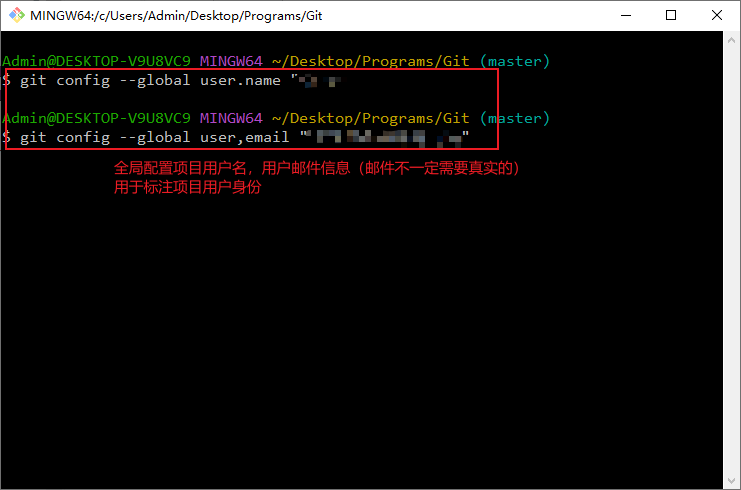


（2）下面来介绍工作区、缓存区、版本仓库的初始化，以及Git用户信息配置：





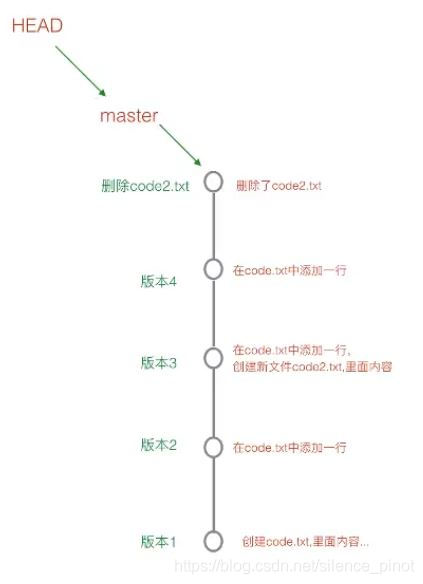
摁shift+Q退出



（3）最后是Git的分支介绍：

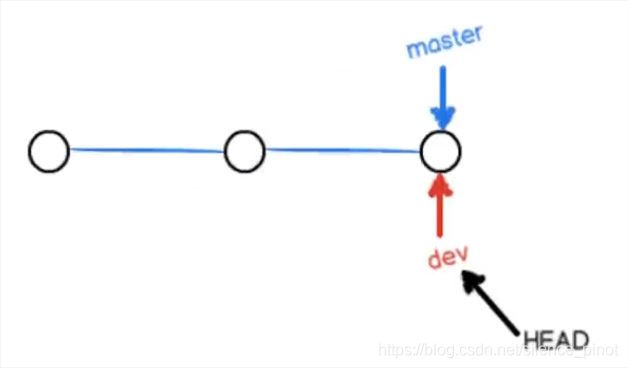
对于一个项目，如果仅仅是一个人开发，或者多人轮流开发，则实际上并不需要使用到Git中的分支（branch）工具，但考虑到多人同时开发一个项目这一现实，常常需要我们熟悉分支这一基本概念：

任一经过git初始化的项目都有一个默认的主干master或main，主干的任务是确认并保存项目的最终修改情况，如下：

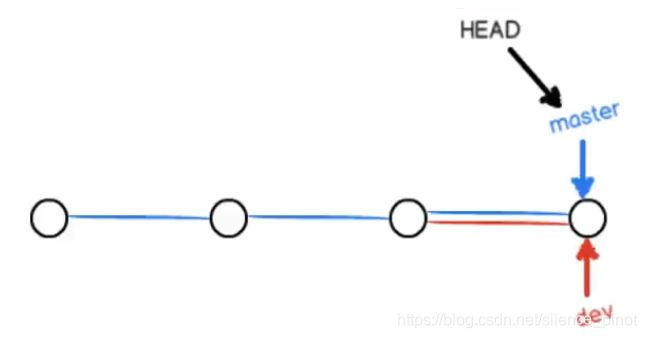


其中HEAD为指针，作用是定位于某分支，或分支的某一历史版本。

若另开一分支dev，则dev会继承当前主干master所有关于项目的历史版本信息。我们可从主干master切换至分支dev，如下：



现已修改并提交dev分支下项目内容，我们想要将dev分支信息完全合并至主干master，则只要两者不产生冲突，即可如下操作：



即切换回主干master（HEAD指针位于master），并将dev合并至master中。

然而当两者产生冲突时（如master和dev同时修改了某一相同模块函数，那么版本信息就产生了内容冲突），则无法合并：



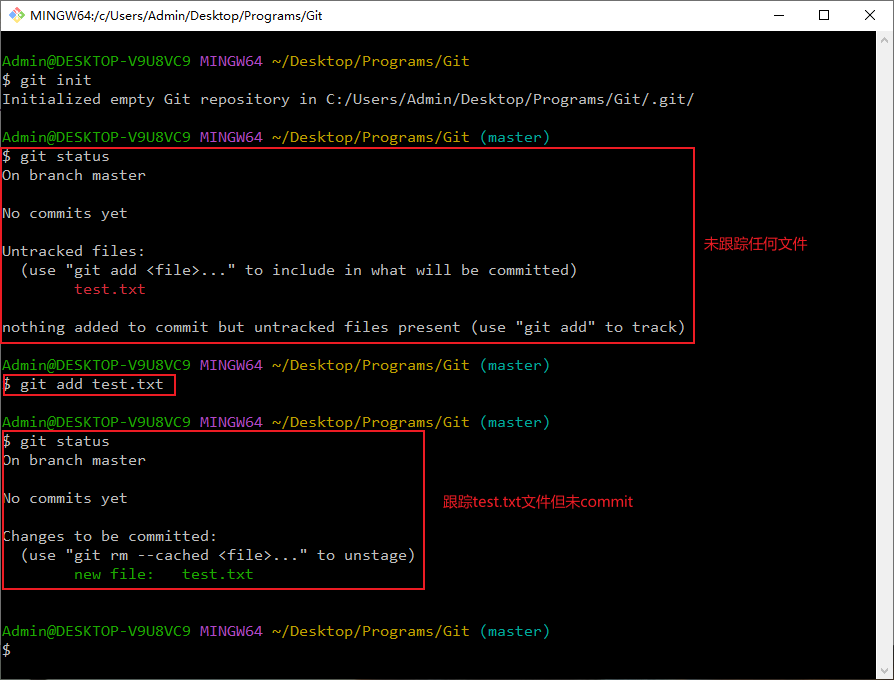
此时需要对比修改两个分支的文件，直至满足合并的要求。

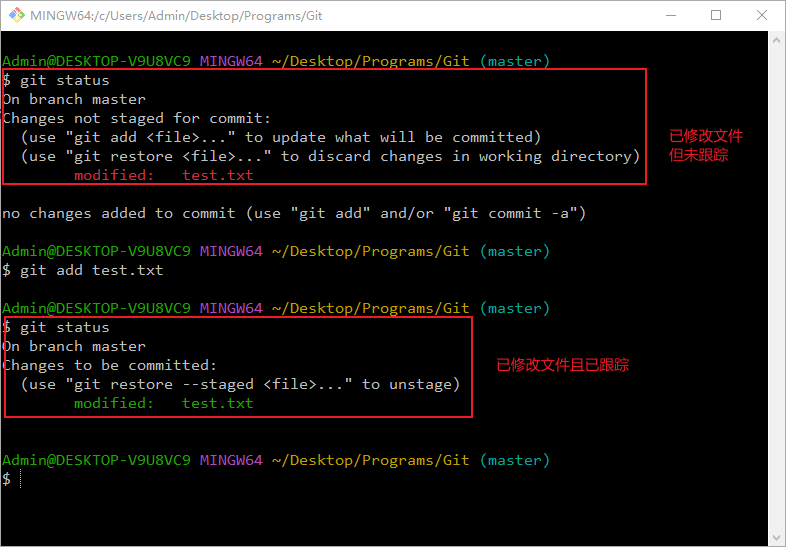
2.Git 本地基本操作

1. 查看当前工作状态

git status

工作状态可分为四种：未跟踪工作文件（即没把文件添加到缓存区）、跟踪但未提交工作文件、文件已修改（指文件信息和版本仓库中记录不一致）但未跟踪、文件已修改且已跟踪。如下：





1. 添加工作文件至缓存区、从缓存区删除工作文件、提交文件

添加部分工作区文件：git add file1 file2 ……

添加工作区所有文件：git add .

删除部分缓存区文件：git reset HEAD file1 file2 ……

清空缓存区所有文件：git reset HEAD .

删除本地工作区文件：git rm file1 file2…… （不推荐直接在本地删除文件）

提交文件更新信息：git commit -m “备注信息”

注意：除初始化git版本仓库外，若工作文件未曾修改则无法跟踪（跟踪没意义）。已经跟踪的文件不能修改，否则需再次跟踪

注意：另一种清除缓存的类似方法为git rm --cached xxx，与 git reset HEAD xxx有所不同。

1. 查看历史版本记录，版本回退

自HEAD起查看历史版本记录：git log 或者git log --oneline （一行展示）

查询HEAD更改前的历史版本记录：git reflog 或者 git reflog --oneline

版本回退操作：git reset --hard 提交编号

注意：版本回退是将工作区回退至指定版本的状态，而不只是针对某一文件！

1. Git的分支管理

查看所有分支：git branch

创建新分支：git branch branch\_name

删除分支：git branch -d branch\_name

切换分支（改变HEAD指针位置）：git checkout branch\_name

合并其他分支至当前分支（HEAD所指向分支）：git merge other\_branch\_name

3.Git+Github线上操作

在实际工作中，我们更多需要与他人协作开发项目，因此了解Git远程项目管理是必须的，也即了解如何远程管理Github上的repository：

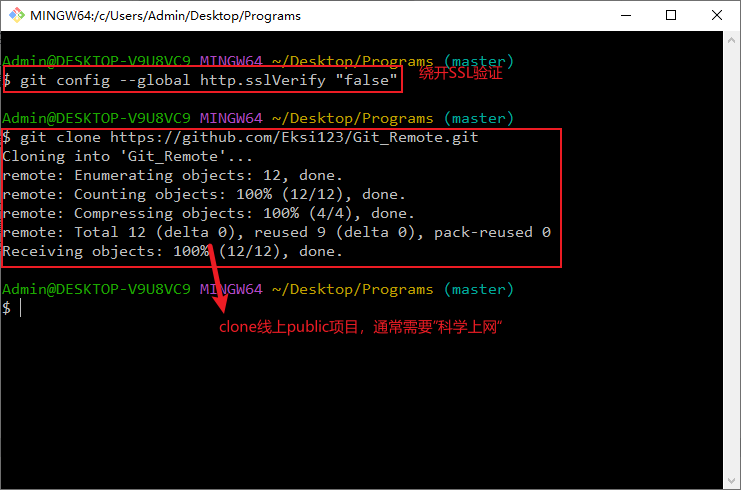
1. 克隆Github中的项目

git config --global http.sslVerify "false"

以public repository为例，由于线上Github与本地Git的连接还需要经SSL验证，因此在非必要前提下，可以直接绕开验证。

git clone https://github.com/Eksi123/Git\_Remote.git

输入git clone + repository的url连接，即可直接将线上项目克隆至本地。如果repository是private，还需要一个身份验证过程



1. Push和Pull

将Github上的项目clone至本地后，我们就可以直接在克隆的项目中操作，基本操作上面已有讲述，下面介绍如何完成克隆项目的线上线下更新：

无其他分支情况下：

git push 将本地克隆项目的修改信息更新至线上原项目

git pull 将线上项目的修改信息更新至本地克隆项目

有其他分支情况下：

git push origin branch\_name 默认将本地分支信息更新至线上同名称分支

git push --set-upstream origin branch\_name 若线上没有与本地相同名称的分支，则执行此代码

git pull origin branch\_name 默认将线上分支信息更新至本地同名称分支

注意：git push origin: branch\_name 表示删除线上分支

# Mysql +Redis（传统数据库）

作为偏底层的应用程序，数据库能够对文件数据的存取，编辑、权限安全等进行高效管理，故而通常被作为服务器级程序部署于计算机上，并作为其他程序，网页开发运营的底层支撑。数据库可分为传统型数据库（如Mysql、Redis、sql server、oracle、MongoDb等）和大数据仓库（如Hadoop），本文主要就开源的传统型数据库Mysql和Redis进行介绍。

Mysql和Redis分别由sql语言（结构化查询语言）和C语言所开发，前者为关系型数据库，适用于磁盘上的数据进行常规存取操作；后者为非关系型数据库（缓存数据库），适用于内存上的数据进行频繁高速存取操作。关于两者的区别与联系，具体可见：https://blog.csdn.net/weixin\_43520450/article/details/108778289?spm=1001.2014.3001.5506

**Mysql**

**1 mysql登录及设置**

1.1 登录mysql客户端，连接本地服务器

Cmd命令框：mysql -hlocalhost -uroot –p187733 （其中h为 服务器所在计算机IP地址 ，服务器在本机上则用localhost。u为user，root为用户名。P为password， 187733为密码）

注1：保险起见，一般先输入用户名mysql -hlocalhost –uroot -p，回车后匿名输入密码

注2：断开客户端与服务器连接（关闭客户端），可在命令行输入 quit、exit、或\q

注3：通常mysql本地服务器都是默认自启动的，若要手动启动与退出，则在cmd管理员模式下命令为：

net start mysql 与 net stop mysql

注4：中断某一命令用 \c

1.2 远程连接到其他计算机的mysql服务器

假如远程计算机（IP地址为110.110.110.110）已为本地计算机申请客户端用户， 用户名为root，密码为abcd123，则远程连接方式为：

Mysql -h110.110.110.110 –uroot –pabcd123

注：如果忘记了客户端登录密码，本机上可以使用免密码登录，具体教程网上有许多，可参考CSDN博客（计算机用户名为中文可能会报错）

1.3 修改mysql用户名和密码

登陆后

修改用户名 update mysql.user set user="new user" where user="old user";

修改密码 update mysql.user set authentication\_string=password('new password') where user='username';

最后刷新权限 flush privileges;

注：登录mysql客户端后，所有命令要加分号

1.4 增删用户及权限设置

查看用户基本信息：select user, host, authentication\_string from mysql.user;

增加用户：create user username@'IP name' identified by 'password'; 创建时依次要求登录用户名，主机IP地址（也可以是本机localhost）和登陆密码

注1：本地的mysql客户端默认具有最高权限，可以为服务器增删用户及权限设置

注2：当新建用户主机为localhost本机时，用户名不可以和原先的重名

修改用户密码：alter user username@'IP name' identified by 'password';

删除用户：drop user username@' IP name';

查看用户权限信息：show grants for username@'IP name';

常见权限信息有：

Usage：无权限

select：可以从数据库中的表中查询数据

insert：可以插入数据到数据库的表中

update：可以对数据库中的表进行更新数据

delete：可以从数据库中的表中删除数据

alter：可以使用 alter table 来改变表的结构

create：可以创建新的数据库或表的语句

drop：能够删除现有数据库，表和视图

grant option：可以向其它用户授予或移除权限

index：可以创建和删除索引

all或all privileges：代表指定权限等级的所有权限

赋予用户权限：grant privilige-name on \*.\* to username@'IP name'; 权限赋予依次要求权限名，作用域和用户

回收用户权限：revoke privilige-name on \*.\* from username@'IP name';

注：作用域\*.\*默认为全域，也可指定为某数据库 database-name.\* ，或某数据库下某表database-name.table-name\*

注：通常来讲，为避免误操作，应尽量减少使用本机上的root用户，最好的办法是新建一个恰当权限的用户

最后别忘了刷新权限 flush privileges;

**2 mysql常用命令**

2.1 数据库，表的简单查询

展示所有数据库：show databases；

切换某数据库：use base-name; 注：select database(); 可用于查看当前所在数据库

展示该库下所有表：show tables;

展示该库下某表的字段信息：describe table-name;

展示该库下某表某字段记录值：select variable-name from table-name; 注：若variable-name = \* 则为展示全部字段的值

2.2 数据库、表的创建

创建数据库：create database base-name character set utf8; 注：考虑到兼容性，通常编码设为UTF8

创建表：use base-name;

create table table-name(variable1 type1, variable2 type2, variable3 type3 ……);

注：variable表示字段名，type表示字段类型，常用字段类型有INT、FLOAT、DOUBLE、CHAR(X)、VARCHAR、DATETIME等

注：若要求表含有自增长的一个主键（一般为id号），则可添加字段id Int Unsigned Primary Key Auto\_Increment。

增加字段：alter table t1 add v-name v-type;

删除字段：alter table t1 drop v-name;

修改字段名称类型：alter table table-name change old-v-name new-v-name new-v-type;

增加记录：insert into table-name(v-name1, v-name2) values(value1, value2); 注：为指定字段增加记录。不指定则如下：

insert into table-name values(value1, value2, ……); 还可同时插入多条记录，如下：

insert into table-name values(……), (……), (……) ……;

修改记录：update table-name set v-name=new-value, …… where ……; 注：where 后面的内容为v-name=value 用于检索

update table-name set v-name=new-value, …… 若无where条件语句，则相当于无条件修改字段值

删除记录：delete from table-name where …… 注：where作用同上，若无where，则清空整个表

删除表：drop table table-name;

注：安全起见，一般不支持修改数据库名称。可修改表名称：alter table old-name rename new-name;

注：若创建表时忘记设置主键，则可设置表中某字段为主键：alter table table-name change var-name var-name int primary key auto\_increment。

另外，经常还需要修改当前主键值以保持各记录ID连续：alter table table-name auto\_increment = xxx

删除数据库：drop database base-name;

2.3 数据的条件查询

上小节中涉及到了where查询语句的简单使用，本小节进一步拓展其应用：

关系运算符：where variable ( = , != , < , <= , > , >= ) value;

in关键字：where variable in (value1, value2, ……); 注：variable在列表中所有出现的取值将视作条件

between and关键字：where variable between value-x and value-y; 注：variable在值x，y之间的取值将视作条件

空值：where variable is (not) null; 注：variable为（不为）空值将视作条件

and or 关键字：where variable variable1 – value1 and (or) variable – value2; 注：且条件，或条件

like 关键字：where variable like ‘xxx’; 注：用于字符型变量匹配

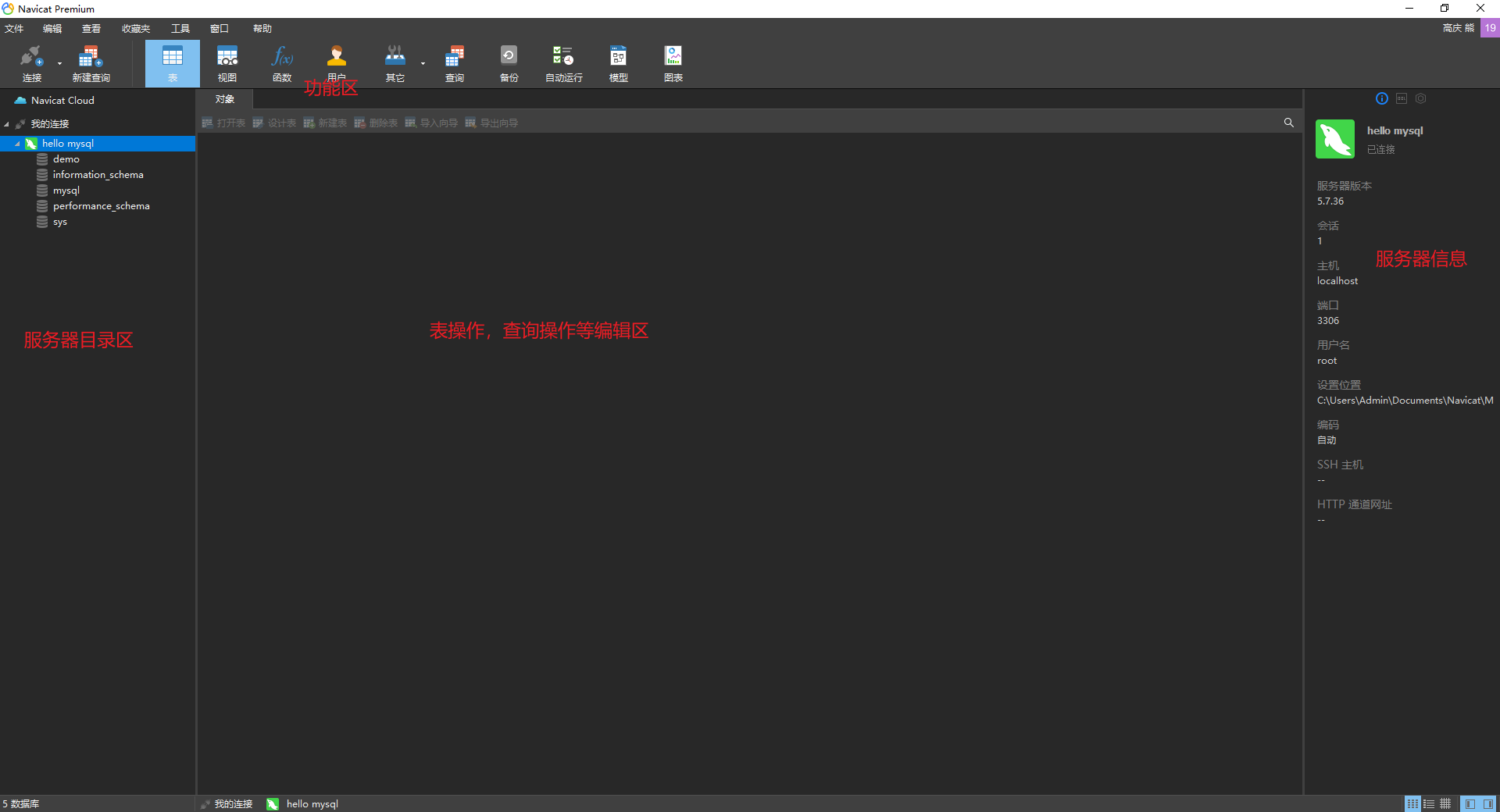
2.4 数据的导入与导出

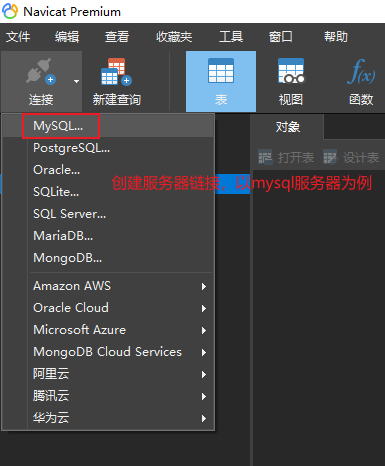
为避免导入sql或其他类型数据出错，该部分建议采用mysql图形界面工具Navicat完成，十分简洁。

**3 Navicat基本操作**

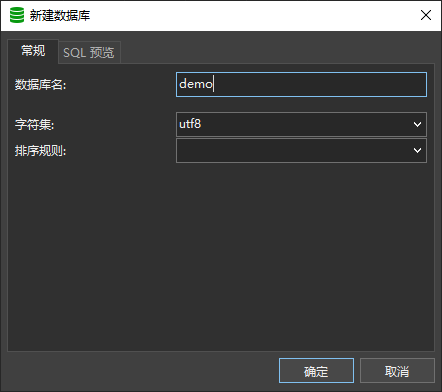
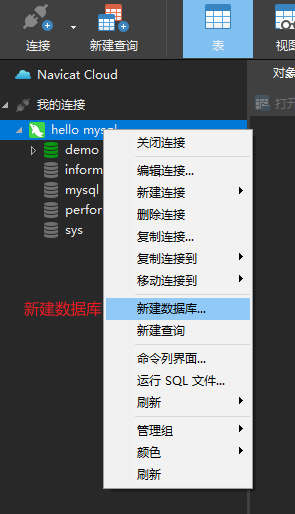
**Navicat是一套可创建多个不同数据库连接的图形界面操作工具，其基本操作如下：**

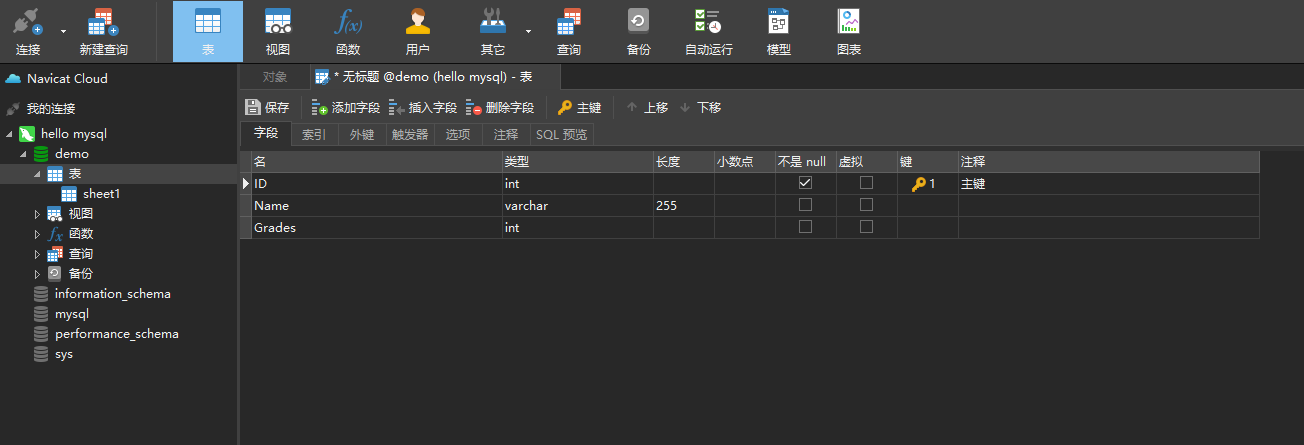
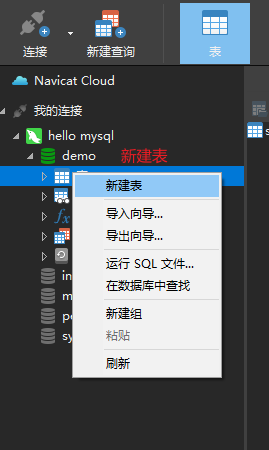
3.1 创建连接

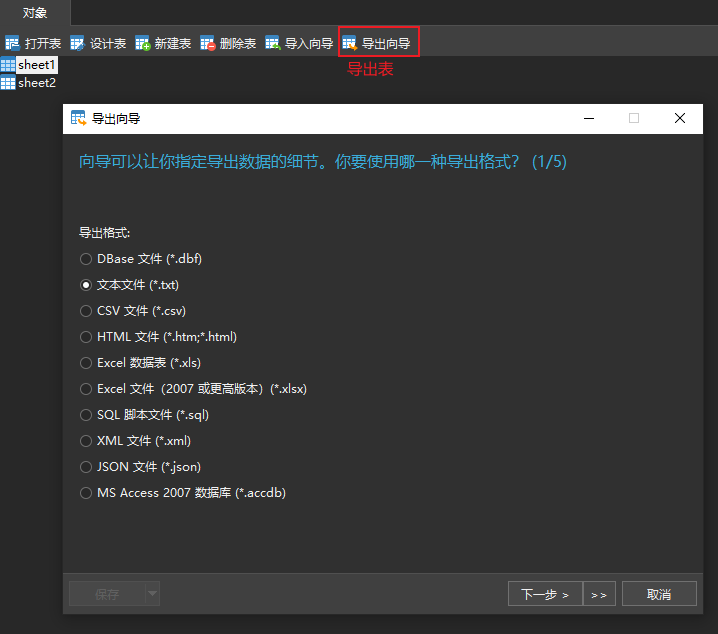




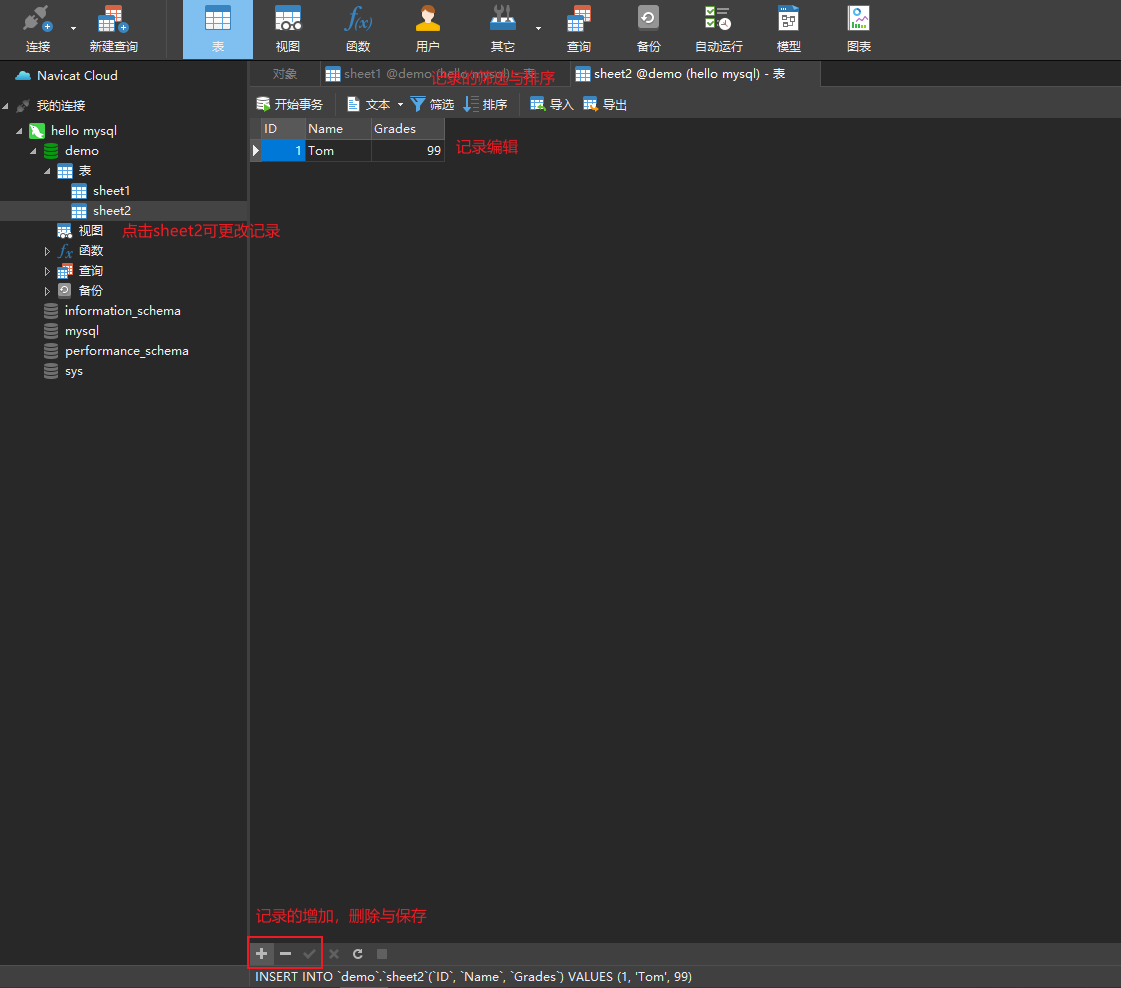
3.2 创建表，导入表，导出表



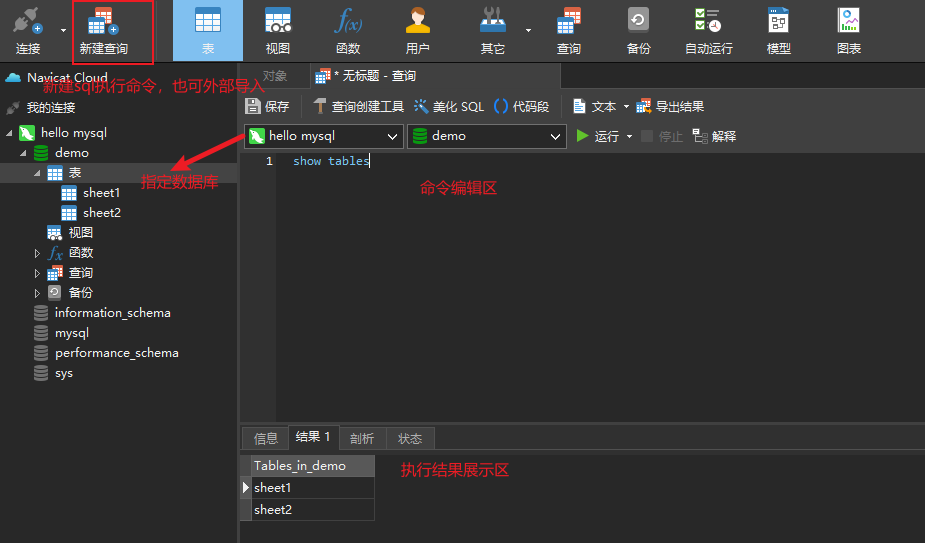




3.3 记录的增删查改



3.4 mysql语句执行



**4 mysql中的锁的介绍与设置**

**Redis**

（1）Redis数据库与Mysql数据库区别：

首先，两者均为数据库，Mysql为关系型数据库（表格型，行列定位）而Redis为非关系型数据库（键值型，行定义）；其次，Mysql一般用于海量数据存取，故数据放在磁盘中，而Redis用于少量且常用的数据存在，故数据放在“缓存”里（读取到内存很快）

（2）登录数据库

Cmd控制台

cd C:\Redis-x64-3.0.504

redis-server.exe redis.windows.conf

打开新的cmd控制台

cd C:\Redis-x64-3.0.504

redis-cli

或者直接在vscode上连接，可查看数据集

（3）键值输入输出

set key “value”

get key

\*其原理类比于散列表，哈希表

# Linux 操作基本操作

作为长期存在的一款开源操作系统，Linux广泛受到程序员的偏爱，原因在于比之其他如Windows、Mac商用操作系统，Linux系统能满足“命令执行更快更稳定”、“系统控制设置更自主”、“开源免费”等项目开发基本要求，这大大拓宽了Linux系统在“程序员”及相关业务圈的发展。

Linux旗下有Centos、Ubuntu、RedHat等系统，本文只对linux基本通用操作做介绍，比较推荐的方式是通过虚拟机软件VMware workstation来建立Linux系统。下面我们主要介绍Linux的远程访问（建议通过FinalShell实现）和Linux的基本操作

注：FinalShell远程连接使用方法：<https://zhuanlan.zhihu.com/p/589218854>



**1 文件（夹）打开显示**

pwd: 查看当前文件（夹）路径

cd : 用于切换文件夹路径（注意不能切换到某个文件）

绝对路径切换：cd /home/xgq/.cache 即切换至/home/xgq/.lcache 这一文件夹路径下

相对路径切换：cd ../.local 即切换至/home/xgq/.local 这一文件夹路径下，其中cd ..用于返回上一级路径 cd.为当前路径

其他：cd - 返回上次的目录； cd / 返回root目录； cd ~ 返回home目录

ls 查看当前目录下文件

ls -l 查看当前目录下文件及详细信息（如大小，修改时间等）

ls -a 查看当前目录下所有文件夹和文件（包括隐藏文件）

vim file: 用于查看编辑文件（目前感觉使用不多），默认进入命令模式，该模式下可对文件内容复制粘贴等

编辑模式：命令模式下+i (insert) 在光标之前插入内容；命令模式下+a (append) 在光标之后插入内容，esc退出编辑

退出：输入:wq+enter 保存内容，输入:q! +enter 不保存内容，输入:q +enter退出，输入:! +enter强制退出

cat file: 仅查看文件

more/less file: 多行文件分页查看

**2文件（夹）创建，删除，复制，移动 ，查找，解压缩（注意以下操作均需明确路径）**

touch file：创建文件

rm file：删除文件

mkdir dir/dir1：在当前目录下创建文件夹（允许多级创建）

rmdir dir/dir1：删除空文件夹

cp file dir：复制文件至文件夹

cp -r dir1 dir：复制文件夹至另一文件夹

mv file dir：移动文件至文件夹

mv file ./file1：将file重命名为file1

mv dir1 dir：移动文件夹至另一文件夹

find -name file/dir：搜索文件（夹）所在路径

grep ‘匹配项’ file：搜索文件内所匹配内容

tar -cvf xx.tar file/dir：把文件（夹）压缩为xx.tar

tar -xvf xx.tar：解压xx.tar

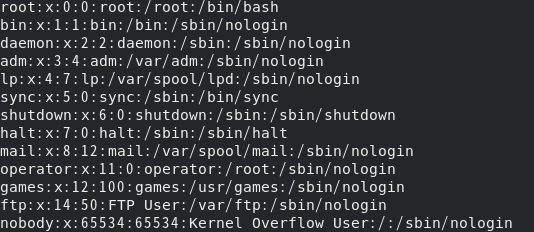
tar -xvf xx.tar -C dir：解压xx.tar至指定文件夹

\*类似压缩和解压还有zip, rar等

**3用户管理、关机重启**

出于安全和实际需求考虑，linux允许赋予不同用户以不同的权限。Linux里面有三种用户：【1】超级管理者：拥有linux系统的一切权限 【2】普通管理者：适用对象如开发人员，拥有部分权限，通常不能修改系统底层配置。【3】程序用户或服务用户，适用对象如客户端用户，通常就基于linux系统所开发的服务进行使用。

cd etc + cat passwd 或 cat etc/passwd：查看linux用户账号信息





更多信息，可查看如下博客：

[https://blog.csdn.net/weixin\_54055099/article/details/125880255?ops\_request\_misc=&request\_id=&biz\_id=102&utm\_term=linux%E7%94%A8%E6%88%B7%E6%9D%83%E9%99%90%E7%AE%A1%E7%90%86&utm\_medium=distribute.pc\_search\_result.none-task-blog-2~all~sobaiduweb~default-0-125880255.142^v86^insert\_down1,239^v2^insert\_chatgpt&spm=1018.2226.3001.4187](https://blog.csdn.net/weixin_54055099/article/details/125880255?ops_request_misc=&request_id=&biz_id=102&utm_term=linux%E7%94%A8%E6%88%B7%E6%9D%83%E9%99%90%E7%AE%A1%E7%90%86&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-0-125880255.142%5ev86%5einsert_down1,239%5ev2%5einsert_chatgpt&spm=1018.2226.3001.4187)

su username：切换用户，如果是切换到root用户需要输入密码。Ctrl+d退出当前用户

sudo + 执行命令 + 输入当前用户密码：临时使用root权限

useradd username + passwd username（需重复输入密码）：在root用户或root权限下可创建新用户

userdel -r username：在root用户或root权限下删除用户及文件目录

shutdown -h：关机命令

shutdown -h +10：10分钟后关机

shutdown -h 12:00：指定12:00关机

shutdown -h now：立刻关机

shutdown -r：重启命令，使用方法同上

**4软件下载与更新**

yum是linux中非常便捷好用的应用管理器，其基本命令如下：

yum list：查看yum中安装的所有应用

yum -y install 应用名：下载应用并回复yes

yum -y update 应用名：更新应用并回复yes

yum -y remove 应用名：卸载应用并回复yes

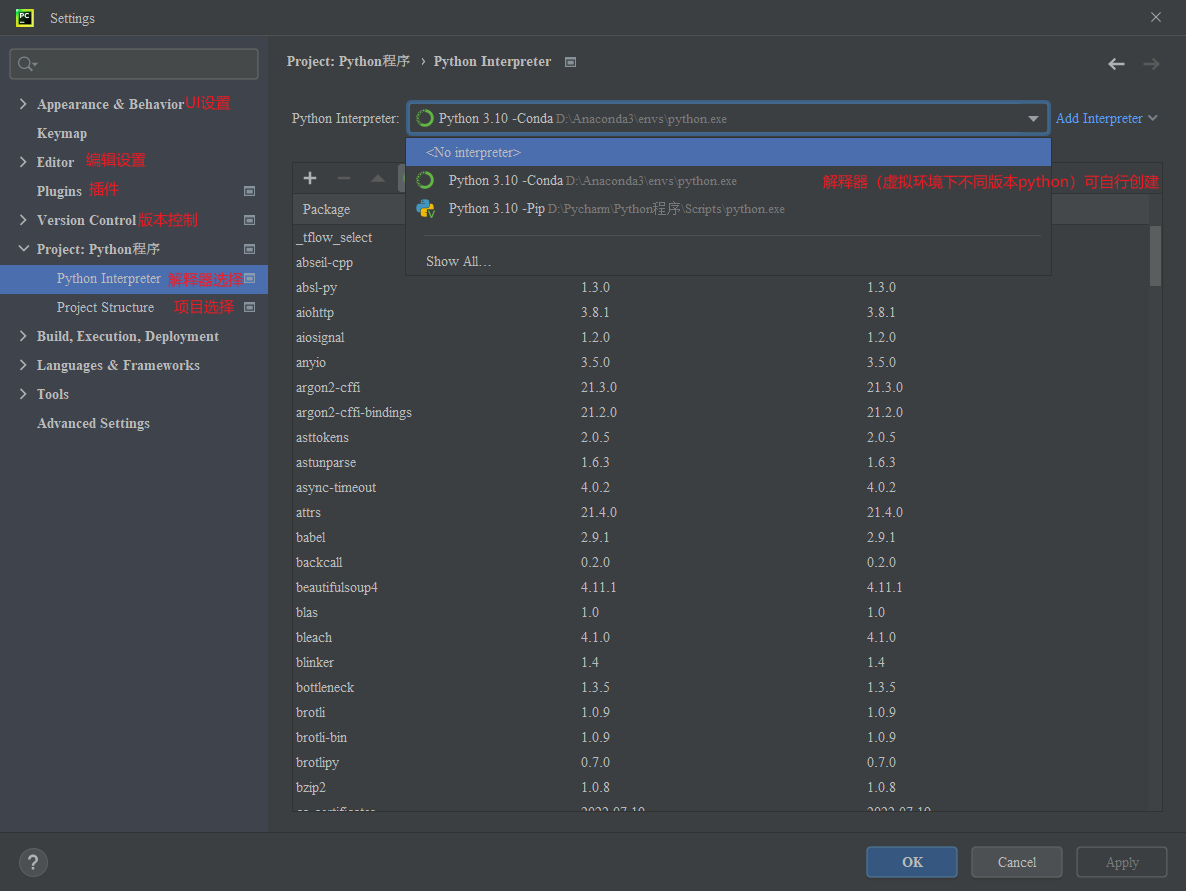
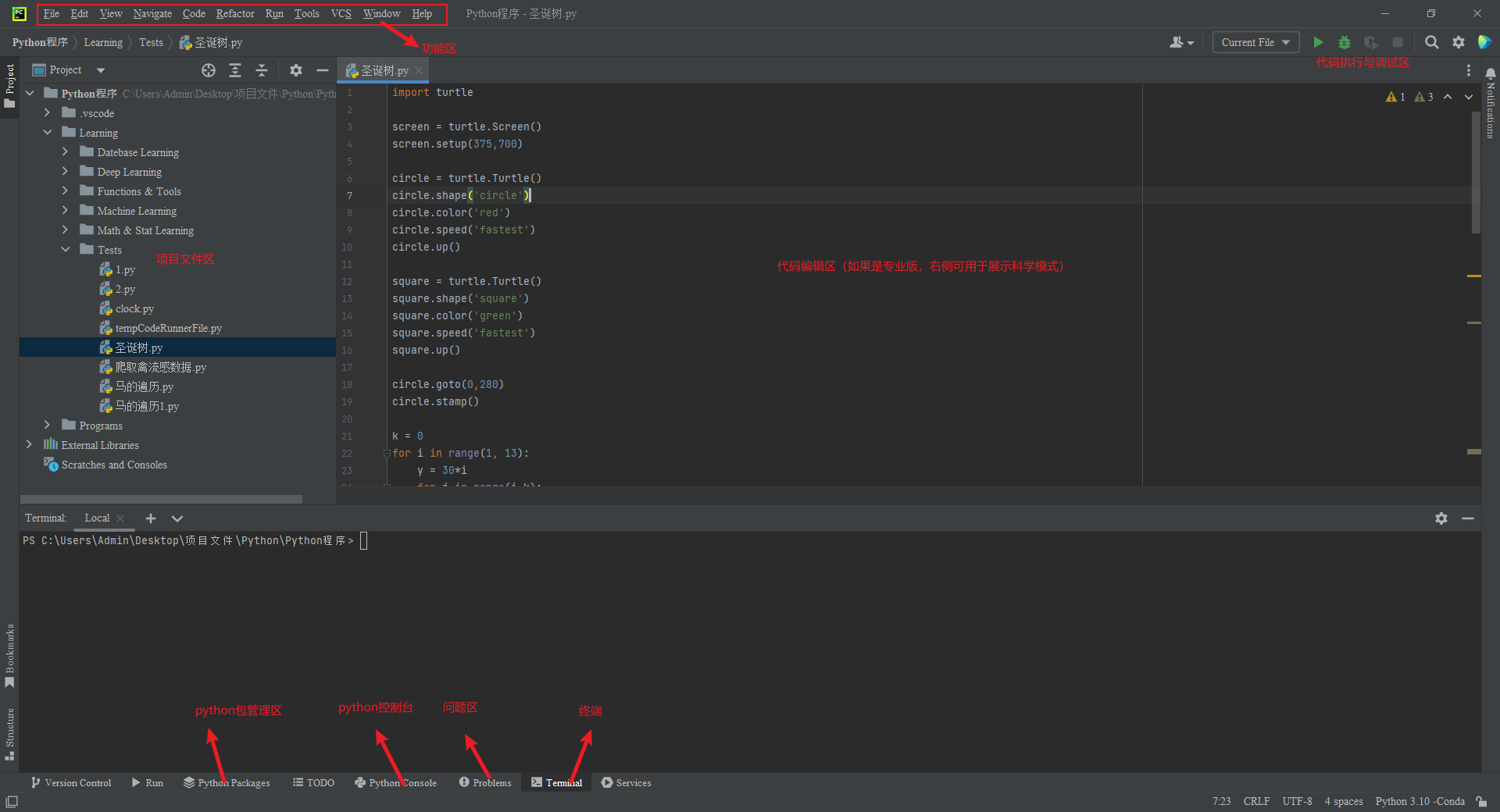
# Pytorch深度学习框架搭建

要想实现如标题所言操作，需要理解如下几个概念：

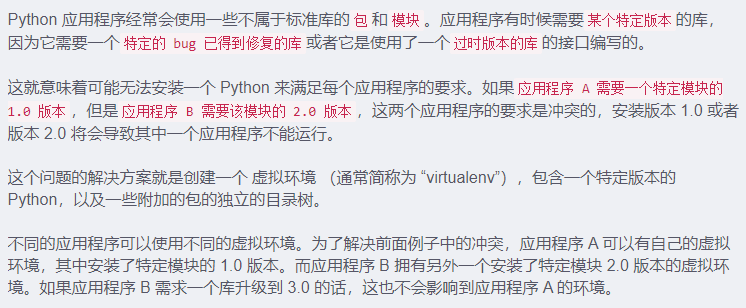
1 Pycharm-IDE 2 conda环境管理器 3 GPU和cuda 4 pytorch深度学习框架

1. Pycharm和conda

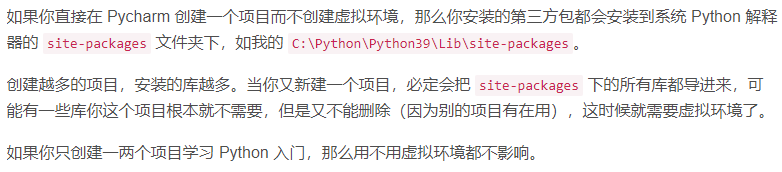
Pycharm是一款高度集成的Python-IDE，表现为Python编译器+Python包管理器+Python代码管理器+Python应用开发工具……，原理与安装了众多插件的vscode差不多，此处介绍一下Pycharm的常用功能：



基本功能与vscode并无太大差别，其主要突出点在于通过创建虚拟环境来调控不同版本下的python及python包，这就引出如下话题：什么是虚拟环境，虚拟环境有什么用？

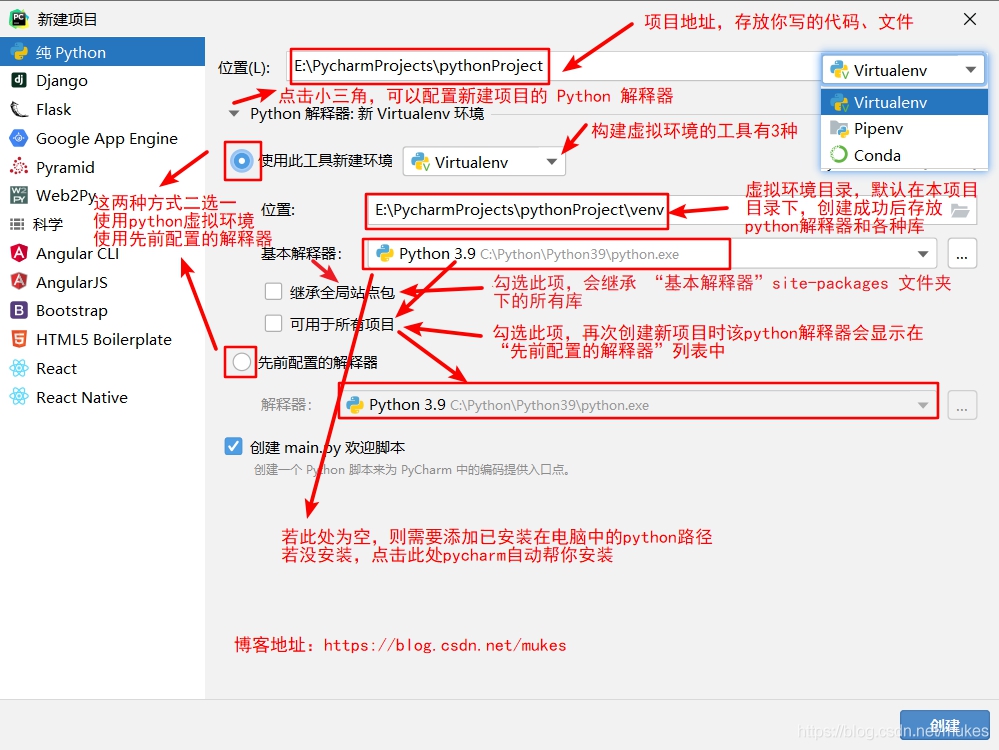


虚拟环境和非虚拟环境如何选择？

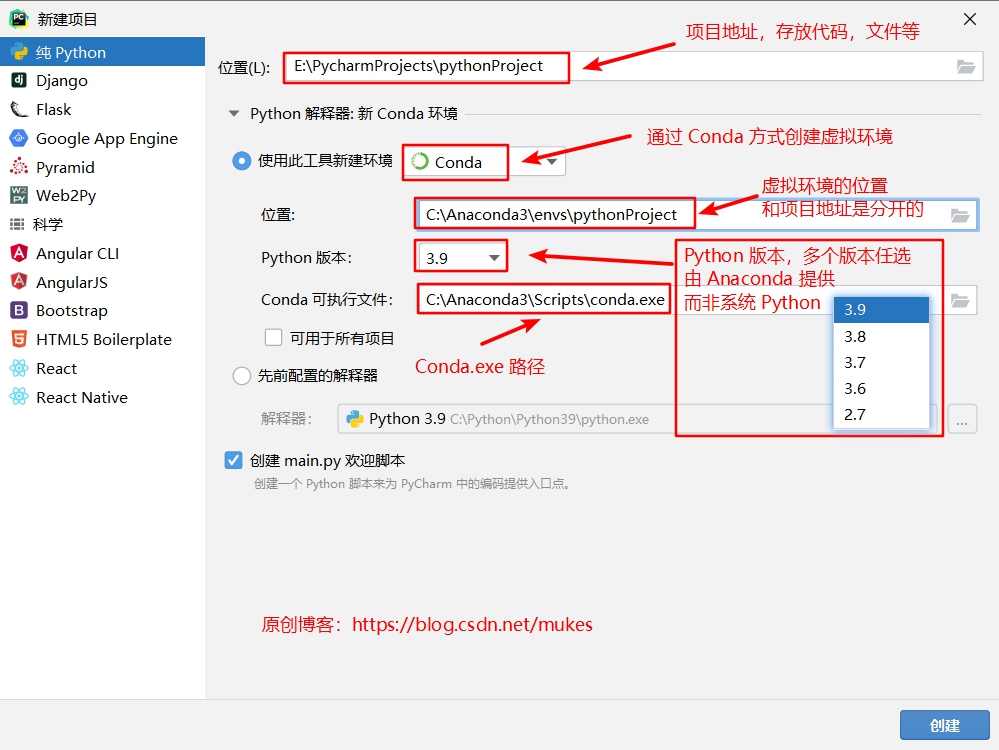


Pycharm中创建虚拟环境主要为如下两种方式：

1. 基于Virtualenv创建虚拟环境（以下为网图）:



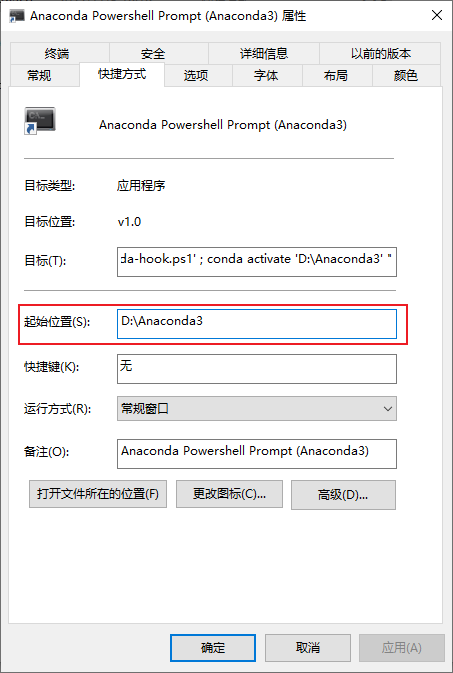
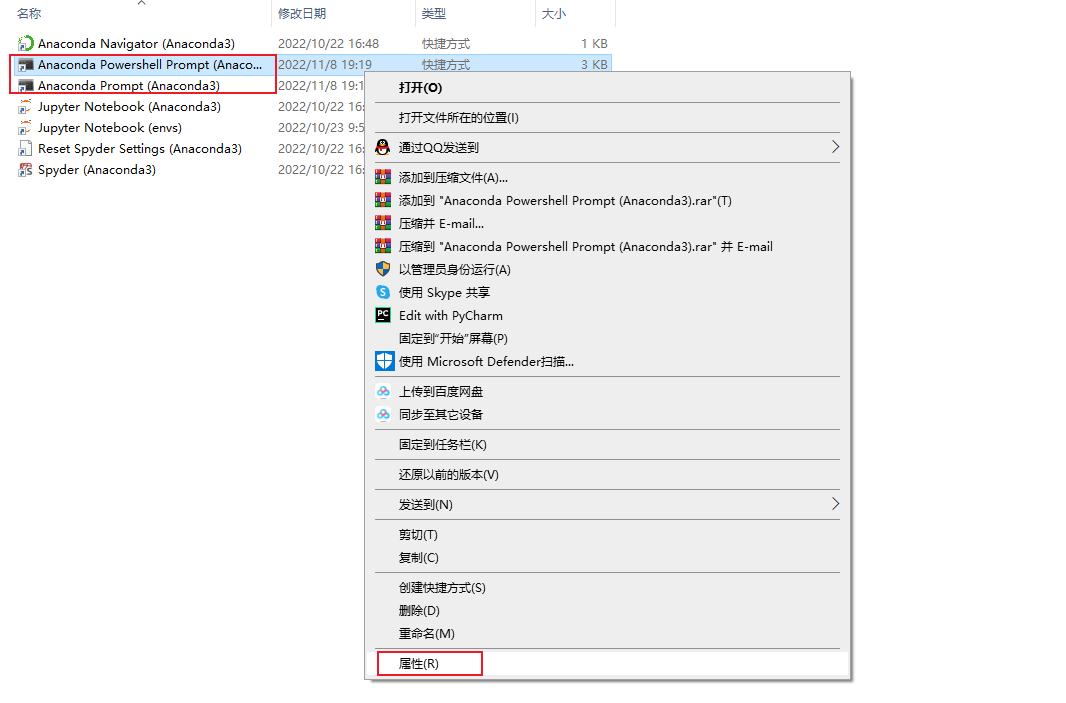
2 基于conda创建虚拟环境（以下为网图）：

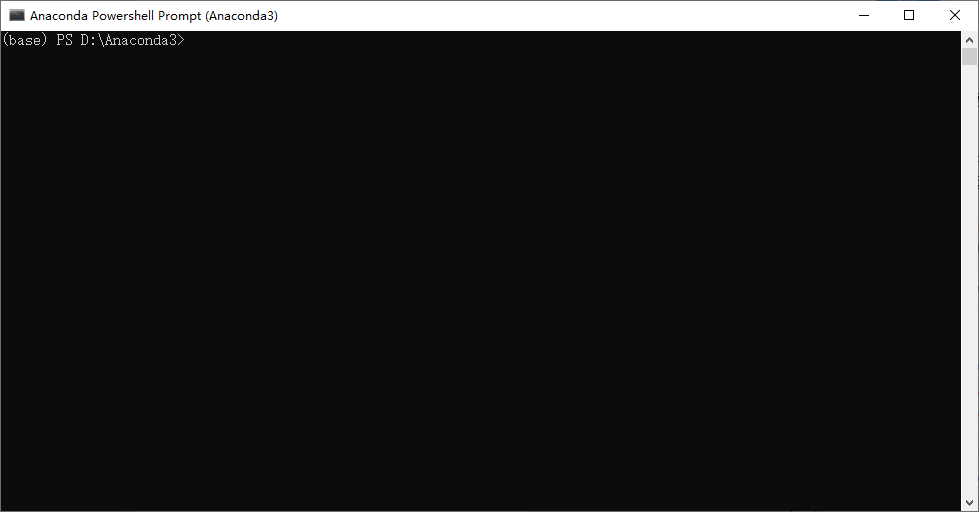


虚拟的python解释环境，以及虚拟环境下python包与项目是直接相关的，可以在Pycharm或conda的envs文件中找到，若某个项目被删除，则附属它的虚拟环境也会被删除。另外，conda包管理器通常嵌入于Anaconda或Miniconda中，因此若想用conda来创建虚拟环境，需要先下载Anaconda或Miniconda。通常我们考虑用conda来创建虚拟环境，一个重要的原因在于通过conda可较方便下载我们所需的深度学习框架（深度学习包）Pytorch和Tensorflow。

除pycharm外，还可用conda自带的命令行控制台实现虚拟环境和包的管理，了解这一点有助于后续我们下载GPU版本下的Pytorch：

**切换文件路径：**conda命令控制台默认初始化文件路径为C盘的user用户，但通常我们将conda安装文件放在其他盘，因此需要设定初始路径为conda文件路径，这一点非常重要，不然容易出现安装的包（如Pytorch）不在对应虚拟环境的情况，如下：





可以看到此时命令台初始文件路径为conda文件路径。下面我们给出常用的命令操作代码：

创建虚拟环境和python编译器：conda create --name env\_name python=xxx （版本）

删除某虚拟环境：conda remove --name env\_name --all

查看所有环境信息：conda env list

激活某虚拟环境（默认激活base）：conda activate env\_name （退出环境用 conda deactivate）

查看当前环境下所有包：conda list

当前环境下安装包：conda install xxx （卸载包用 conda uninstall xxx，另外还可用conda中的pip安装卸载包 ）

查看conda包含的下载源（用于加速下载）：conda config --show-source

添加源：conda config --add channels xxx（常见源如下）

清华镜像conda源：<https://mirrors.tuna.tsinghua.edu.cn/anaconda/pkgs/free/>

<https://mirrors.tuna.tsinghua.edu.cn/anaconda/pkgs/main/>

https://mirrors.tuna.tsinghua.edu.cn/anaconda/cloud/pytorch/

中科大镜像conda源：<https://mirrors.ustc.edu.cn/anaconda/pkgs/main/>

<https://mirrors.ustc.edu.cn/anaconda/pkgs/free/>

<https://mirrors.ustc.edu.cn/anaconda/cloud/conda-forge/>

<https://mirrors.ustc.edu.cn/anaconda/cloud/bioconda/>

豆瓣conda源：<http://pypi.douban.com/simple/>

阿里云conda源：<https://mirrors.aliyun.com/pypi/simple/>

清华镜像pip源：https://pypi.tuna.tsinghua.edu.cn/simple/

阿里云镜像pip源：<http://pypi.douban.com/simple>

中科大pip源：<https://pypi.mirrors.ustc.edu.cn/simple>

豆瓣pip源：http://pypi.douban.com/simple

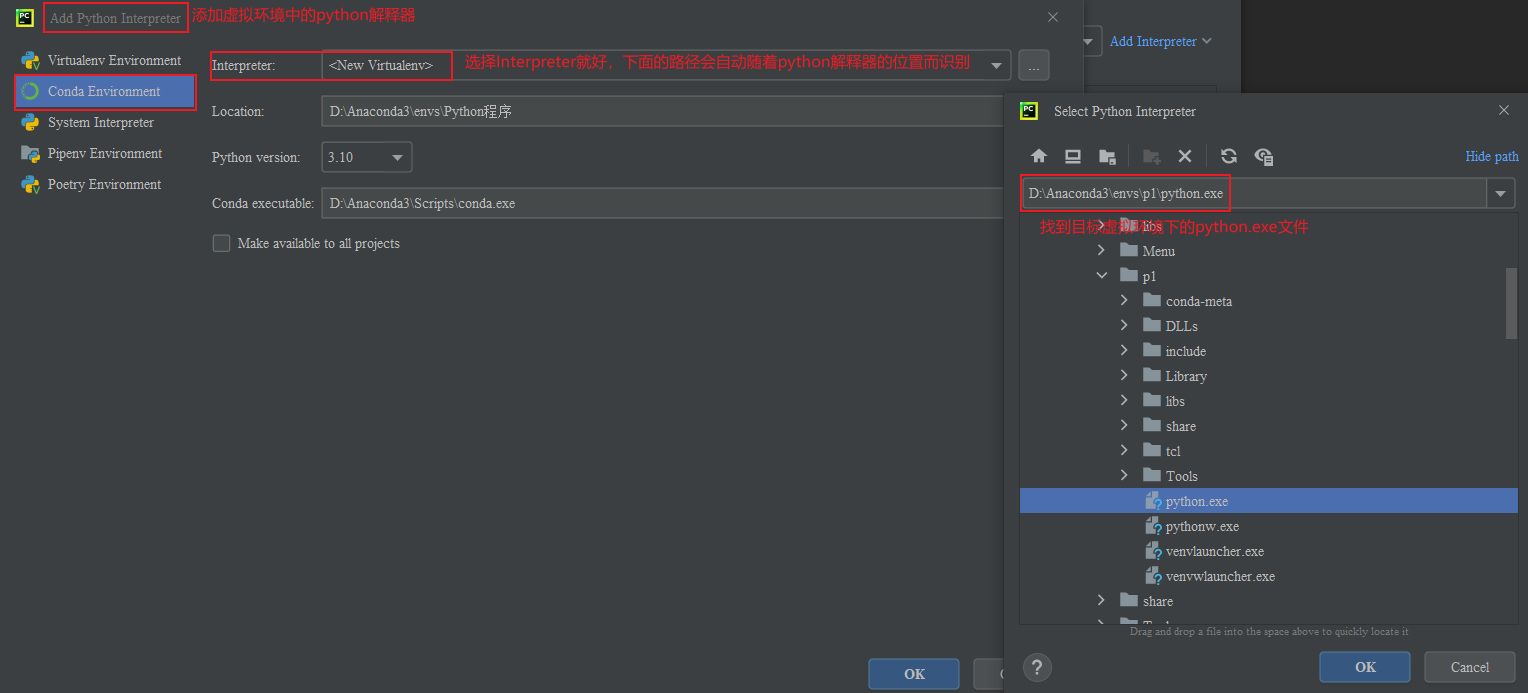
删除源：conda config --remove channels xxx

实际中，通常这样运用： conda/pip install xxx <https://……（源链接）>

升级conda/anoconda/python/package：conda update conda/anaconda/python/xxx （升级anaconda前需升级conda）

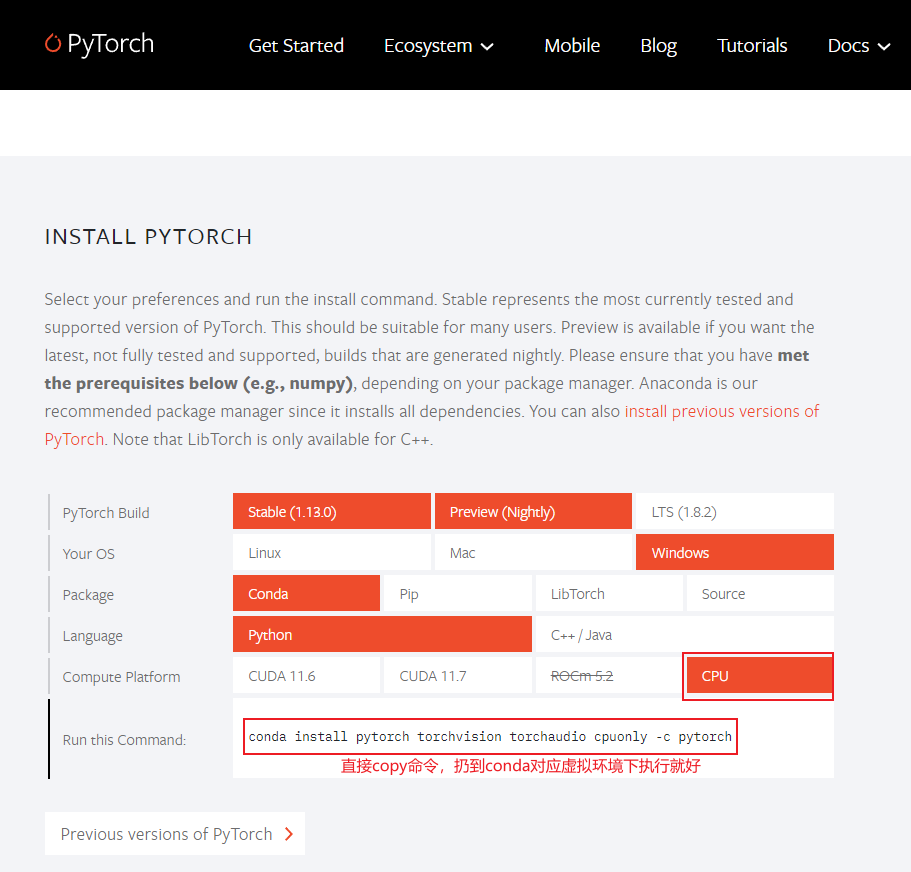
或者 pip install –upgrade/package

最后，在conda命令行创建好虚拟环境后，我们还需要将其导入pycharm中，如下：



1. Pytorch、cuda、对应版本安装问题

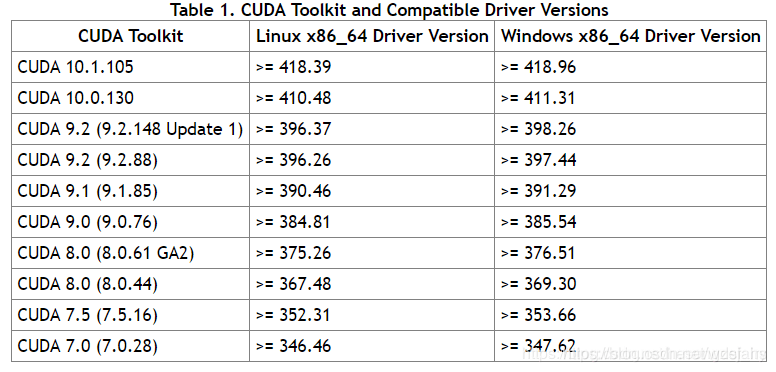
首先我们需要明白，Pytorch虽然被称作深度学习框架，但本质上和numpy、pandas这些包并没有什么不同，因此理论上我们可以直接用conda里面的conda install 或pip install 命令下载，实际上若只想在cpu环境下跑深度学习，我们就可以这样做：



Pytorch官网截图

当然，主流更多还是基于GPU来跑深度学习，毕竟谁也不希望一个模型跑数个礼拜或数月……，但想用GPU跑模型也不容易，须遵循如下规则：

1. 电脑是否有独立显卡：任务管理器中查看是否有 GPU 1这一栏，若有则表示电脑主机内嵌独立显卡，GPU位于显卡内
2. 查看显卡驱动版本：显卡是物理硬件，显卡驱动为系统软件，用于控制显卡运作，查看方式：NVIDIA控制面板 -> 系统信息 -> 驱动程序版本
3. 找到与驱动相对应的CUDA版本：CUDA为NVIDIA推出的并行计算框架……，总之必须得有这个，CUDA版本与驱动版本的对应关系如下：



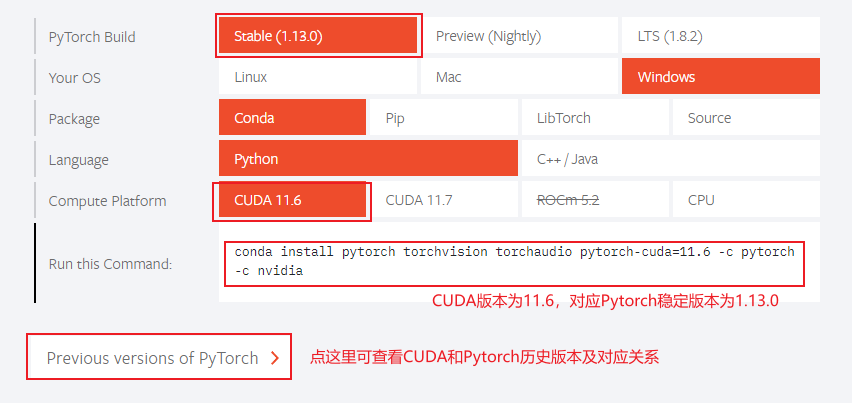
早期CUDA与驱动版本对应关系，最新对应关系可自行查找

需满足驱动版本大于CUDA版本，若嫌弃电脑驱动版本过低，可尝试用NVIDIA的GeForce Experience升级

1. 下载CUDA和Cudnn并配置：根据所需版本下载对应的CUDA，再基于CUDA版本下载对应的Cudnn，具体教程可参考网上其他资料

如：https://zhuanlan.zhihu.com/p/416712347

1. 最后，基于不同版本CUDA下载对应版本的Pytorch，如下：



同样复制粘贴命令到conda命令行执行，若下载太慢可在命令后添加源链接。注意此时需切换至目标虚拟环境

1. 至此，基于GPU的Pytorch框架就搭建起来了，我们在Pycharm里验证一下是否可行，命令如下：

import torch # 如果pytorch安装成功即可导入

print(torch.cuda.is\_available()) # 查看CUDA是否可用，可用返回True

print(torch.version.cuda) # 查看CUDA的版本号

# Vscode + Remote-SSH + Docker 远程深度学习环境

# Hadoop大数据生态