学习中遇到的困难以及希望杨老师解答的问题：

1. 如何从内存角度理解函数调用中的参数传递，区分形式与实质参数？（图1、2）2022-6-27讲过  
   形参【函数定义处的参数是形参】： arg1 ,arg2

实参【函数调用处的参数是实参】： fun(n1,n2) n1,n2为实参

n1的数据类型是 int 🡪不可变对象

n2是列表 list 🡪可变对象

arg1=n1 ，所以arg1为int类型🡪不可变对象

arg2=n2 ,所以arg2为list类型🡪可变对象

开执行函数体：

arg1=100 （函数中的局部变量,只在函数内部起作用） 🡪 n1依然是11，

arg2.append(10) 🡪增加了n2中元素的个数，并没有为list重新开辟空间

函数执行完毕

变量就剩 n1与n2 , n1的值是11 ， n2=[22,33,44,100]

1. 有的代码打的与老师一样，却有问题，重新打一遍就好了，有疑惑  
   计算机是不会骗人的，肯定有细微的差别，只是你没发现这个细微的差别而已

(中英的符号，缩进问题 ，忘记写冒号的问题……)

1. 想老师讲一下这串代码（图3）  
   enumerate()是python的一个内置函数,函数作用是获取个可迭代对象的每个元素的序号值及该元素值
2. 爬虫如何获取某url对应的headers。首先我有疑问，就是不同网页的headers相不相同。老师,我不知道我想的对不对,一般情况(图片上那样)他们的headers我都是写同一个，如果每个url对应的headers不同，这种要怎么获取headers呢,我只会复制。（图4、5、6）

Headers 是表示请求头里的内容，基本上请求头里的内容都相同，除了cookie,refer等，但是一次会话的cookie是一样的，不会根据网页而变化,refer会跟据网页而变

1. object这种数据类型的转化，object数据讲解，怎么处理第二张的数据（图7、8）  
   需要提供代码进行调试
2. 为什么我这个none对象的布尔值为true（图9）  
   你数错了，是False
3. 在数据分析课程中（杨老师的），有几个疑问：Q1. pandas的merge和concat方法的区别   应用场合

Merge与concat的作用都是用于合并DataFrame数据用的  
concat 只是 pandas 下的方法，而 merge 即是 pandas 下的方法，又是DataFrame 下的方法

concat 可以横向、纵向拼接，又起到关联的作用

merge 只能进行关联，也就是纵向拼接

concat 可以同时处理多个数据框DataFrame，而 merge 只能同时处理 2 个数据框

concat的源码

|  |
| --- |
| **def** concat(  objs: Iterable[NDFrame] | Mapping[Hashable, NDFrame],  axis=0,  join=**"outer"**,  ignore\_index: bool = **False**,  keys=**None**,  levels=**None**,  names=**None**,  verify\_integrity: bool = **False**,  sort: bool = **False**,  copy: bool = **True**, |

objs：合并的数据集，一般用列表传入，例如：[df1,df2,df3]

axis：指定数据拼接时的轴，0是行，在行方向上拼接；1是列，在列方向上拼接

join：拼接的方式有 inner，或者outer，与sql中的意思一样

pd.merge()源码

|  |
| --- |
| **def** merge(  left: DataFrame | Series,  right: DataFrame | Series,  how: str = **"inner"**,  on: IndexLabel | **None** = **None**,  left\_on: IndexLabel | **None** = **None**,  right\_on: IndexLabel | **None** = **None**,  left\_index: bool = **False**,  right\_index: bool = **False**,  sort: bool = **False**,  suffixes: Suffixes = (**"\_x"**, **"\_y"**),  copy: bool = **True**,  indicator: bool = **False**,  validate: str | **None** = **None**, |

left：左表

right：右表

how：关联的方式，{'left', 'right', 'outer', 'inner', 'cross'}, 默认关联方式为 'inner'

on：关联时指定的字段，两个表共有的

left\_on：关联时用到左表中的字段，在两个表不共有关联字段时使用

right\_on：关联时用到右表中的字段，在两个表不共有关联字段时使用

 Q 2. 时间序列的pd.resample方法没太懂，能否举几个例子

Pandas中的resample，重新采样，是对原样本重新处理的一个方法，是一个对常规时间序列数据重新采样和频率转换的便捷的方法

|  |
| --- |
| **def** resample(  self,  rule,  axis=0,  closed: str | **None** = **None**,  label: str | **None** = **None**,  convention: str = **"start"**,  kind: str | **None** = **None**,  loffset=**None**,  base: int | **None** = **None**,  on=**None**,  level=**None**,  origin: str | TimestampConvertibleTypes = **"start\_day"**,  offset: TimedeltaConvertibleTypes | **None** = **None**, |

1. 想老师讲一下相对定位，绝对定位

老师使用的软件叫HBuilder  
CSS的内容

1. \_\_new\_\_与\_\_init\_\_传递过程不清楚。【比较难理解，如果初学者不太理解，不影响后续学习】

\_\_new\_\_ 与\_\_init\_\_ 都是object类中的方法

\_\_new\_\_ :用于创建对象 （实际上就是在内存中开辟内存空间，开辟内存空间的目的就是 存储对象）,该 方法是由程序运行时，系统自动调用，什么时候调用 ，创建对象（实例）时，不建议手动调用， 因为手动调用无意义

\_\_init\_\_ : 创建完对象之后 ，系统自动调用，调用的时间点是\_\_new\_\_之后，作用是为这个对象的属性进行赋值 ，不需要手动调用

执行顺序 : 先执行 \_\_new\_\_ ，再执行\_\_init\_\_

\_\_new\_\_ :生孩子 （一个新生儿诞生了）

\_\_init\_\_ :给孩子起名字，上户口，写家庭地址，身份证号…

在类中定义的方法，第一个参数都是self, 这个self第一次使用在\_\_init\_\_中, 是谁给\_\_init\_\_中的self进行赋值的？是\_\_new\_\_方法，因为该 方法有一个返回值return obj , self的值又传递给了创建的对象 p , p=Person(‘marry’,20)

\_\_new\_\_-->return obj-🡪\_\_init\_\_-🡪self-🡪 创建的对象(p)

1. Python定义类的时候，赋值了参数，比如self.name，这个是不是在后面调用的时候，就相当于是一个变量，这个具体怎么使用？跟一般的定义一个变量

如name=xxx感觉有点不太一样，但是又说不出来，希望老师详细说一下。

如果使用name=xxx，那么这个变量只能在定义的方法里使用，不能在整个类里使用,使用self.name=xxx，那么这个变量可以在这个类的任意方法中都使用

1. python中类的浅拷贝，浅拷贝，深拷贝（图10）  
   浅拷贝：外(id)变，里(id)不变

深拷贝：外(id)变，里(id)也变

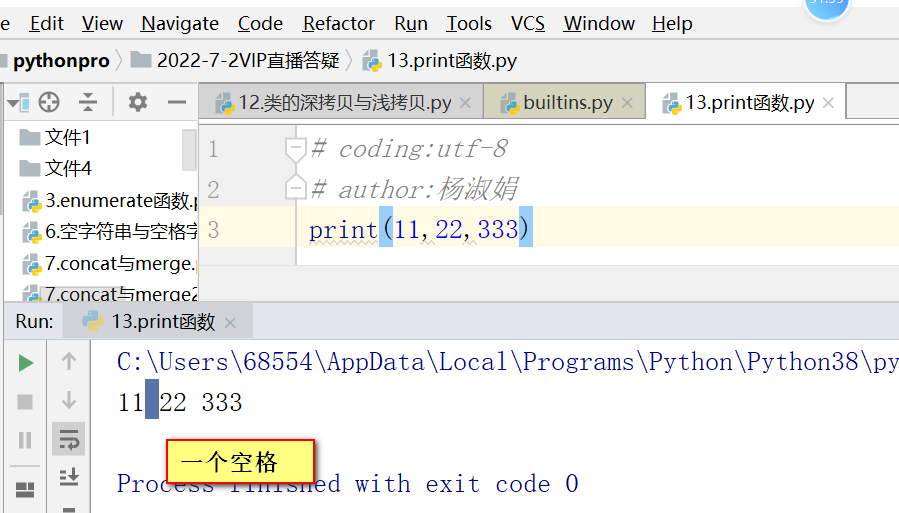
1. python的输入函数print详细用法  
    按ctrl 点击print函数，进行函数的定义处

**def** print(self, \*args, sep=**' '**, end=**'\n'**, file=**None**):

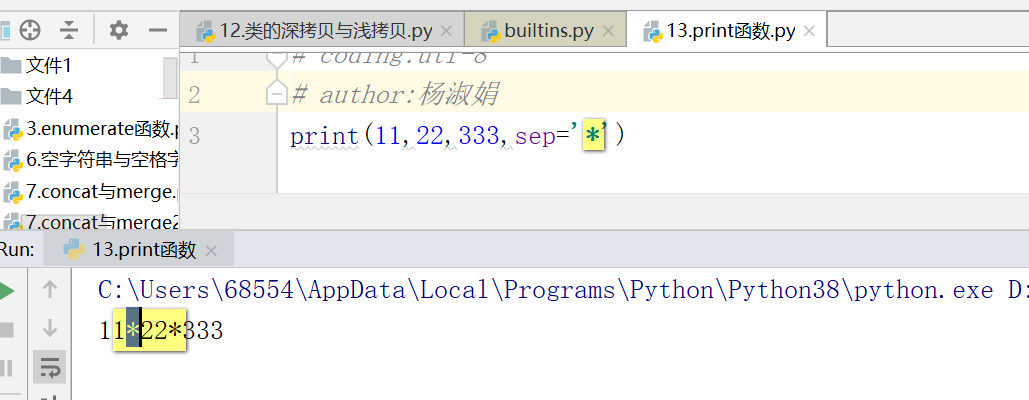
self 🡪不用管

\*args :这是一个个数可变的位置参数

sep : 是一个默认值参数，值为一个空格，所以，使用 逗号分隔传递的实参时，输出结果带一个空格



可不可以修改 sep的值？可以



end: 默认值参数，默认的结束符是 ‘\n’, \n是换行的意思，所以单独写一个print()一个参数都没有的情况下，作用就是换行，可不可以修改end的值？可以，最常用的使用的是 print(end=’\t’)

1. 下载的桌面版，这个不知道要如何设置，只能在虚拟机里打开，如何能弄到外面，老师的看没在虚拟机打开就可以进入无版图的应用（图11、12）  
   这个私下微信下肖老师，杨老师的Linux蓝屏
2. 作业：给定一个4维矩阵，如何得到最后两维的和？（提示，指定axis进行计算）， 但是路老师的视频中，只有一个结果，我想知道计算过程  
   需要提示代码（视频链接）
3. 作业的问题（图13、14）（文件1）  
   图13 问题 所在：缩进问题 ，代码的层次结构
4. 哈希表不了解，课堂上面草草带过去 这里不明白  
   哈希原理 <https://zhuanlan.zhihu.com/p/439157543>

哈希函数有一个 叫“除留取余法” y=k(x) 🡪 假设 y=x%N的余数

[10,23,25,45] 假设有4个整数，使用哈希表存储 🡪 开的数组为

0 1 2 3

23(k,v)

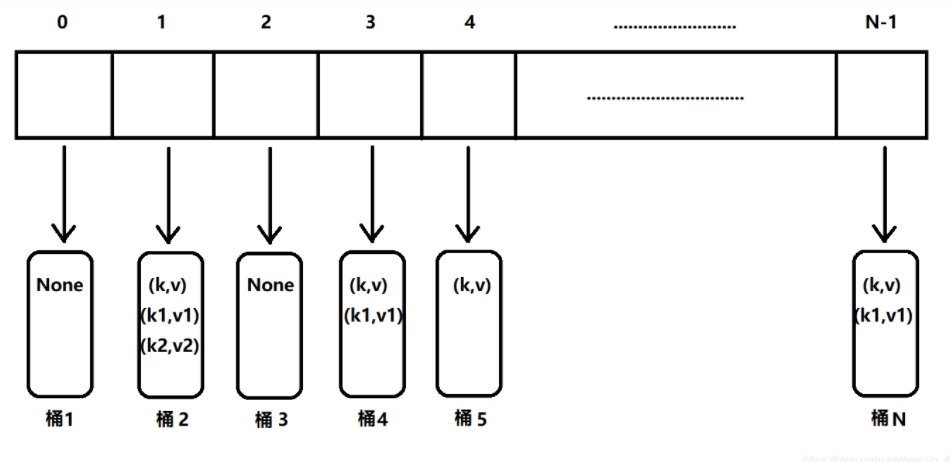
10(k,v)

25(k,v)

10

10

45(k,v)



1. 循环语句的使用  
   具体题目 (不清楚你是语法没记住，还是代码报错)
2. Q1:希望杨老师讲一下爬虫中动态数据和异步加载的数据怎样抓包？  
   提示具体网址

Q2:用网上的免费IP做成一个python文件，然后在自己的爬虫文件导入这个文件，实现ip代理爬取，但是IP的有效期有限，希望杨老师帮我讲一下怎样随机生成一个有效的IP来爬？

IP无法自动生成，需购买，比特代理  
https://www.feiyuip.com/help

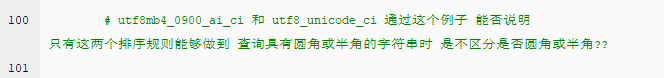
Q3:就是以后的爬虫文件，只要调用这个IP接口文件就可以用代理IP爬取文件，但是就是不知道IP有没有效，所以请杨老师帮我讲一下怎样判断!（图15)(文件2）

1. 想要达到的目的：  
   一、抓取指定文件夹中文件的名字（已实现）  
   二、把抓取到的文件名字按类别区分放入excel（已实现）  
   三、把抓取到的数据进行线性回归（部分实现）  
         会抱一个不影响运行的错误：warnings.warn("kurtosistest only valid for n>=20 ... continuing "  
   四、把线性回归图片和线性方程（y=ax+b，a和b值从线性回归结果中抓取），保存到之前保存的excel的sheet2里面（还未实现）  
   麻烦老师帮忙看看  
   1）目前的代码是否有冗余的地方  
   2）第三个目的报错的原因是啥  
   3）第四个目的还未实现的功能如何完善（文件3）  
   （这个需要问下路老师）

20. matplotlib画图时如何设置中文标注的乱码问题

提供代码好测试 （截图，让我看下效果）

|  |
| --- |
| import matplotlib.pyplot as plt  plt.rcParams["font.sans-serif"]=["SimHei"] #设置字体  plt.rcParams["axes.unicode\_minus"]=False #该语句解决图像中的“-”负号 |

1. 问题在这里，里面的执行包括了创建库表（文件4）  
   

https://blog.csdn.net/myqq1418/article/details/118354670