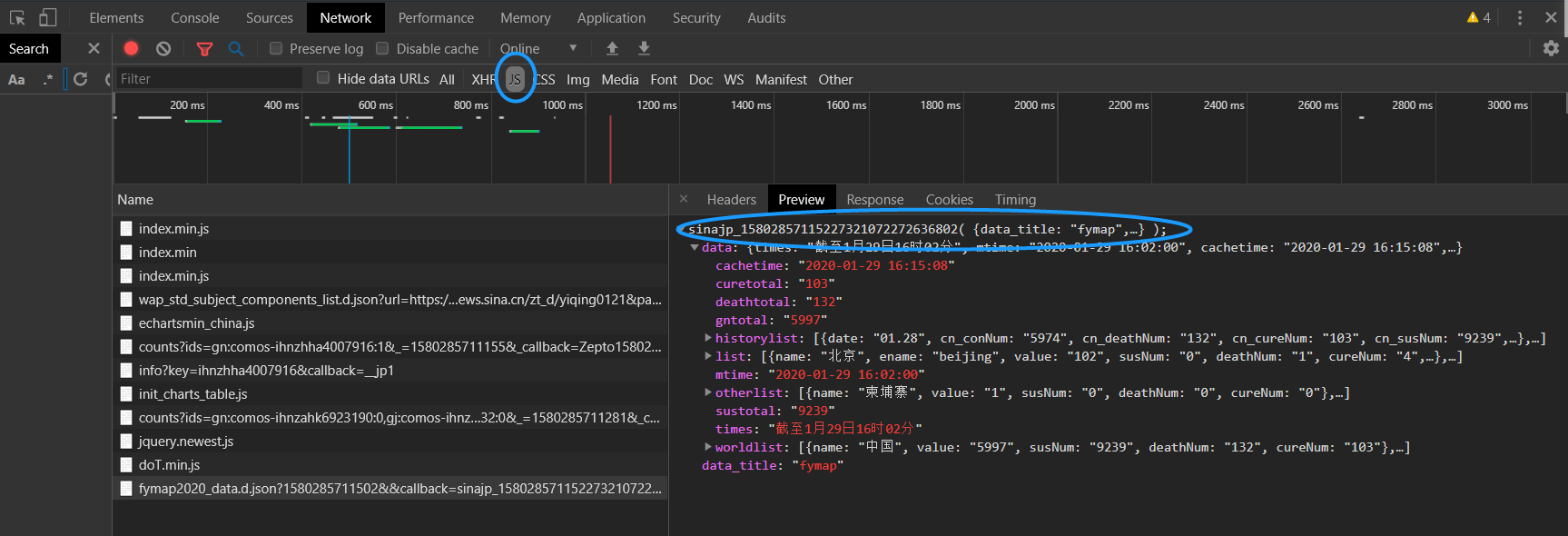
一．文件寻找的不同寻常

1. 一般情况下储存信息的文件存放在XHR文件夹之中，而这一次居然在JS文件夹中发现储存信息的文件的
2. 与XHR文件夹中的json数据不同，虽然JS文件与dic很相似，不过这个字典是包装在一个类似元组的东西里面的sinajp\_1508239812412598(dict)



解决方法：

提取字典成分，得到数据类

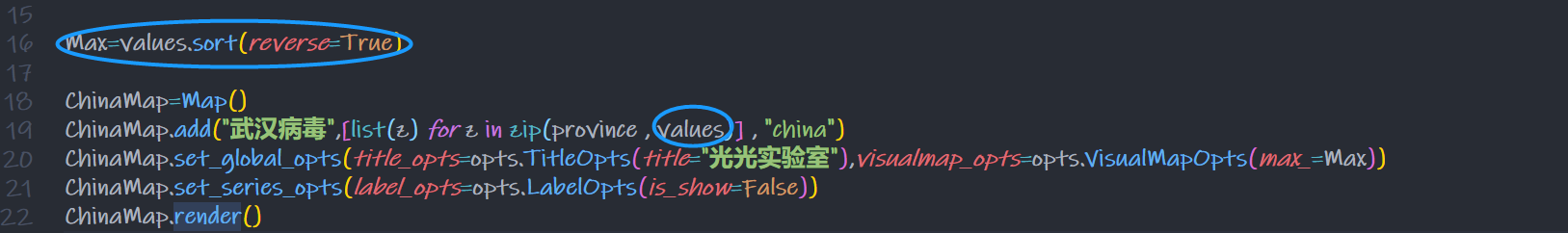
**\*json模块的作用：**

**1.将json数据转化为dict/list**

**2.将json数据中的unicode编码转化为utf-8编码**

**二．**List的sort（）方法，特别要注意List的会改变原来的List，所以最好使用sorted(),而不是List；以免出现意外

在培养病毒的过程中，我试图找出病毒发育的最凶猛的数，可是却无意间改动了list本身，造成bug



三．测试了一下生成器yield的用法，发现Python还是挺聪明的

Def generator0(times):

a=times

while(times):

times-=1

yield a

print(list(generator0(3))

>>[3,3,3]

Def generator1(times):

While(times):

Times-=1

Yield times

Print(list(generator1(3))

>>[2,1,0]

通过上面我们知道了，在Python Genertor中

唯一的可变变量来自于形式参数

在Python Generator中定义的局部变量只定义一次

Def generator2(times):

A=3

While(times):

Times-=1

a-=1

yield a

print(list(generator2(3))

>>[2,1,0]

四．Python time模块使用的注意事项

Python time 中返回的时间戳是以秒为单位的时间戳，而不是我们在Java New date函数中看到的毫秒级时间戳

1s=1000ms!

五．并不是很万能的 列表推导式[x for x in list ]

实际上我们说列表推导式非常的快，但实际上你要注意 其中的x 表示返回值！

A= [[2,3],[3]]

[item.append(3) for item in A]

>>[None , None , None]

六．注意数学运算中 数据类型 的隐式转换

4-2.//2=3.0

4-2//2=3

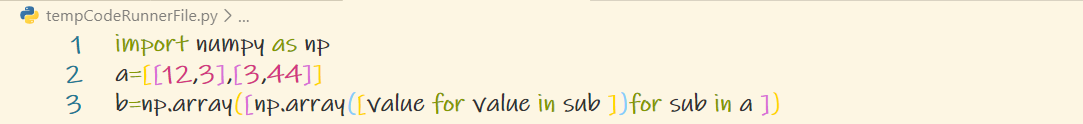
七．神奇的文字类型：

在爬虫中，要注意到，不是爬下来的所有的变量，Python中都有对应的数据类型的

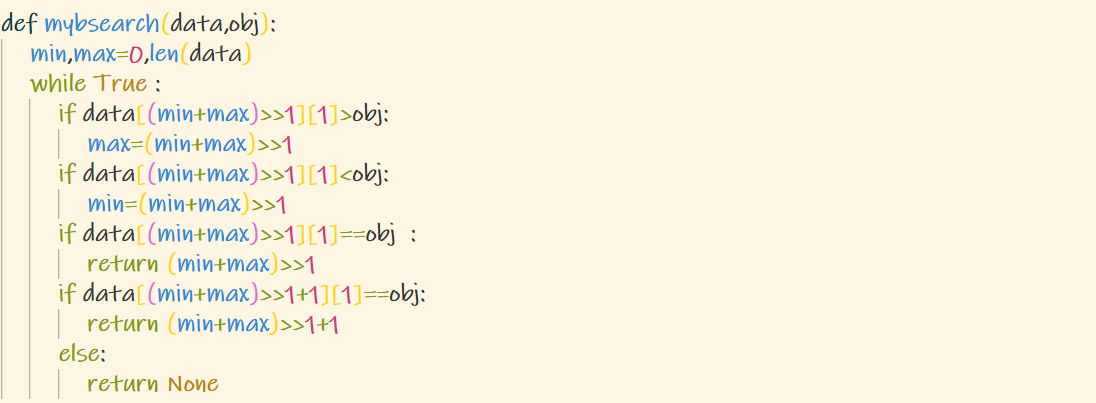
同样是‘2’，可能在网络中就不一定是Python内置的str变量！

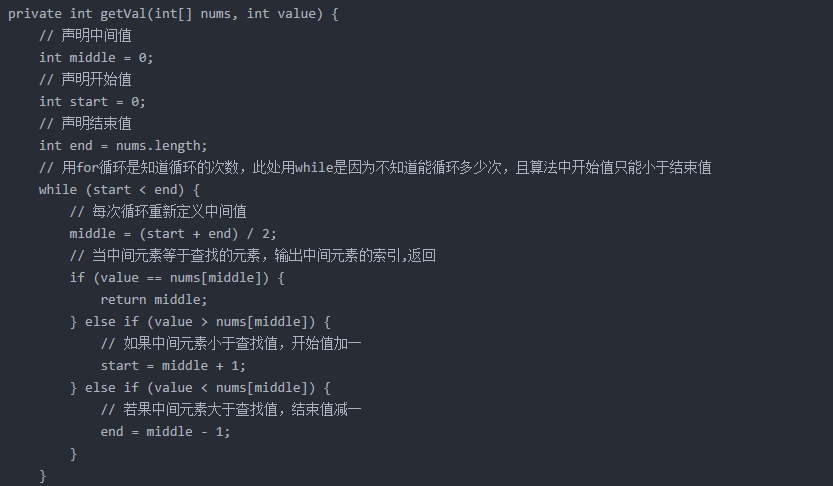
八．

对于复合列表而言，想要把它转化为完完全全的array,必须把每一个子列表与唯一的父列表都转化为array 才可以



九．二分查找





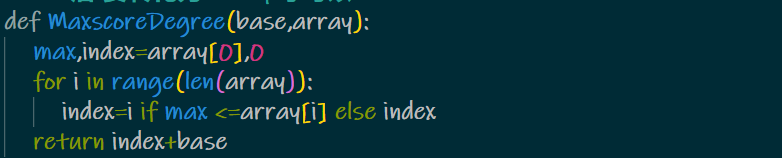
在二分查找的终止判断条件中，要注意到两类情况，

第一类：（min-max）近邻，这时候min，max有一者是obj/都不是

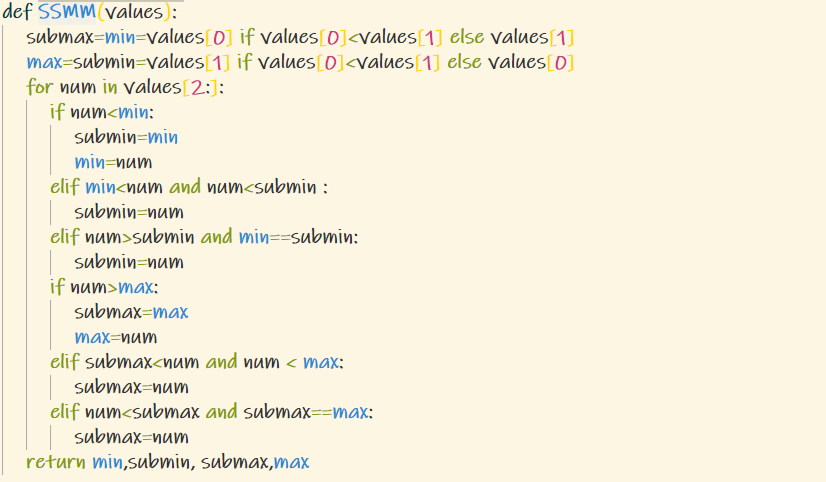
第二类：（min-max）隔着一个数，这时候要么 中间的数是obj，要么不是obj

\*\*如果查找的是浮点数的话，那么 abs( obj – num )<1e-5

十．找出第二大，找出第二小



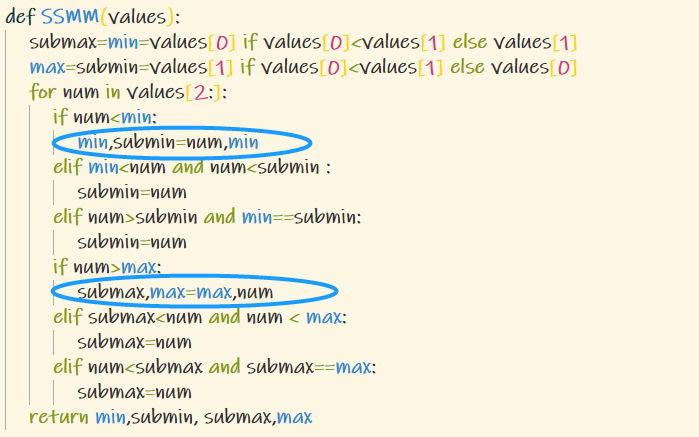
上面这个代码错在，没有改变最大值啊啊啊啊啊！！！



核心：在num<submax的时候，判断submax==max？，如果True，那么抓住机会拉开关系！

当然，前两行代码同样重要！

十一.动态语言的静态问题

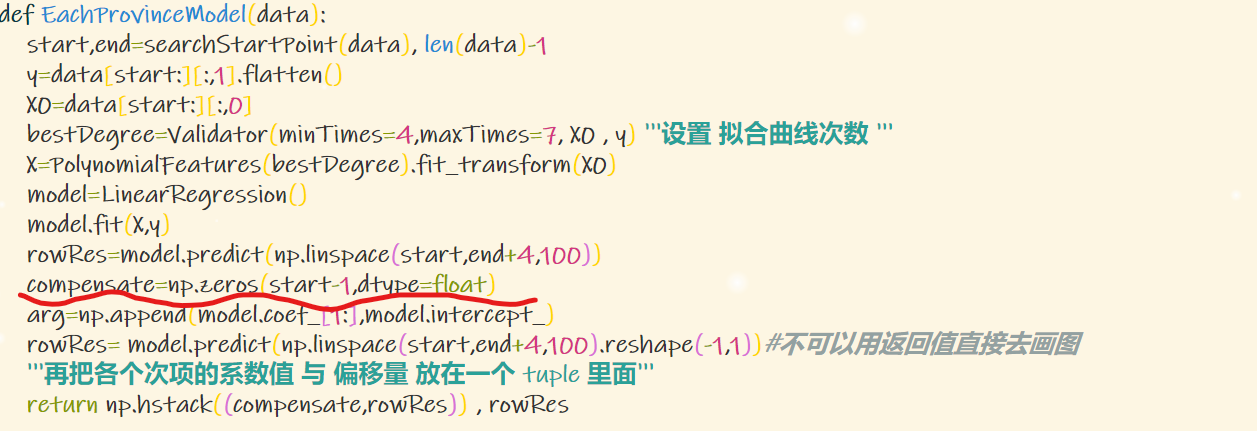


min , submin = num , min

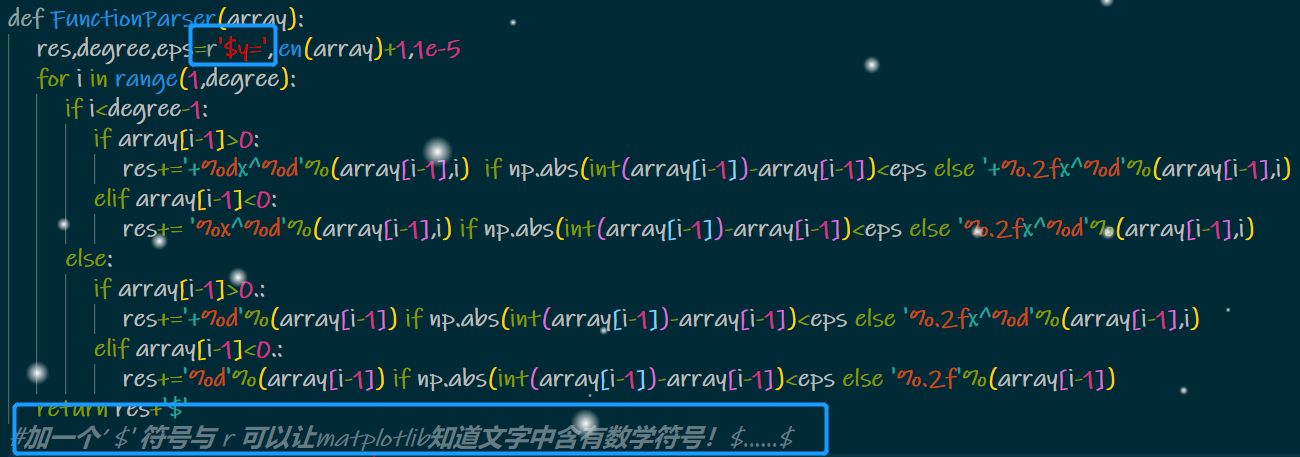
>>这时候，num，min在赋新值的时候 不会随变而变

十二.对于交叉检验 cross\_val\_score(cv),cv值越大，越能接近全数据拟合的训练结果

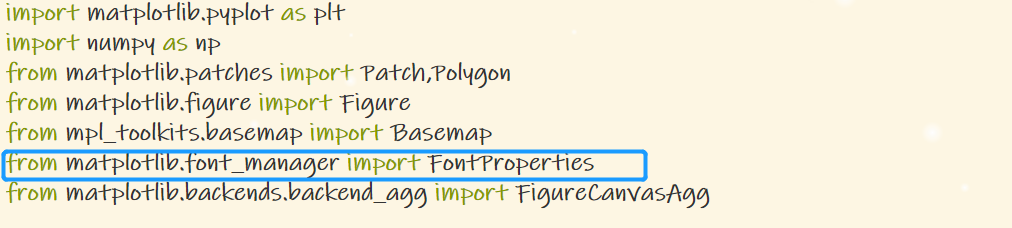
十三.对于有的参数项，如果有再另外出现加减，那么我们常常要进行特判

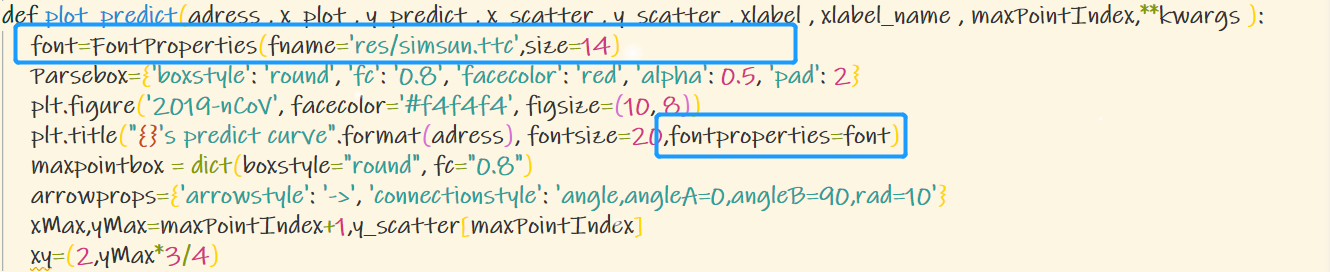


十四.在画图的时候，如何让 matplotlib知道你用了数学公式呢？



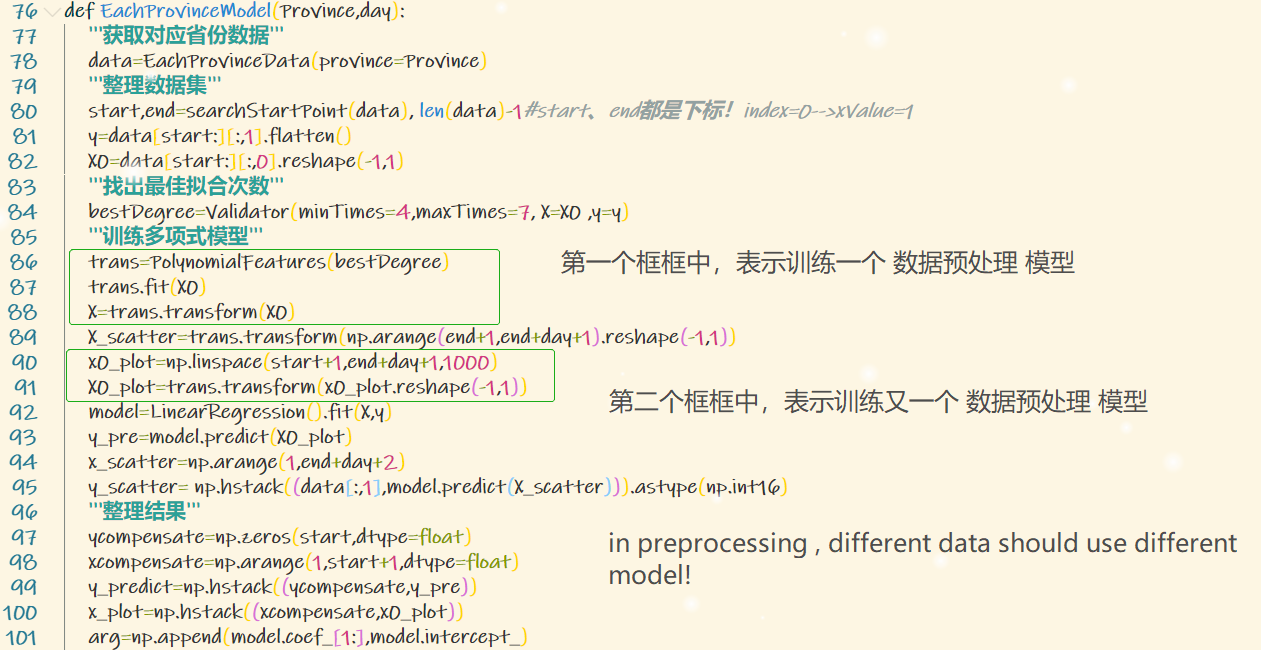
十五. 在画图的时候，如何让matplotlib显实中文呢？





实例化 FontProperties 对象！

十六.sklearn.preprocessing

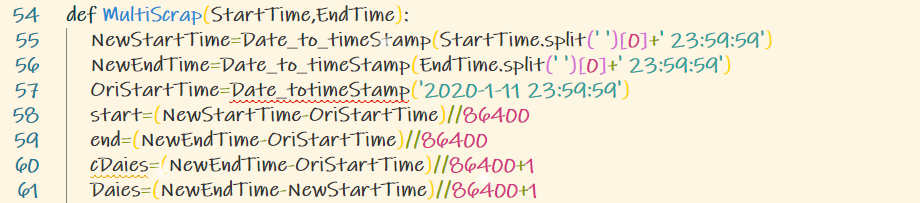


在数据预处理中，不同的数据，不可以使用同一个实例化对象！

十七.一些有用的结论

1.arrange(a,b),一共插入了 b-a 个数字

2.z注意一下start ，end 他们的计算方式！可以就出index



十七.在拟合数据的时候，一开始以为差不多做完了，结果检验了一下r2值；默默再看了一遍统计学……首先我们要有一种意识，就是这个数据集有很明显的趋势特征；可是我们关注的是数据的周期特征，我们需要去趋势化。其次就是r2\_score为负，那么说明出现了高方差。综合以上两点，我们要通过 对数化 来优化这个模型！

