python环境搭建

在python的官网下载完之后会有一个IDLE和一个launcher

IDLE是一个编码运行器，还可以自行新建一个编码器

编码器里面对于格式和显示有自动的识别和缩进功能（类似于vba）

而在使用编码器进行数据分析之前

要现在电脑的系统终端搭建python环境

这里使用python3.6

参考 https://www.cnblogs.com/i-am-lvjiazhen/p/6264354.html

总结一下：

1.打开终端窗口，输入以下命令

/usr/bin/ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"

即可成功安装HomeBrew

2.在终端输入以下命令：

brew install python3.6

此时python3已经安装在应用程序里面，此时会看到IDLE和launcher

brew linkapps python3.6

可以查找他们的位置

对于数据分析需要一下几个包

参考：https://www.cnblogs.com/mandy-study/p/7780213.html

1.Numpy：

　　Numpy是python科学计算的基础包，它提供以下功能（不限于此）：

　　　　(1)快速高效的多维数组对象naarray

　　　　(2)用于对数组执行元素级计算以及直接对数组执行数学运算的函数

　　　　(3)用于读写硬盘上基于数组的数据集的工具

　　　　(4)线性代数运算、傅里叶变换，以及随机数生成

　　　　(5)用于将C、C++、Fortran代码集成到python的工具

2.pandas

　　pandas提供了使我们能够快速便捷地处理结构化数据的大量数据结构和函数。pandas兼具Numpy高性能的数组计算功能以及电子表格和关系型数据（如SQL）灵活的数据处理能力。它提供了复杂精细的索引功能，以便更为便捷地完成重塑、切片和切块、聚合以及选取数据子集等操作。

　　对于金融行业的用户，pandas提供了大量适用于金融数据的高性能时间序列功能和工具。

　　DataFrame是pandas的一个对象，它是一个面向列的二维表结构，且含有行标和列标。

　　ps.引用一段网上的话说明DataFrame的强大之处：

　　Excel 2007及其以后的版本的最大行数是1048576，最大列数是16384，超过这个规模的数据Excel就会弹出个框框“此文本包含多行文本，无法放置在一个工作表中”。Pandas处理上千万的数据是易如反掌的事情，同时随后我们也将看到它比SQL有更强的表达能力，可以做很多复杂的操作，要写的code也更少。 说了一大堆它的好处，要实际感触还得动手码代码。

3.matplotlib

　　matplotlib是最流行的用于绘制数据图表的python库。

4.Scipy

　　Scipy是一组专门解决科学计算中各种标准问题域的包的集合。

5.statsmodels： https://github.com/statsmodels/statsmodels

6.scikit-learn： http://scikit-learn.org/stable/

首先要在终端里面安装pip

参考：https://www.runoob.com/w3cnote/python-pip-install-usage.html

curl https://bootstrap.pypa.io/get-pip.py -o get-pip.py

sudo python3.6 get-pip.py

接下来就可以安装各种包了

pip install SomePackage （package名称）

这里是

pip install pandas

pip install numpy

pip instal matplotlib

sudo pip install ipython

然后就可以打开IDLE进行数据分析啦

用到包的时候使用import sompackage

字符串 String

hello world

print("hello world!")

字符串的生成，单引号与双引号是等价的：

s = "hello world"

s

s="""hello

...: world"""

print(s)

s = '''hello

world'''

print s

s = "hello" + " world"

s

%matplotlib inline

from matplotlib.pyplot import plot

need to first:

%pip install ipykernel

Python提供了一个很像MATLAB的绘图接口。

%matplotlib

from matplotlib.pyplot import plot

plot(a, a\*\*2)

Python用关键词def来定义函数。

用Numpy数组做参数x

可以在定义时指定参数的默认值：

Python中使用import关键词来导入模块。

import os

pid=os.fork() #fork反复拷贝

if pid==0:

print("A",os.getpid(),os.getppid())

else:

print("B",os.getpid(),os.getppid())

# os.getpid()获取当前进程id os.getppid()获取父进程id

用class来定义一个类。 Person(object)表示继承自object类； \_\_init\_\_函数用来初始化对象； self表示对象自身，类似于C Java里面this。

python 3.3.2报错：No module named 'urllib2' 解决方法

2017年09月06日 14:50:15 hacker\_LeeDa 阅读数 21011 标签： python 更多

个人分类： Python

python代码：

import urllib2

response = urllib2.urlopen('http://www.baidu.com/')

html = response.read()

print html

报错如下：

Traceback (most recent call last):

File "<pyshell#0>", line 1, in <module>

import urllib2

ImportError: No module named 'urllib2'

解决方法：

在python3.3里面，用urllib.request代替urllib2，另外python3之后，不能再用，print html

注意：print 的东西要用（）括起来。

这样的方式，因为print这个时候已经是一个方法了。必须使用下面的方法

可以将代码换成：

import urllib.request

resp=urllib.request.urlopen('http://www.baidu.com')

html=resp.read()

print(html)