IPython：一种交互式计算和开发环境

1、Tab键自动完成、内省（？，？？），%run命令

2、脚本：在一个空的命名空间中运行，没有import，没有定义任何其他变量，与标准命令环境中执行时一样。

3、执行粘贴板中的代码：%paste,%cpaste

4、键盘快捷键：C-u清除当前行的所有文本，C-L清屏

5、魔术命令：P58

Numpy基础：数组和矢量计算

1、ndarray：多维数组对象

2、shape:表示各维度大小的元祖；dtype：说明数组数据类型的对象

3、np.array()：创建一维或多维数组

np.zeros()：创建全0数组

np.ones()：创建全1数组

np.empty()：创建没有任何具体数值的数组

np.arange()：内置函数range的数组版

更多数组创建函数请查看P85

4、ndarray的数据类型：浮点数、复数、整数、布尔值、字符串、普通的python对象

更多数据类型请查看P86

5、转换数据类型：.astype(要转换的类型)

6、数组与标量的算术运算会将标量传播到各个元素

7、大小不同的数组之间的运算叫广播，大小相等的数组之间的任何算术运算都会将运算应用到元素级。

8、与列表的区别在于，数组切片是原始数据的视图，意味着数据不会被复制，视图上的修改都会直接反映到原数组上。

9、.copy()：显示地进行复制操作

10、切片索引，布尔型索引，花式索引

11、通过布尔型索引选取数组中的数据，总是创建数据的副本。

12、花式索引：利用整数数组进行索引

13、np.ix_：将两个一维数组转化为一个用于选取方形区域的索引器，总是将数据复制到新数组中。

14、数组的转置：.T，.transpose()，.swapaxes()

15、计算矩阵内积，即矩阵乘矩阵转置：np.dot()

16、通用函数：

np.sqrt()，np.exp()，np.add()，np.maximum(),

np.modf()用于浮点数数组的小数和整数部分

更多通用函数请查看P99~100

17、矢量化：用数组表达代替循环

18、将条件逻辑表示为数组运算：np.where()

19、std：标准差；var：方差；argmin：最小元素索引；cumsum：所有元素累计和；cumprod：所有元素累计积

更多基本数组统计方法请参考P104~105

20、.sort()：排序

21、np.unique()：找出数组中唯一值并返回已排序结果。

22、np.in1de()：测试一个数组中的值在另一个数组中的成员资格

更多数组的合集运算请参考P107

23、CSV：逗号分隔文件

24、常用的numpy.linalg函数请参考P110

25、np.seed()：确定随机数生成器的种子

numpy.random函数请参考P111

Panda入门

1、Series:由一组数据以及一组与之相关的数据标签组成

2、DataFrame:一个表格型数据结构，含有一组有序的列，每列可以是不同的值类型，既有行索引也有列索引。

3、.reindex()函数的参数参考P129

4、DataFrame的索引选项参考P132~133

5、对不同索引的对象进行算术运算，结果索引是该索引的并集；对不同DataFrame进行算术运算，索引和列为原来那两个DataFrame的并集。

6、add：加法；sub：减法；div：除法；mul：乘法

7、排序：.sort_index(axis=0/1,ascending=False/True,by=列名)

按值排序：.order()

8、排名：.rank() ps:我也搞不懂

9、DataFrame、Series描述和汇总统计相关方法请参考P144

10、NaN（not a number）：表示浮点和非浮点数组中的缺失数据（missing data），只是一个便于被检测出来的标记。

11、NA处理方法：dropna，fillna,，snull，notnull。

12、.dropna()：默认丢弃任何含有缺失值的行。传入how='all',丢弃全为NA的行。传入axis=1，how='all',丢弃全为NA的列。

13、只留下一部分观测数据：.dropna(thresh=3)

14、填充函数：.fillna()

参数请参考P152~153

15、层次化索引

16、.set_index()：将一个或多个列转换为行索引。

17、.reset_index()：层次化索引的级别会被转移到列里面

数据加载、储存与文件格式

1、pd.read_csv()，pd.read_table()，其参数请参考P163

更多解析函数请参考P162

2、nrows，chunksize

3、from_csv

4、csv.reader()，参数请参考P172

5、JSON数据：通过HTTP请求在Web浏览器和其他应用程序之间发送数据的标准格式之一。

数据规整化：清理、转换、合并、重塑

1、pd.merge()：根据一个或多个键将不同的DataFrame中的行连接起来

参数请参考P190

2、pd.concat()：沿一条轴将多个对象堆叠到一起

参数请参考P198

3、.join()：实现索引上的合并

4、np.where()：索引全部或部分重叠的两个数据集，Series的combine_first方法实现一样的功能。

5、重塑层次化索引：

stack：将数据的列旋转为行

unstack：将数据的行旋转为列

6、时间序列数据通常是以所谓的长格式或堆叠格式存储在数据库和CSV中的。

7、.pivod()
8、.duplicated()；.drop_duplicates()移除重复数据
9、.map()：利用函数或映射进行数据转换
10、替换值：fillna，replace
11、重命名轴索引：.index.map()；.rename()
12、离散化和面元划分：pd.cut()；pd.qcut()
13、np.random.permutation()：随机重排
14、pd.get_dummies()
15、.split()：拆分逗号分隔符的字符串
.strip()：修剪空白符
16、子串定位：in，index，find
16、返回指定子串的出现次数：.count()
17、python内置的字符串方法请参考P218~219
18、正则表达式请参考P219
19、pd中矢量化的字符串方法

数据聚合与分组运算

1、对时间数据的聚合（groupby的特殊用法之一）也称作重采用
2、Groupby技术

时间序列