第 1 章 Python 程序库入门

基本的数据分析库
- numpy
- pandas
- scipy

可视化
- ggplot
- matplotlib
- seaborn
- Bokeh
- Mayavi
- Nodebox
- igraph
- pandas rplot
- googleVis
- pygal

其他语言
- R：rpy2
- Java：Jython、java2python
- golang：pygo

数据格式
- HDFS：pyHDFS、py4hdfs、tinyHDFS
- Text Files：xreadlines、readLines、pandas read
- SQL：SQLAlchemy、pysqlite3、pyodbc
- NoSQL
  - MongoDB：PyMongo、mongomodels
  - CouchDB：couchdbkit、couchdb-python
  - HBase：PyHBase、pdhbase
  - Cassandra：pycassa、Lazyboy
- JSON：simplejson、anyjson、djson、hjson
- XML：dcxml、simplexml、xmlr

Python数据分析

机器学习
- scikit-learn
- statsmodels
- pylearn2
- NuPIC
- Nilearn
- PyBrain
- Pattern
- fuel
- skdata
- MILK
- nltk

MapReduce
- Hadoop：pydoop、ziggy、mrjob、pyhadoop、dumbo
- Spark：PySpark、spark-sklearn、pyspark-pandas
- Python MapReduce：mapreduce JobX、mrs-mapreduce、pymapreduce、tinymr

高性能
- Libraries：Numba、Blaze
- GPGPU：PyCUDA、NumbaPro
- Parallel：ipyparallel、dispy、dask、pp
- MPI：pyMPI、mpi4py、mpiutils
- Python Systems：Cython、pypy

云
- Amazon：boto、aws、awscli
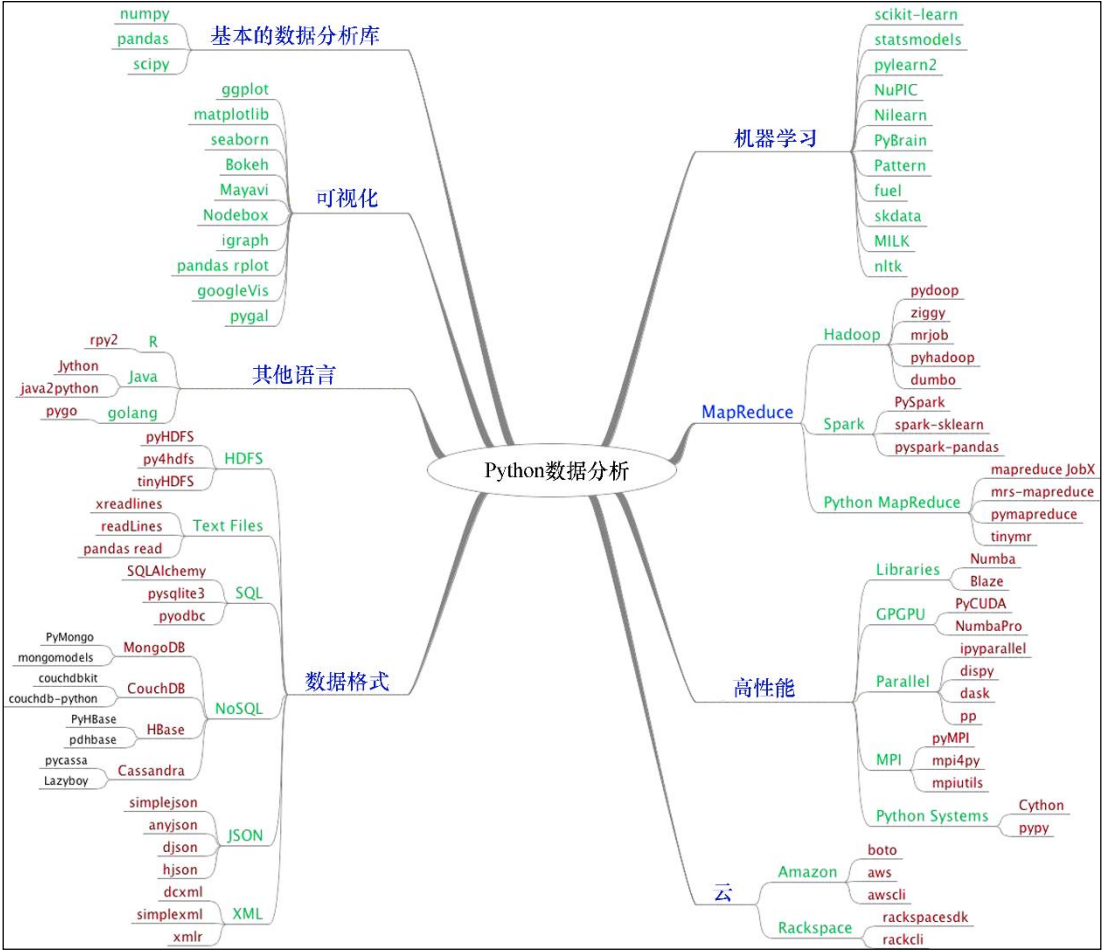- Rackspace：rackspacesdk、rackcli

图 1-1

```
[In [1]: import numpy

[In [2]: help(numpy.ar
  numpy.arange       numpy.arctan      numpy.argpartition  numpy.array2string
  numpy.arccos       numpy.arctan2     numpy.argsort       numpy.array_equal
  numpy.arccosh      numpy.arctanh     numpy.argwhere       numpy.array_equiv
  numpy.arcsin       numpy.argmax      numpy.around         numpy.array_repr
  numpy.arcsinh      numpy.argmin      numpy.array          numpy.array_split
```
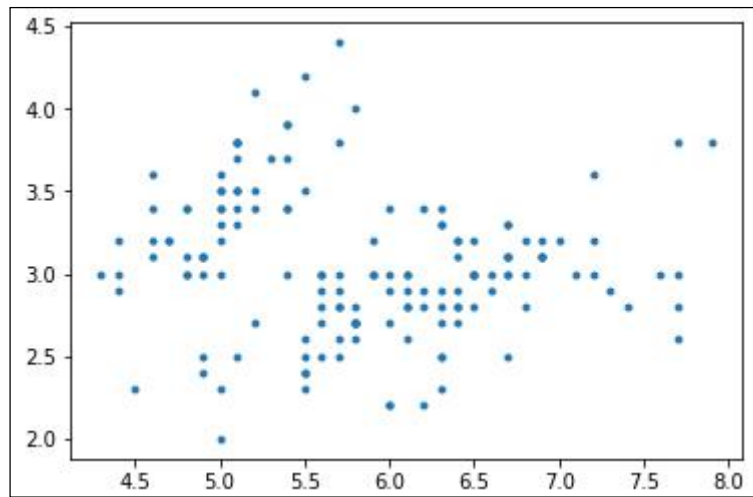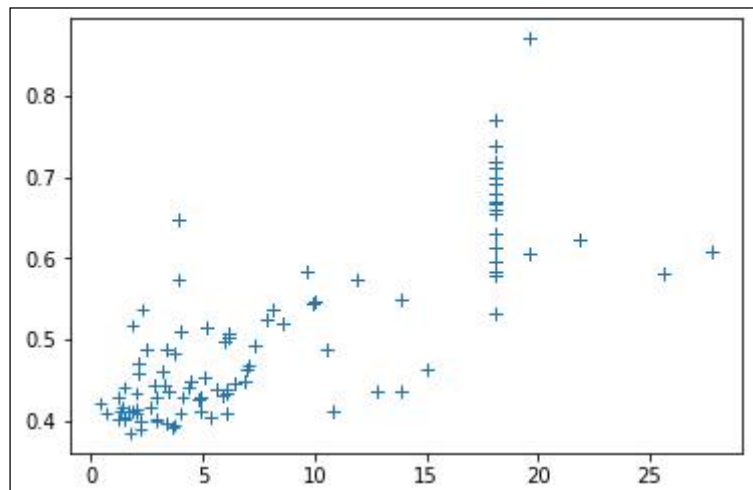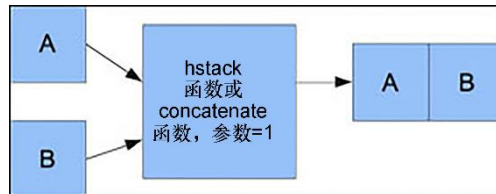
图 1-2

图 1-3

图 1-4

第 2 章　NumPy 数组

图 2-1

图 2-2
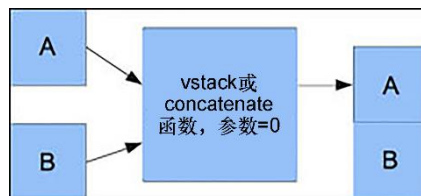
图 2-3
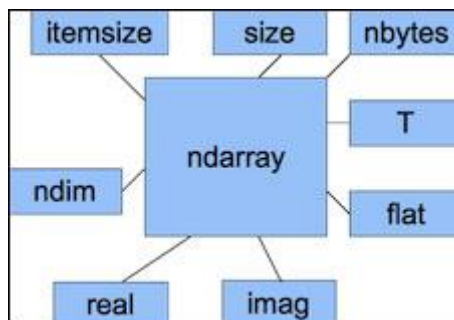
图 2-4
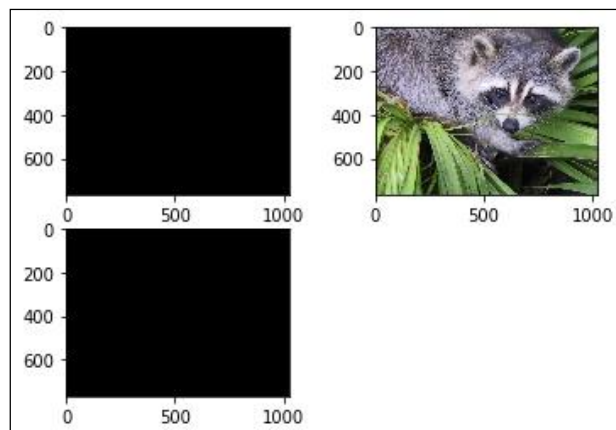
图 2-5

图 2-6

图 2-7

图 2-8

图 2-9

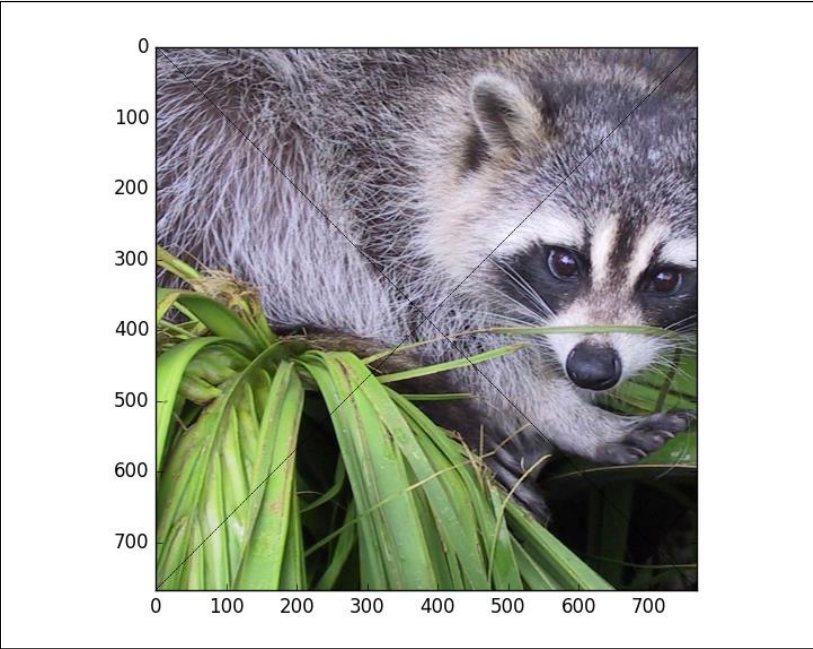# 第 3 章　Pandas 入门

```
Data types Country                                                object
CountryID                                                         int64
Continent                                                         int64
Adolescent fertility rate (%)                                   float64
Adult literacy rate (%)                                         float64
Gross national income per capita (PPP international $)          float64
Net primary school enrolment ratio female (%)                  float64
Net primary school enrolment ratio male (%)                    float64
Population (in thousands) total                                float64
```
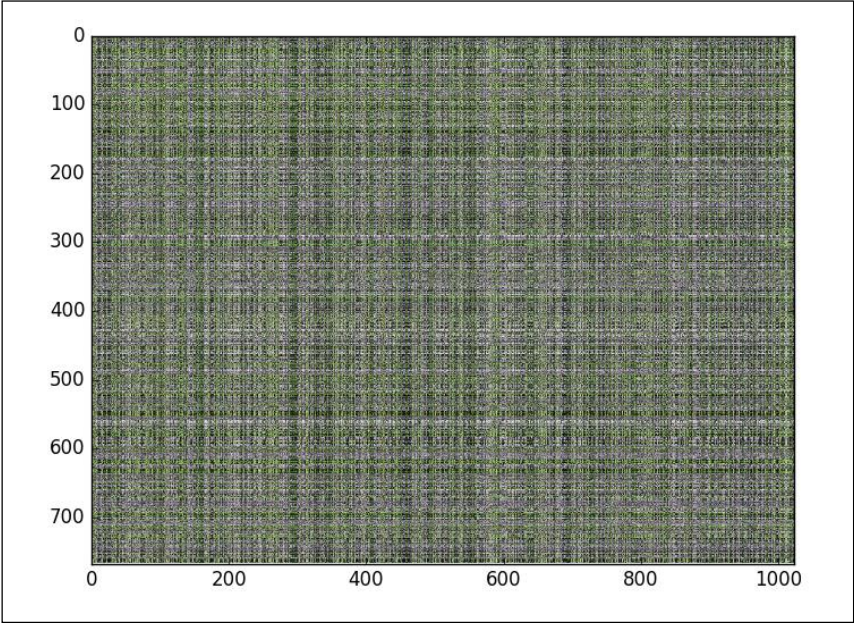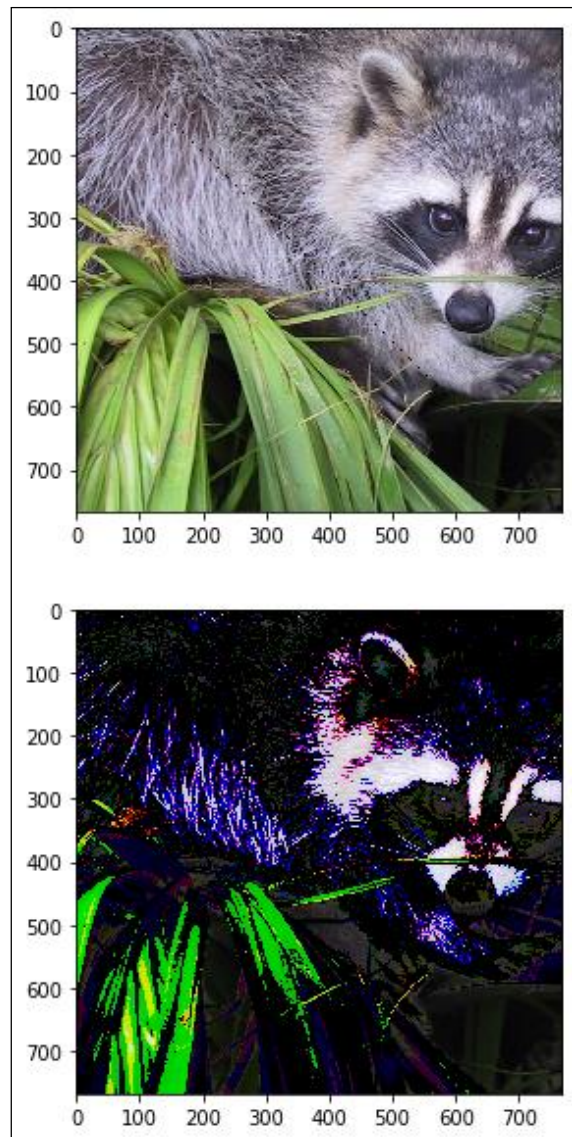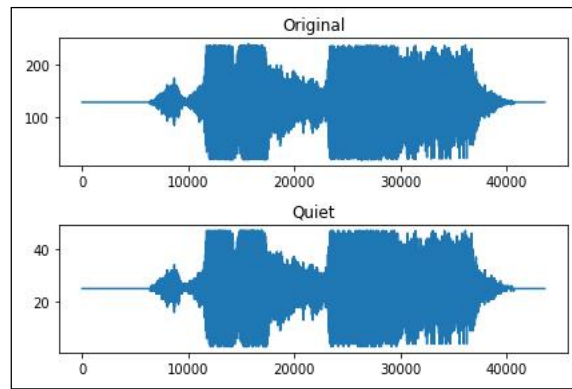
图 3-1　输出结果

```
Describe       Yearly Mean Total Sunspot Number  Yearly Mean Standard Deviation  \
count                       316.000000                       198.000000
mean                         79.503481                         8.030303
std                          62.057114                         3.807299
min                           0.000000                         1.700000
25%                          25.050000                         4.725000
50%                          66.700000                         7.700000
75%                         116.400000                        10.475000
max                         269.300000                        19.100000

        Number of Observations  Definitive/Provisional Indicator
count            198.000000                           316.0
mean            1424.888889                             1.0
std             2394.898980                             0.0
min              150.000000                             1.0
25%              365.000000                             1.0
50%              365.000000                             1.0
75%              366.000000                             1.0
max             8903.000000                             1.0

Non NaN observations Yearly Mean Total Sunspot Number     316
Yearly Mean Standard Deviation        198
Number of Observations                198
Definitive/Provisional Indicator      316
dtype: int64

MAD Yearly Mean Total Sunspot Number       50.987620
Yearly Mean Standard Deviation         3.125375
Number of Observations              1777.463524
Definitive/Provisional Indicator       0.000000
dtype: float64

Median Yearly Mean Total Sunspot Number     66.7
Yearly Mean Standard Deviation         7.7
Number of Observations               365.0
Definitive/Provisional Indicator       1.0
dtype: float64
```

图 3-2

```
Min Yearly Mean Total Sunspot Number        0.0
Yearly Mean Standard Deviation             1.7
Number of Observations                   150.0
Definitive/Provisional Indicator           1.0
dtype: float64

Max Yearly Mean Total Sunspot Number      269.3
Yearly Mean Standard Deviation            19.1
Number of Observations                  8903.0
Definitive/Provisional Indicator           1.0
dtype: float64

Mode    Yearly Mean Total Sunspot Number  Yearly Mean Standard Deviation  \
0                               18.3                              9.2

   Number of Observations  Definitive/Provisional Indicator
0                   365.0                               1.0

Standard Deviation Yearly Mean Total Sunspot Number       62.057114
Yearly Mean Standard Deviation            3.807299
Number of Observations                 2394.898980
Definitive/Provisional Indicator          0.000000
dtype: float64

Variance Yearly Mean Total Sunspot Number     3.851085e+03
Yearly Mean Standard Deviation           1.449552e+01
Number of Observations                   5.735541e+06
Definitive/Provisional Indicator         0.000000e+00
dtype: float64

Skewness Yearly Mean Total Sunspot Number     0.799452
Yearly Mean Standard Deviation           0.555067
Number of Observations                   1.876098
Definitive/Provisional Indicator         0.000000
dtype: float64

Kurtosis Yearly Mean Total Sunspot Number    -0.143733
Yearly Mean Standard Deviation          -0.244310
Number of Observations                   1.783261
Definitive/Provisional Indicator         0.000000
dtype: float64
```

图 3-3　脚本运行结果

第 4 章　统计学与线性代数

```
A
[[ 1 -2  1]
 [ 0  2 -8]
 [-4  5  9]]
b
[ 0  8 -9]
```

图 4-1

图 4-2

图 4-3

原图　　　　　　　掩码处理过的原图

对数图　　　　　　掩码处理过的对数图
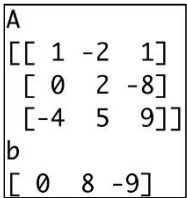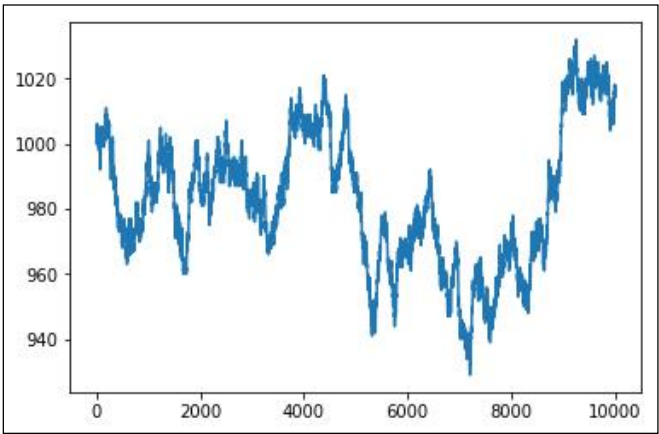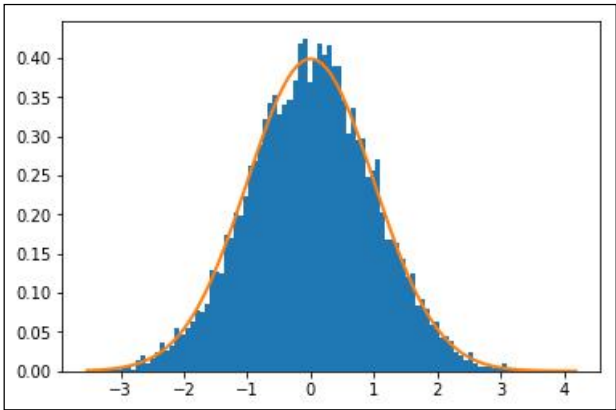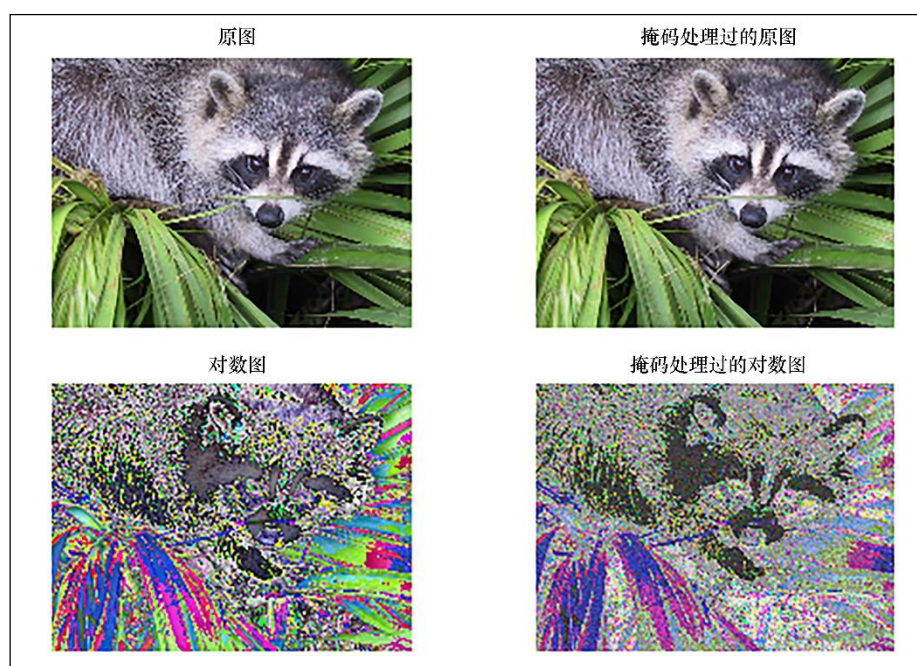
图 4-4
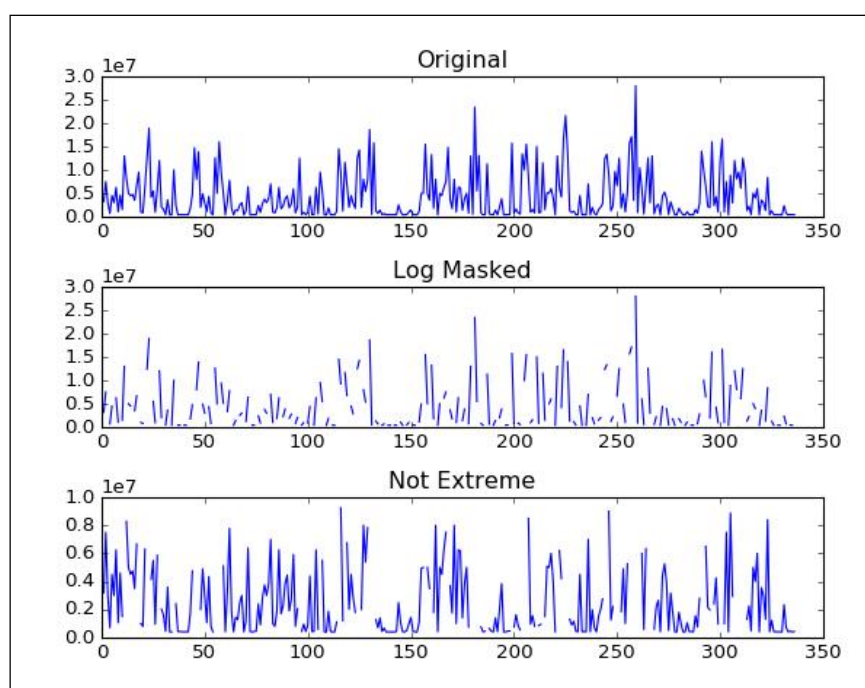
图 4-5

第 5 章　数据的检索、加工与存储
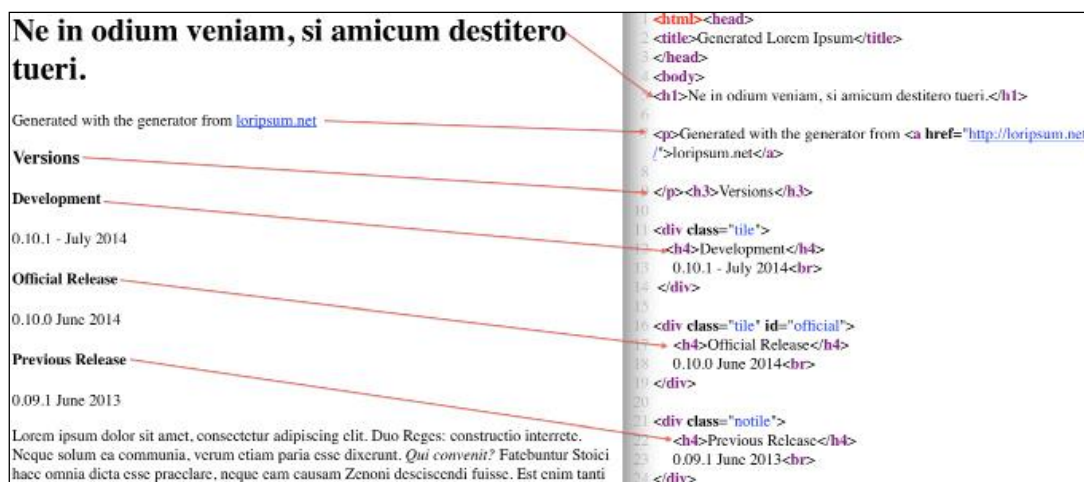
图 5-1
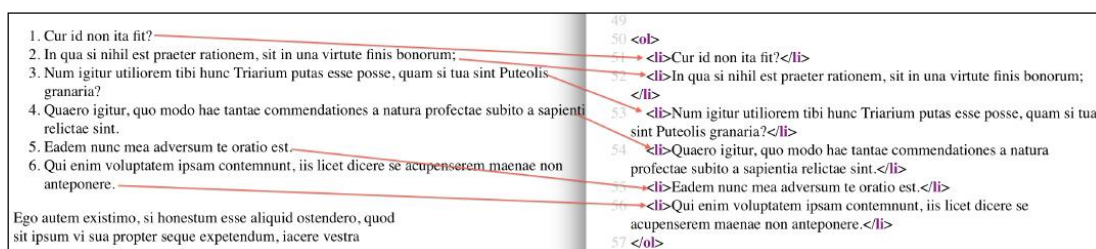
图 5-2

第 6 章　数据可视化
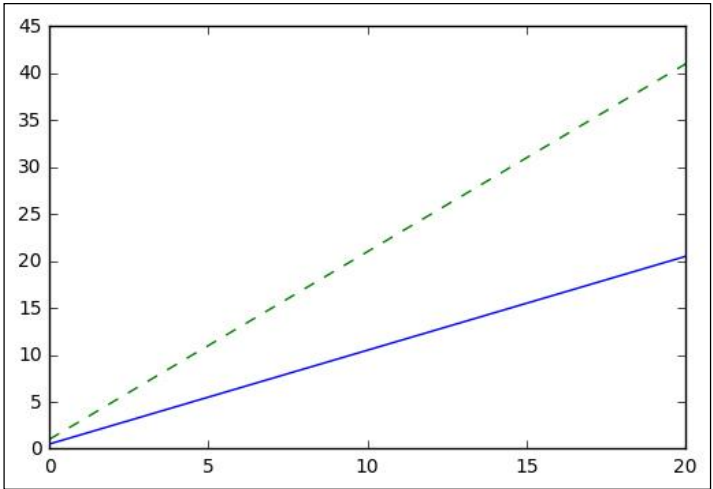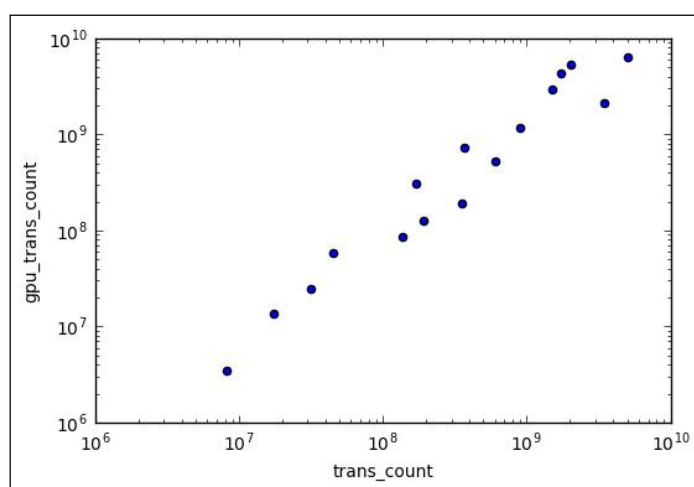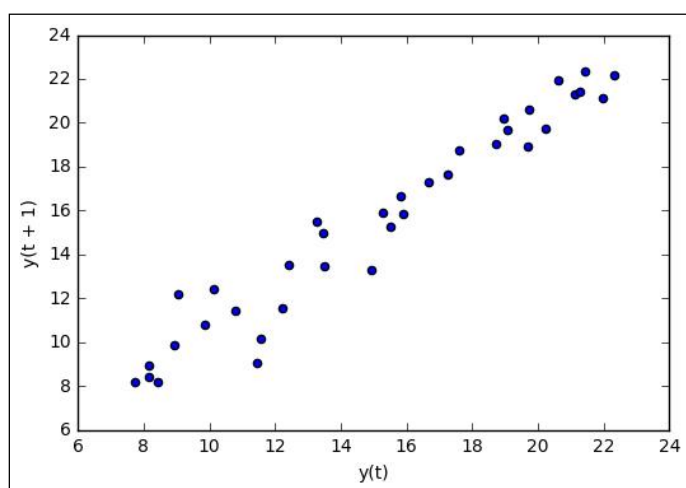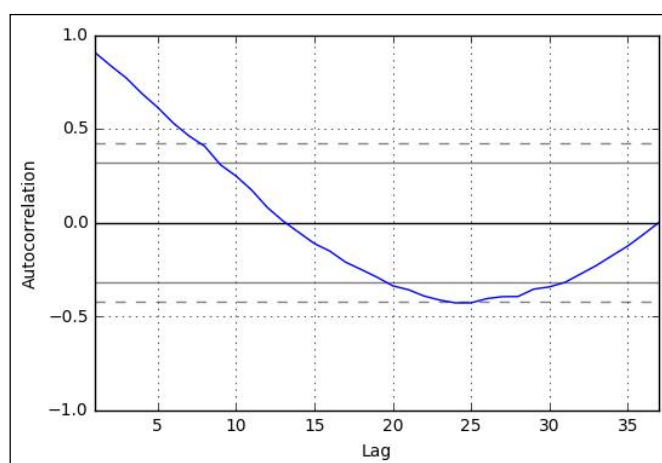

图 6-1
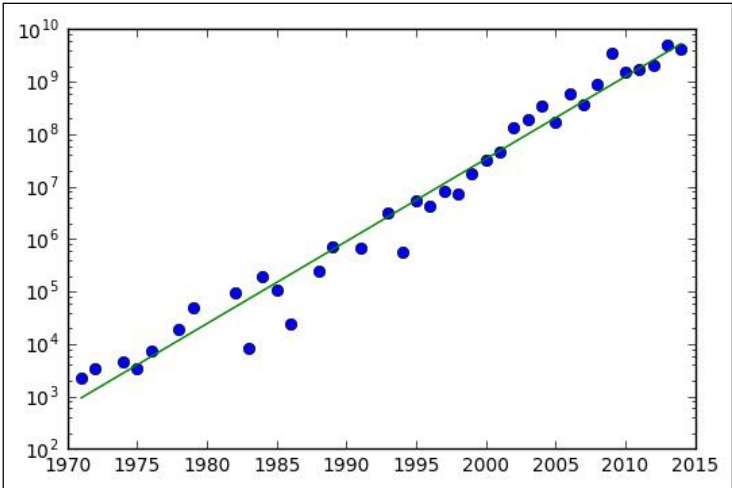
图 6-2
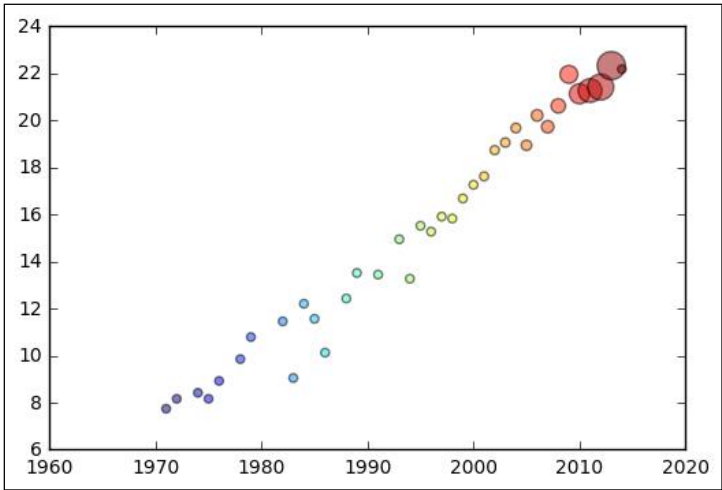
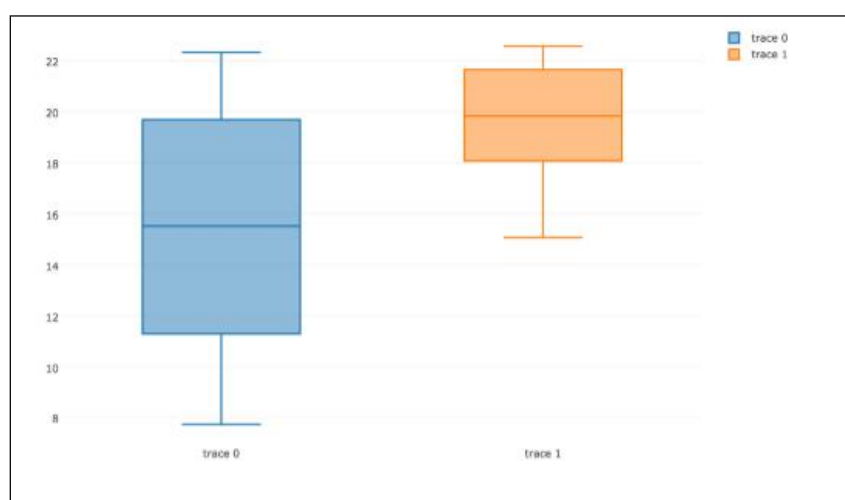图 6-3
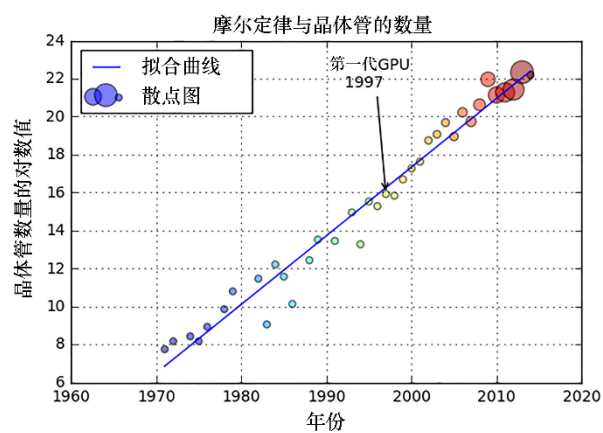
图 6-4
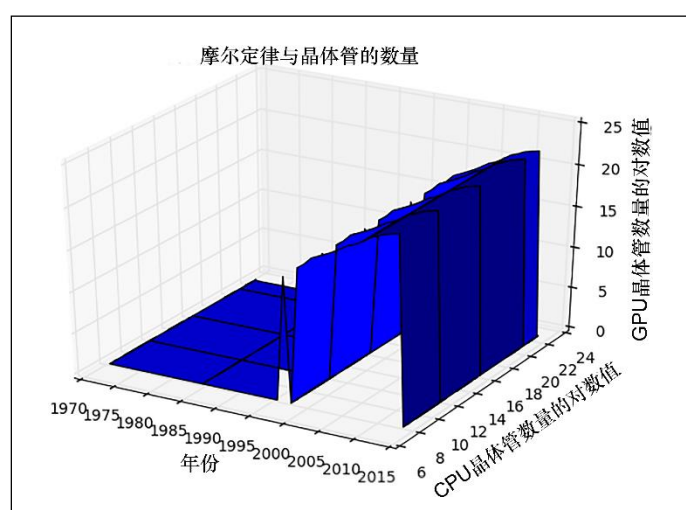
图 6-5
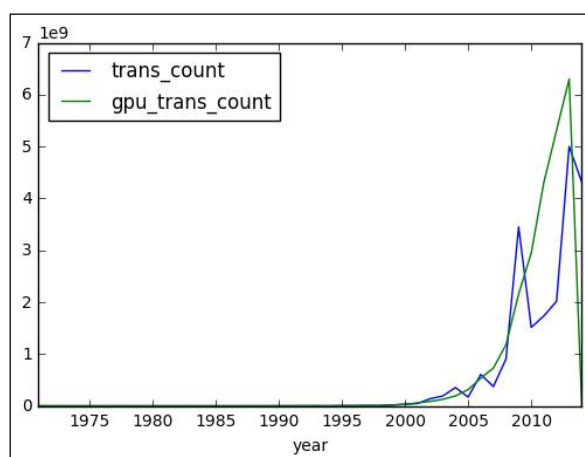
图 6-6
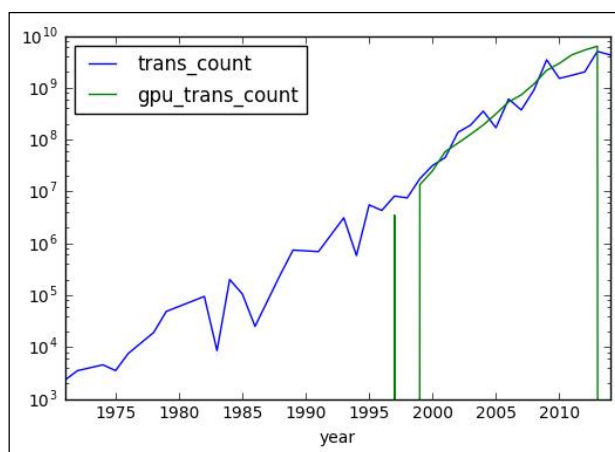
图 6-7

图 6-8

图 6-9

图 6-10

图 6-11

第 7 章 信号处理与时间序列
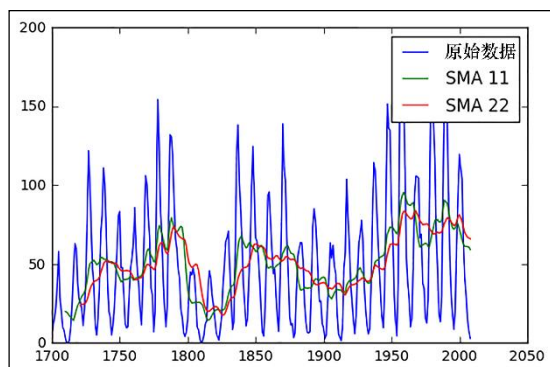
图 7-1
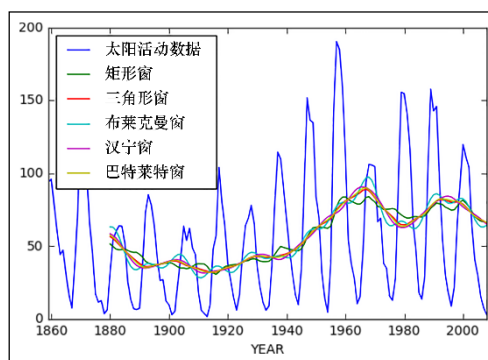
图 7-2
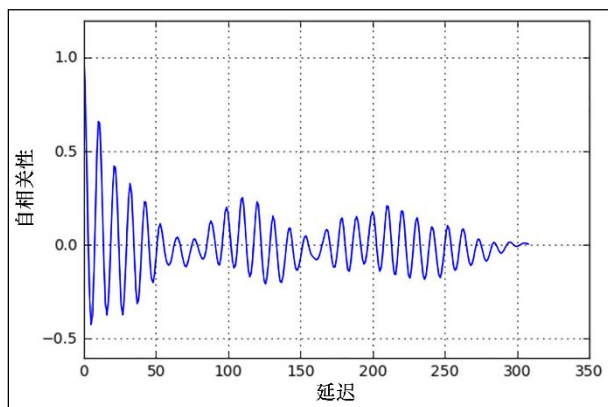
图 7-3
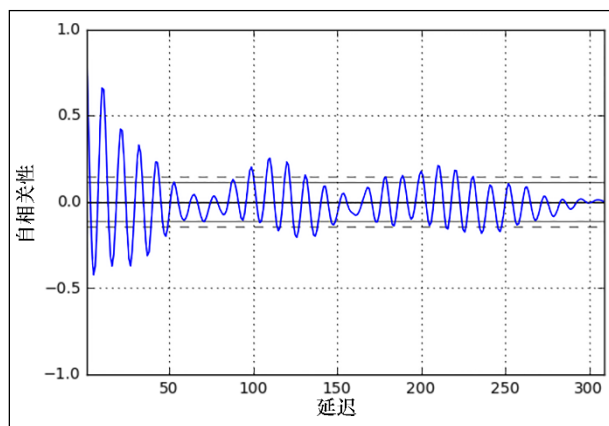
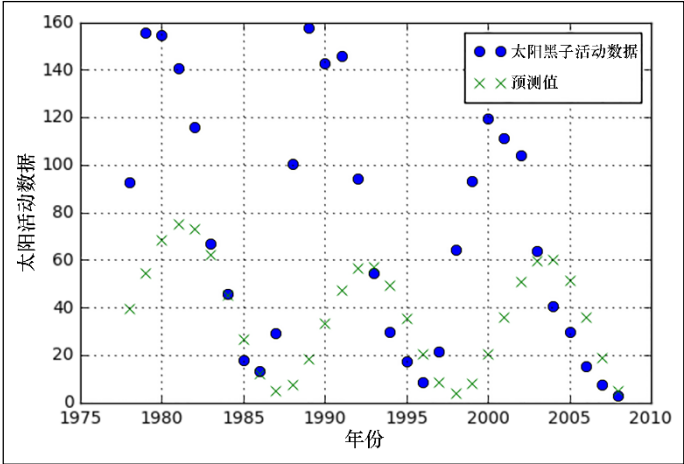图 7-4
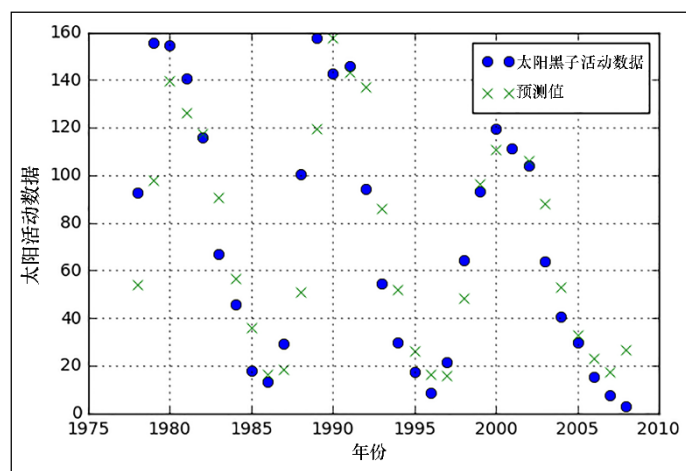
图 7-5
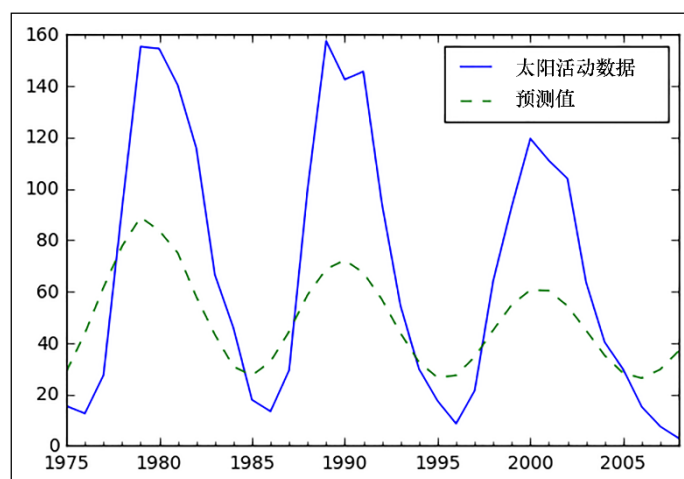
图 7-6

图 7-7
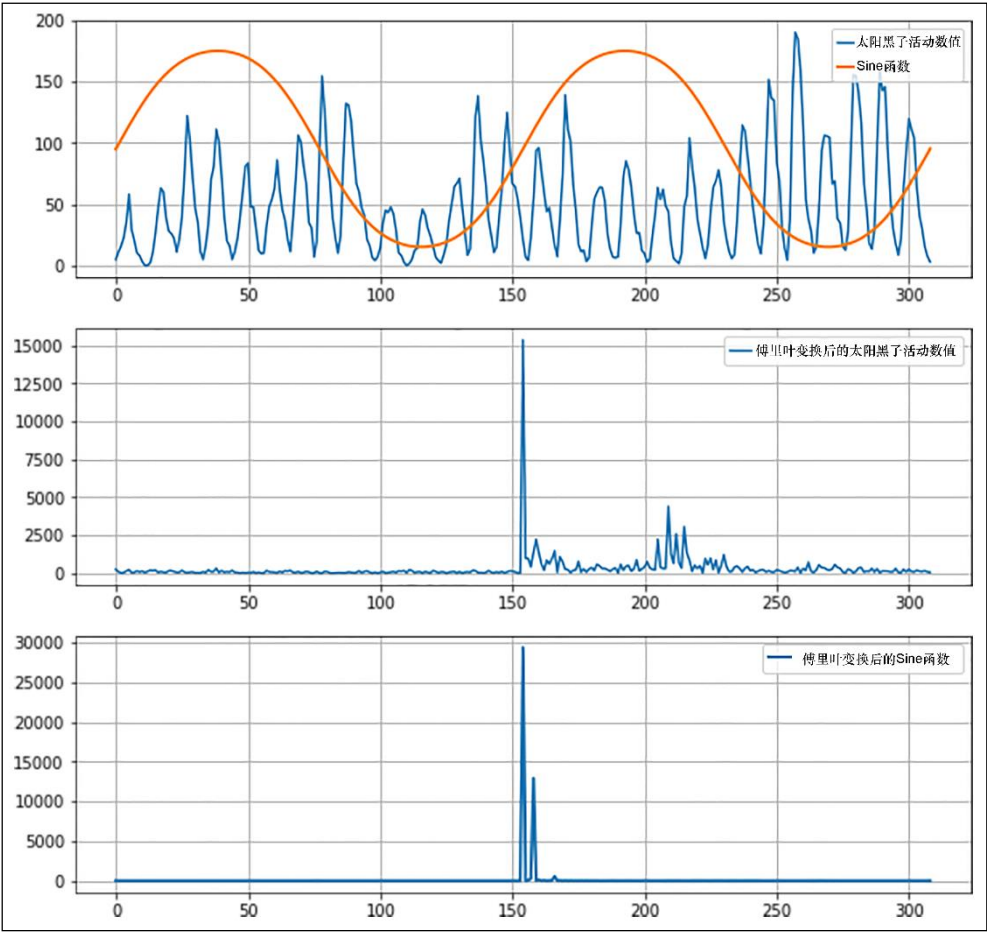
图 7-8
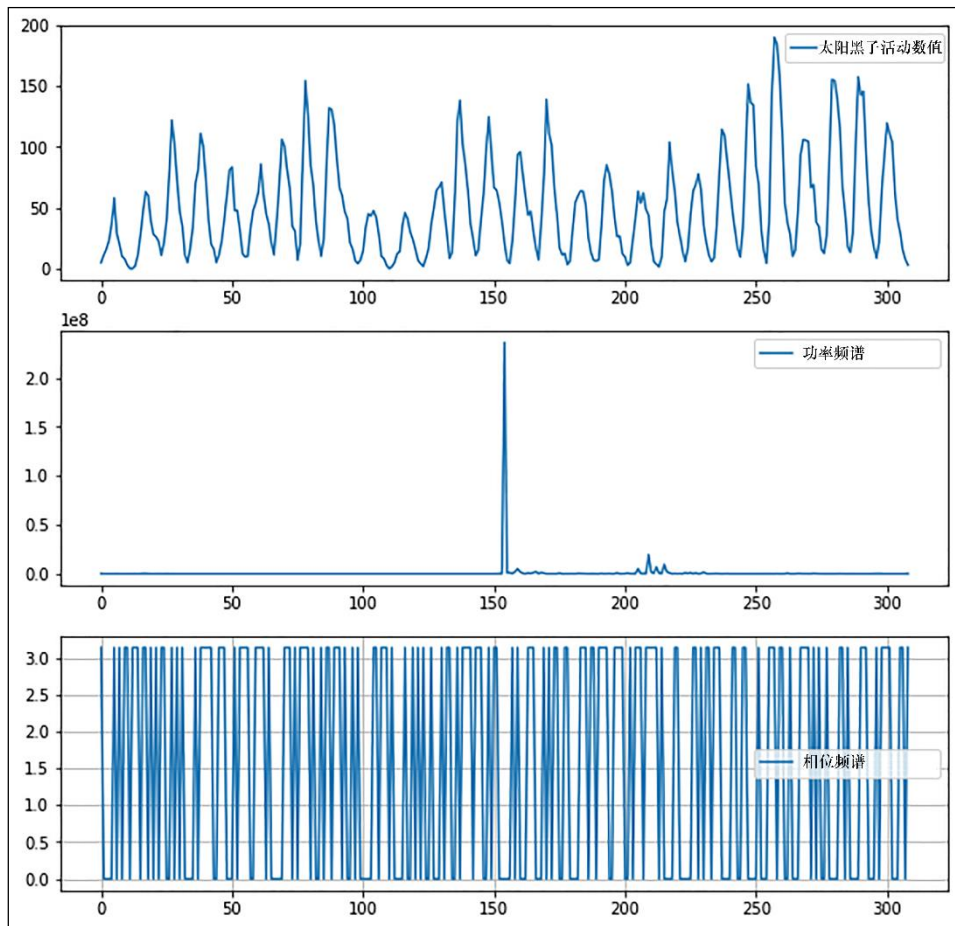
图 7-9
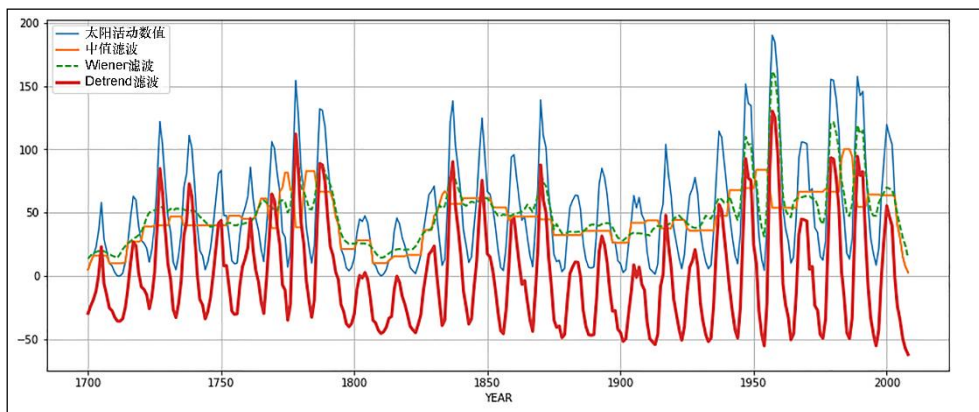
图 7-10

第 9 章 分析文本数据和社交媒体
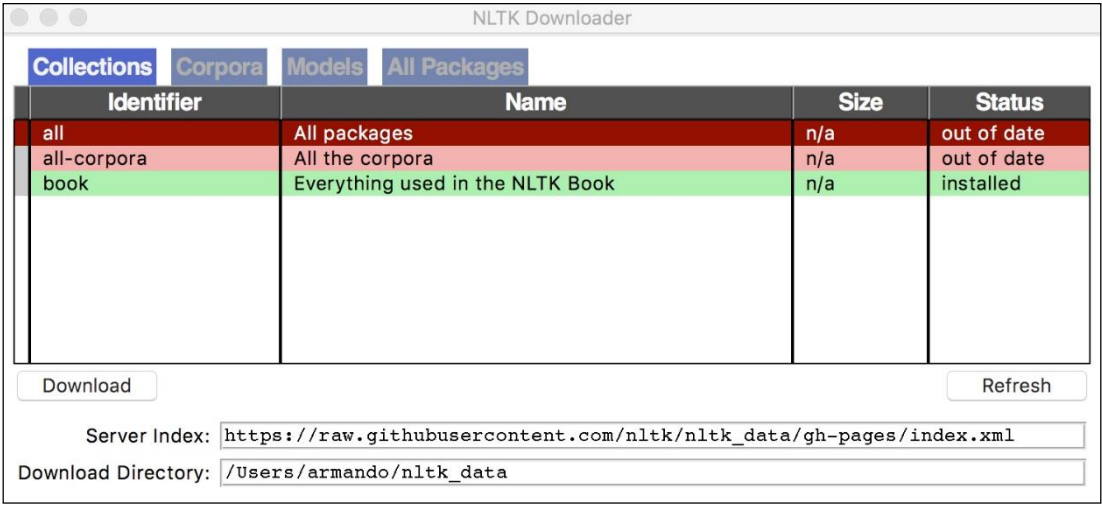
图 9-1

```
len = 7            False : True   =     62.7 : 1.0
len = 6            False : True   =     49.1 : 1.0
len = 1             True : False  =     12.0 : 1.0
len = 2             True : False  =     10.7 : 1.0
len = 5            False : True   =     10.4 : 1.0
```

图 9-2

```
     count (wonderful) = 2         pos : neg   =     14.7 : 1.0
  count (outstanding) = 1          pos : neg   =     11.2 : 1.0
          count (bad) = 5          neg : pos   =     10.8 : 1.0
       count (stupid) = 2          neg : pos   =     10.8 : 1.0
       count (boring) = 2          neg : pos   =     10.4 : 1.0
       count (nature) = 2          pos : neg   =      8.5 : 1.0
    count (different) = 2          pos : neg   =      8.3 : 1.0
          count (bad) = 6          neg : pos   =      8.2 : 1.0
   count (apparently) = 2          neg : pos   =      8.0 : 1.0
         count (life) = 5          pos : neg   =      7.6 : 1.0
```
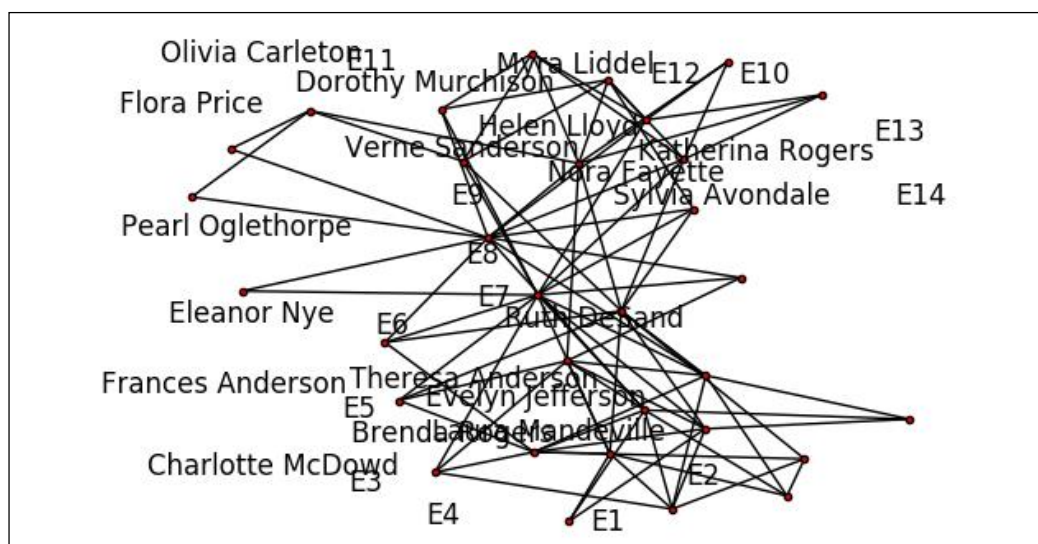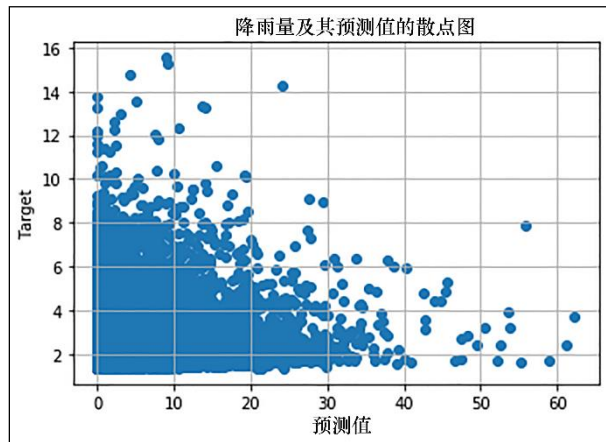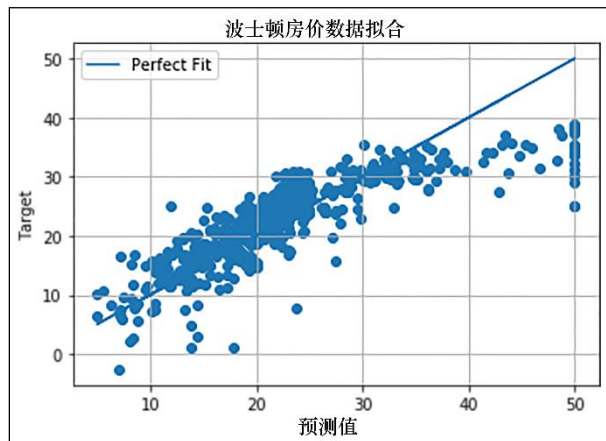
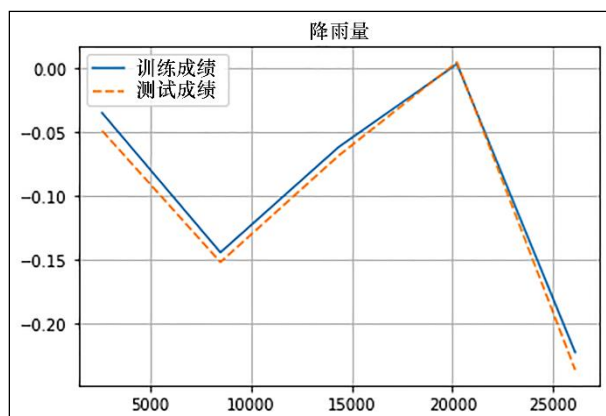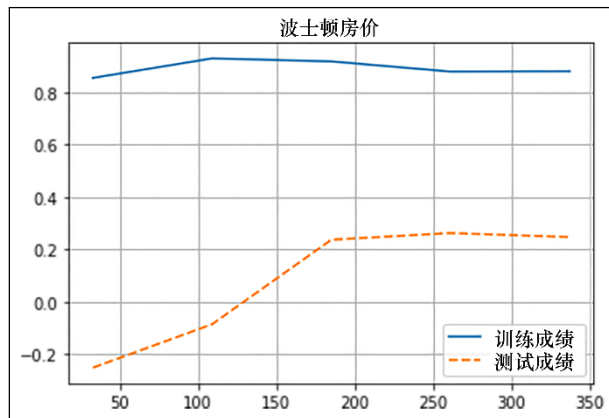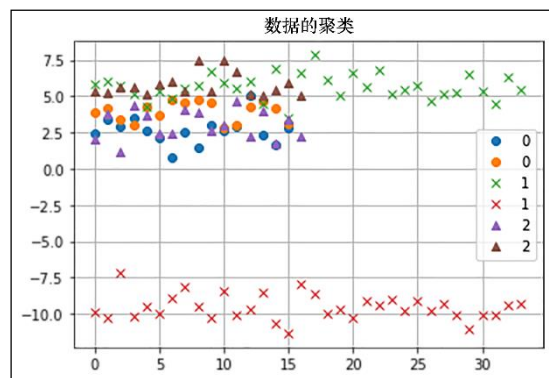图 9-3

图 9-4

图 9-5

图 9-6

图 9-7

第 10 章 预测性分析与机器学习

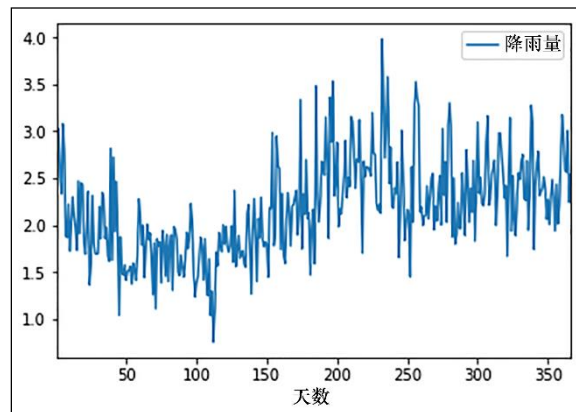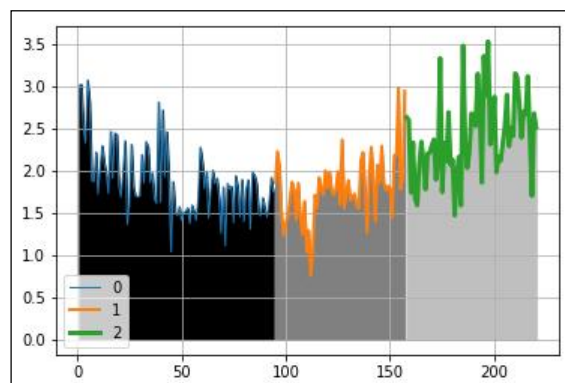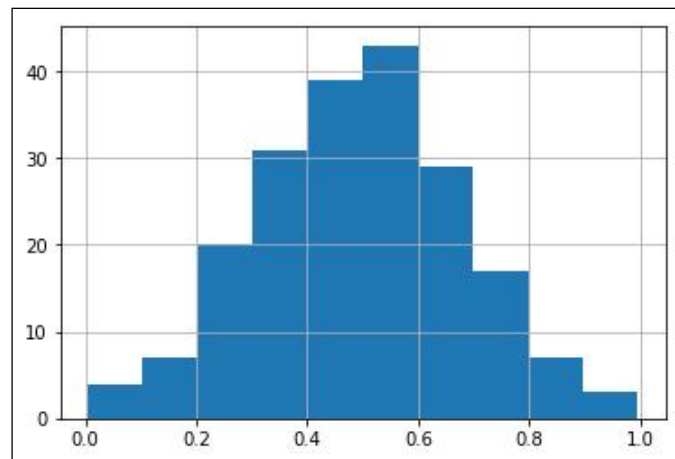
图 10-1

图 10-2

图 10-3

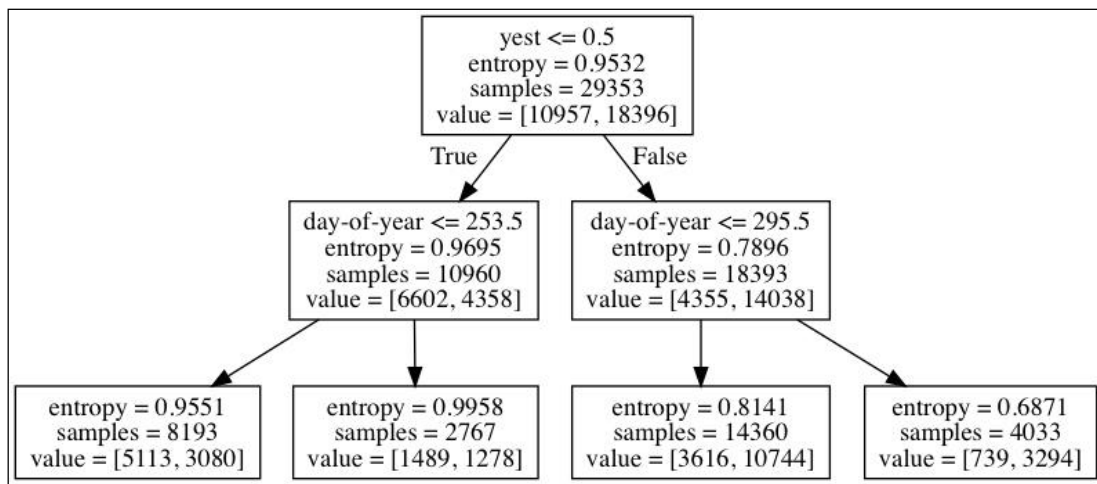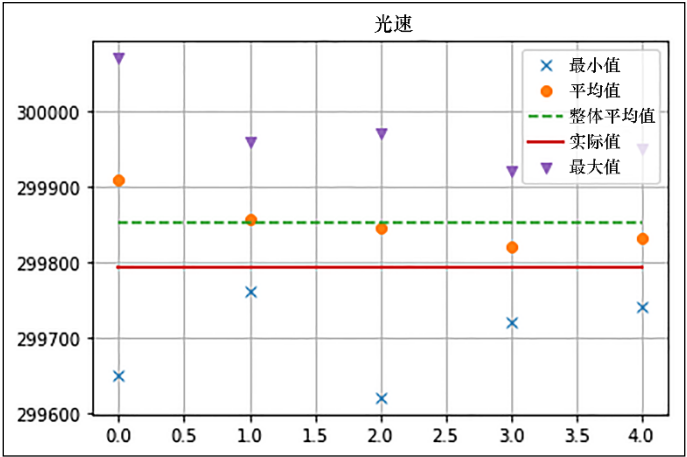图 10-4

图 10-5

图 10-6

图 10-7

图 10-8

图 10-9

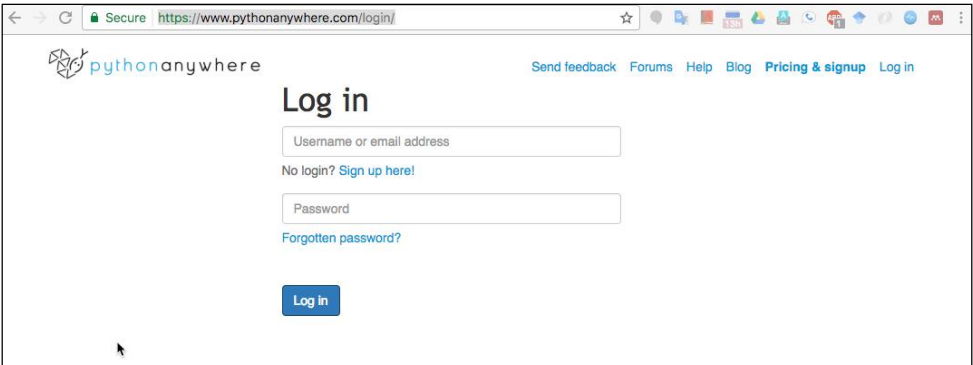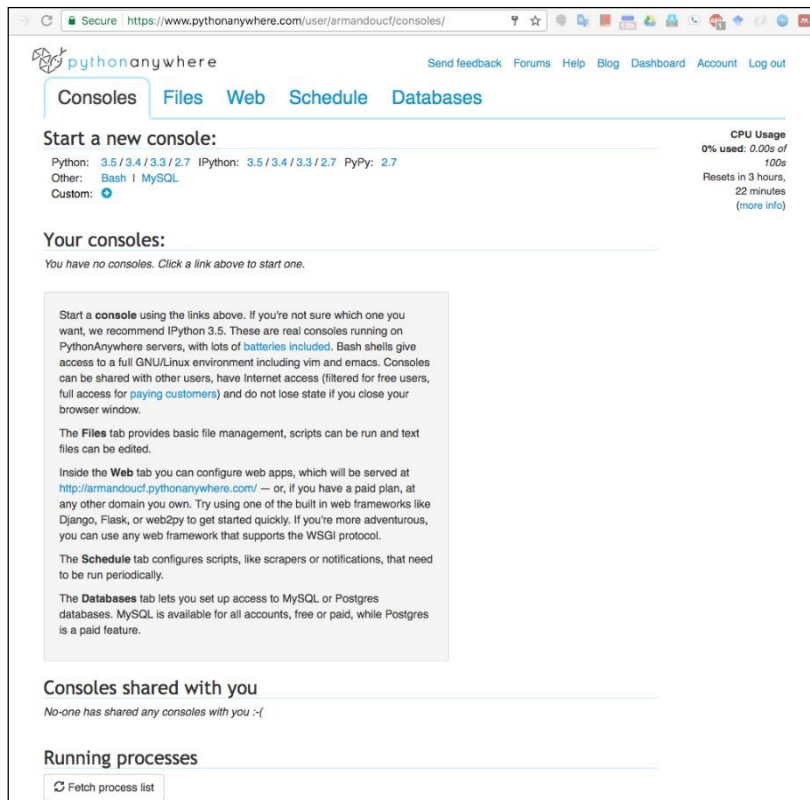# 第 11 章 Python 生态系统的外部环境和云计算
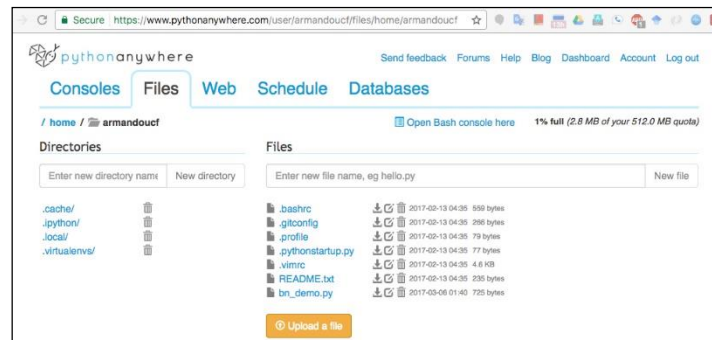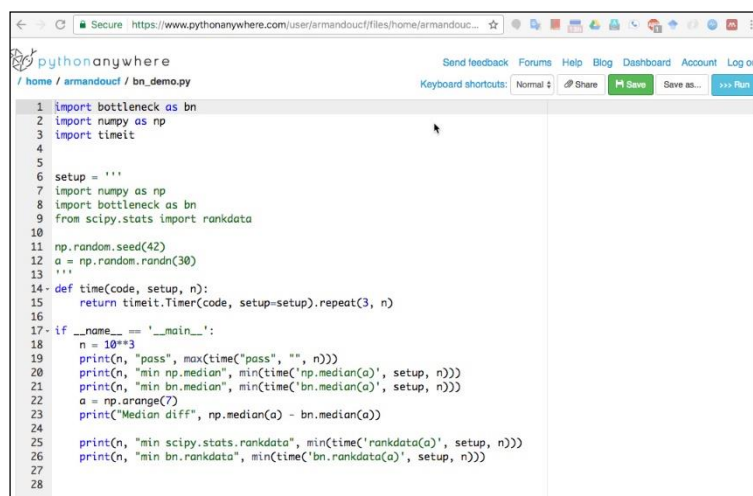

图 11-1

图 11-2

图 11-3

图 11-4

图 11-5


```
1000 pass 9.993906132876873e-06
1000 min np.median 0.057233922998420894
1000 min bn.median 0.0007603260455653071
Median diff 0.0
1000 min scipy.stats.rankdata 0.11264373897574842
1000 min bn.rankdata 0.0016925800059170723
>>>
```

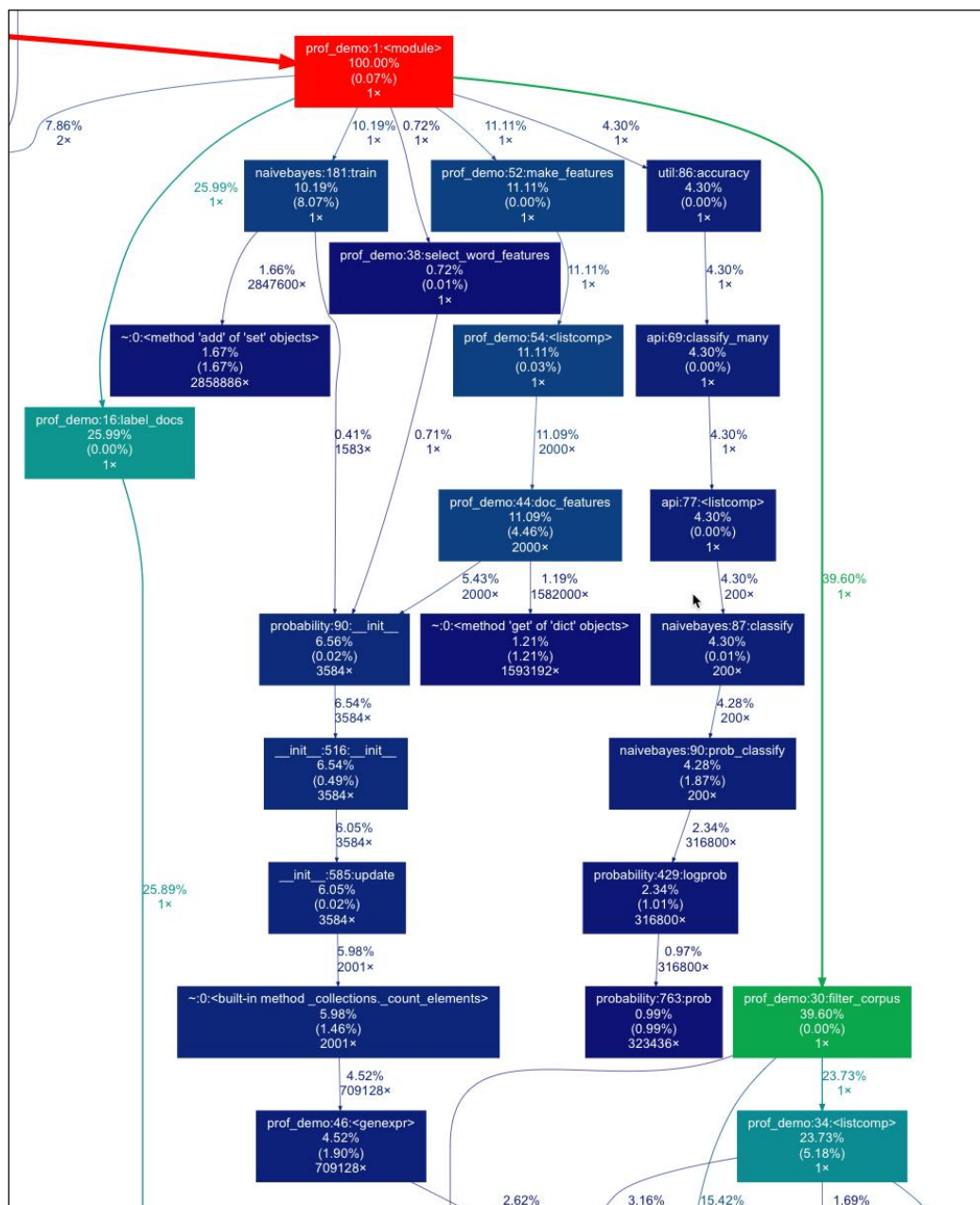图 11-6

第 12 章 性能优化、性能分析与并发性

图 12-1

```
[fluid:ch-12 armando$ python3 -m pstats /tmp/stat.prof
 Welcome to the profile statistics browser.
[/tmp/stat.prof% strip
[/tmp/stat.prof% sort time
[/tmp/stat.prof% stats 10
 Sun Feb  5 18:24:49 2017    /tmp/stat.prof

          30643998 function calls (30123080 primitive calls) in 15.502 seconds

   Ordered by: internal time
   List reduced from 3823 to 10 due to restriction <10>

   ncalls  tottime  percall  cumtime  percall filename:lineno(function)
   319962    2.397    0.000    2.397    0.000 {method 'findall' of '_sre.SRE_Pattern' objects}
        1    1.251    1.251    1.580    1.580 naivebayes.py:181(train)
   319960    1.056    0.000    2.748    0.000 data.py:1114(readline)
  6343280    0.847    0.000    7.338    0.000 util.py:261(iterate_from)
        1    0.803    0.803    3.678    3.678 prof_demo.py:34(<listcomp>)
  3167640    0.741    0.000    0.896    0.000 prof_demo.py:26(isStopWord)
     2000    0.692    0.000    1.719    0.001 prof_demo.py:44(doc_features)
  3167642    0.628    0.000    4.361    0.000 util.py:388(iterate_from)
   371223    0.393    0.000    0.898    0.000 data.py:1353(_read)
3885294/3376152    0.359    0.000    2.696    0.000 {built-in method builtins.len}


/tmp/stat.prof% □
```
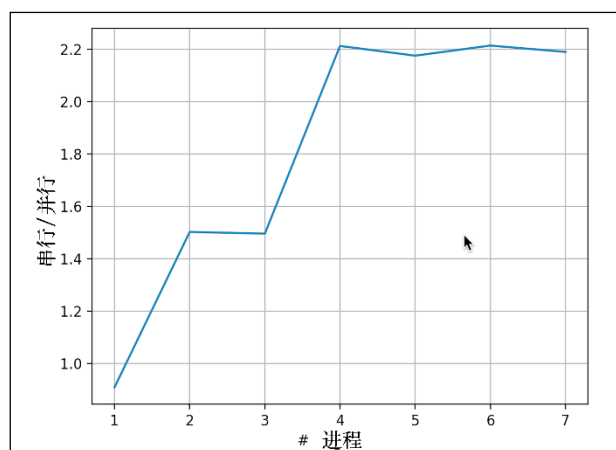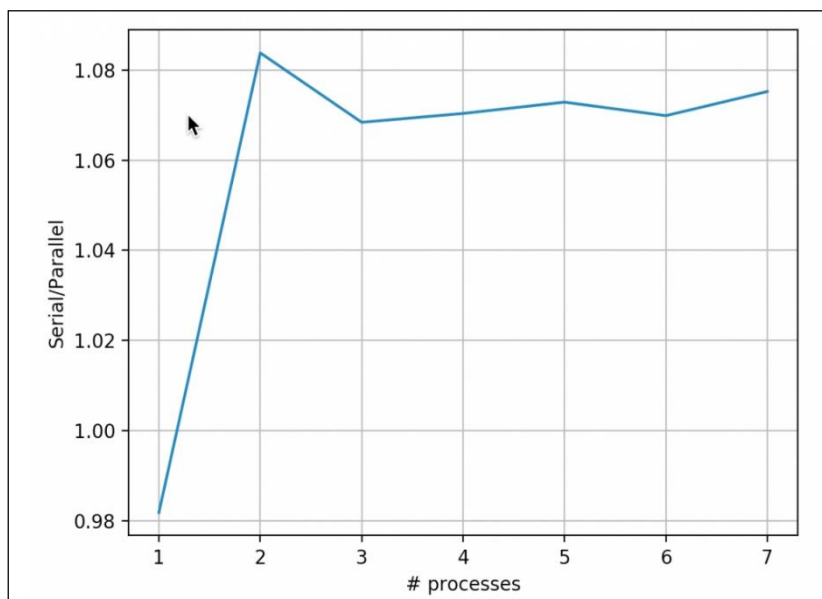
图 12-2

图 12-3

图 12-4