1.NumPy

Travis Oliphant的作品。Python主力分析工具。为用户提供了多维数组，以及对这些数组进行多种数学操作的大型函数集。

pip install numpy

import numpy as np

2.SciPy

Travis Oliphant、Pearu Peterson和Eric Jones等人的原创作品。完善了NumPy的功能，为多种应用提供了大量科学算法，如线性代数、稀疏矩阵、信号和图像处理、最优化、快速傅里叶变换等。

pip install scipy

3.pandas

由Wes McKinney创作。可处理包含不用类型数据的复杂表格，亦可轻松又顺利的加载各种形式的数据。然后对数据进行切片、切块、处理缺失元素、添加、重命名、聚合、整形和可视化等操作。

pip install pandas

4.Scikit-learn

最初是SciKits（SciPy工具包）的一部分，是python数据运算的核心。提供了所有机器学习可能用到的工具，如数据预处理、监督学习和无监督学习、模式选择验证和误差指标等。原为谷歌编程之夏（Google Summer of Code）的一个项目，由David Cournapeau于2007年发起，自2013年开始，被INRIA（法国国家信息与自动化研究所）的研究人员接管。

pip install scikit-learn

5.IPython

科学方法需要对不同的假设可再现地进行快速的验证。Fernando Perez创建了IPython，满足了python交互式shell命令的需要，它是基于shell、web浏览器和应用程序接口的python版本。具有图形化继承、自定义指令、丰富的历史记录（JSON格式）和并行计算等增强功能。它通过监本、数据和相应结果清晰又有效地说明了各种操作。

pip install “ipython[notebook]”

6.Matplotlib

由John Hunter原创开发，是一个包含各种绘图模块的库，能根据数组创建高质量的图形，并交互式地显示它们。

Matplotlib提供了pylab模块，包含许多像MATLAB一样的绘图组件。

pip install matplotlib

import matplotlib.pyplot as plt

7.Statsmodels

之前为SciKits的一部分，是SciPy统计函数的补充。包含通用线性模型、离散选择模型、时间序列分析、一系列描述统计以及参数和非参数校验等特性。

pip install statsmodels

8.Beautiful soup

由Leonard Richardson创建，是一个很棒的HTML/XML解析器，用来分析从互联网上抽取的HTML和XML文档。选择解析器之后（一般情况下，Python标准库中的HTML解析器效果就很好），就可以对页面上的对象定位，并提取文本、表格以及其他有用的信息。

pip install beautifulsoup4

from bs4 import BeautifulSoup

9.NetworkX

由美国洛斯阿拉莫斯国家实验室（Los Alamos National Laboratory）开发。是一个专门进行现实生活网络数据创建、操作、分析和图示的软件包，它可以轻松的进行具有百万个节点和边的图操作。除了专门的图数据结构和良好的可视化方法（2D和3D），它为用户提供了许多标准的图的度量方法和算法，如最短路径、中心性、成本、群体、聚类和网页排名。

pip install networkx

import network as nx

10.NLTK

自然语言工具箱（NLTK）能够访问语料和词汇库，提供从分词到词性标注、从树模型到命名实体识别等统计自然语言处理（Natural Language Processing，NLP）的一整套函数。可进行自然语言处理系统的原型开发和搭建。最初，该软件是Steven Bird和Edward Loper为宾夕法尼亚大学自然语言处理教学设施CIS-530开发的。

pip install nltk

11.Gensim

由Radim Rehurek开发的开源软件包，在并行分布式在线算法的帮助下，能进行大型文本集合分析。它具有许多高级功能，如实现了潜在语义分析（Latent Semantic Analysis，LSA）、通过LDA（Latent Dirichlet Allocation）进行主题建模等。Gensim还包含功能强大的谷歌word2vec算法，能将文本转换为矢量特征，再使用此矢量进行监督和无监督学习。

pip install gensim

12.PyPy

PyPy不是软件包，它是Python2.7.8的替代品。一个主要优势是提高了运行速度及内存处理能力，因此非常适用于大型数据上的繁重操作。

http://pypy.org/download.html