查看使用包的路径

import numpy as np

np.\_\_file\_\_

第三方库下载地址：

[**Unofficial Windows Binaries for Python Extension Packages**](http://www.lfd.uci.edu/~gohlke/pythonlibs/)

<http://www.lfd.uci.edu/~gohlke/pythonlibs/>

一、常用第三方库

|  |  |  |
| --- | --- | --- |
| 常用Python第三方库 | | |
| **分类** | **库名称** | **库用途** |
| *Web框架* | **Django** | 开源web开发框架，它鼓励快速开发,并遵循MVC设计，我以前用过很多次，比较好用，开发周期短 |
| ActiveGrid | 企业级的Web2.0解决方案 |
| Karrigell | 简单的Web框架，自身包含了Web服务，py脚本引擎和纯python的数据库PyDBLite |
| webpy | 一个小巧灵活的Web框架，虽然简单但是功能强大 |
| CherryPy | 基于Python的Web应用程序开发框架 |
| Pylons | 基于Python的一个极其高效和可靠的Web开发框架 |
| Zope | 开源的Web应用服务器 |
| TurboGears | 基于Python的MVC风格的Web应用程序框架 |
| Twisted | 流行的网络编程库，大型Web框架 |
| Quixote | Web开发框架 |
| *科学计算* | Matplotlib | 用Python实现的类matlab的第三方库，用以绘制一些高质量的数学二维图形 |
| SciPy | 基于Python的matlab实现，旨在实现matlab的所有功能 |
| NumPy | 基于Python的科学计算第三方库，提供了矩阵，线性代数，傅立叶变换等等的解决方案 |
| *GUI* | PyGtk | 基于Python的GUI程序开发GTK+库 |
| PyQt | 用于Python的QT开发库 |
| WxPython | Python下的GUI编程框架，与MFC的架构相似 |
| **Tkinter** | Python下标准的界面编程包，因此不算是第三方库了o(∩\_∩)o |
| *其它* | **BeautifulSoup** | 基于Python的HTML/XML解析器，简单易用 |
| PIL | 基于Python的图像处理库，功能强大，对图形文件的格式支持广泛 |
| **MySQLdb** | 用于连接MySQL数据库 |
| cElementTree | 高性能XML解析库，Py2.5应该已经包含了该模块，因此不算一个第三方库了 |
| PyGame | 基于Python的多媒体开发和游戏软件开发模块 |
| **Py2exe** | 将python脚本转换为windows上可以独立运行的可执行程序 |
| **pefile** | Windows PE文件解析器 |

二、使用时安装（第三方库）

conda install scikit-learn

conda install numpy 升级或安装包、库

conda install scipy

conda install pandas

conda install matplotlib

conda install pyreadline

conda install pyzmq

conda install tornado

conda install mathjax

conda install sympy

conda install xlrd

conda install xlwt

conda install mayavi

conda install vtk

conda install tvtk

conda install mathjax

conda install statsmodels

conda install jinja2

conda install networkx

conda install pandoc

pip install pydot

pip instal ggplot

pip install python-louvain

pip install pandas\_datareader

pip install seaborn

pip install scrapy

pip install Mysqlclient

pip install gensim

# conda install mathjax tvtk vtk

[Scikit-learn](http://scikit-learn.org/) 机器学习库

numpy 数值计算库

scipy 数值计算库

numpy提供了矩阵和向量类，而scipy依赖numpy，它封装了科学计算的算法。简单的说， numpy提供对象，scipy提供了算法。

pandas

matplotlib 2D绘图库

pyreadline 补全用

pyzmq 并发用的,是基于消息的网络模型的东东(ZeroMQ)

tornado Http服务器

MathJax 渲染用的可以将公式直接渲染出来

sympy 使用符号运算

xlrd Excel读取

xlwt Excel写入

mayavi 可视化软件，其面向脚本的API：mlab快速绘制三维图

（20150809）Mayavi还没有py3的支持

VTK(Visualization Toolkit) 三维的数据可视化工具，由C++编写

在Python下有标准的绑定，不过其API和C++相同，不能体现出Python作为动态语言的优势

TVTK 三维可视化数据

在标准的VTK库之上用Traits库进行封装，提供了Python风格的API、支持Trait属性和numpy的多维数组

mathjax 数学公式的显示

pydot This module provides with a full interface to create handle modify and process graphs in Graphviz’s dot language

statsmodels 统计建模和计量经济学工具包，包括一些描述统计、统计模型估计和推断

jinja2 能将模板和用于填充模板的数据源组合在一起的一个Python库

ggplot 类似R中的ggplot2的绘图工具

NetworkX 是一个用Python语言开发的图论与复杂网络建模工具，内置了常用的图与复杂网络分析算法，可以方便的进行复杂网络数据分析、仿真建模等工作。

Python-louvain 是一个目的单纯的很小的Python库，作用是使用Louvain方法来发现社群。

pandas\_datareader 远程数据获取工具 Functions from pandas\_datareader.data and pandas\_datareader.wb extract data from various Internet sources into a pandas DataFrame.

seaborn 绘图，类似matplotlib

rpy2 Python调用R做统计分析

Pandoc 是由[John MacFarlane](http://baike.baidu.com/item/John%20MacFarlane)开发的[标记语言](http://baike.baidu.com/view/329009.htm)转换工具，可实现不同标记语言间的格式转换，堪称该领域中的“[瑞士军刀](http://baike.baidu.com/view/4053.htm)”

scrapy 网络爬虫框架（先安装twisted）

[Mysqlclient](http://www.lfd.uci.edu/~gohlke/pythonlibs/#mysqlclient) is a Python 3 compatible fork of MySQL-python.

Gensim是一个相当专业的主题模型Python工具包。在文本处理中，比如商品评论挖掘，有时需要了解每个评论分别和商品的描述之间的相似度，以此衡量评论的客观性。

pip install mne （速度太慢）

mne MEG和EEG数据处理库

下载包之后安装：<https://github.com/mne-tools>

第三方库下载后安装方法：（Windows）

1 确保环境变量中有：

D:\software\Anaconda3\envs\py27、

D:\software\Anaconda3\envs\py27\Scripts、

D:\software\Anaconda3\envs\py27\Library\bin

三个路径（我的电脑-属性-高级系统设置-环境变量-用户变量-编辑）

2 打开Windows shell，cd到有setup.py的文件夹，如：

cd D:\XH\analysis\_prog\ML\_Multiclass\_Prog\_Python27\_20161028\MNE\_Python\mne\_python

3 运行以下语句，prefix后面是指定安装路径。

一般，python setup.py install 即可安装到默认路径（环境变量里最上面的）

python setup.py install --prefix=D:\software\Anaconda3\envs\py27

4 import mne 验证安装成功与否

whl文件安装：

1 下载whl 文件

2 打开Windows shell

3 cd 到文件目录（cd D:\software）

4 pip 安装（pip install Twisted-17.1.0-cp35-cp35m-win\_amd64.whl）

5 验证（import twisted）

rpy2 安装过程：

1 安装R

2 配置环境

这是比较重要的一般，很多安装失败的原因就是没有很好配置环境变量。

1. 将R目录添加至path环境变量

path：D:\Program Files\Microsoft\MRO\R-3.2.3\bin\x64;

2. 将R.dll添加至path环境变量

path：D:\Program Files\Microsoft\MRO\R-3.2.3\bin\x64\R.dll;

3. 设置R\_HOME环境变量

R\_HOME：D:\Program Files\Microsoft\MRO\R-3.2.3

4. 设置R\_USER环境变量

R\_USER：yishikeji-01

R\_USER就是你电脑的用户名，如果没有设置过用户名就是Administrator，可以通过cmd命令行输入 net user查看。

3 下载非官方版包

（需与所使用的Python匹配，如anaconda中的Python是3.5，则下载rpy2‑2.8.5‑cp35‑cp35m‑win\_amd64.whl）

<http://www.lfd.uci.edu/~gohlke/pythonlibs/#rpy2>

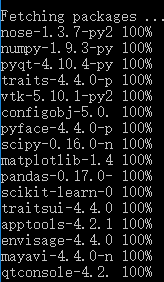
4 打开windows 的shell命令窗

cd D:\software\R （下载文件所在目录）

pip install rpy2-2.8.5-cp35-cp35m-win\_amd64.whl

5 import rpy2 验证

安装mayavi时，会同时安装的库包括：（见下图）



三、常用模块（python 自带）

import sys

import codecs

import csv

import statsmodels

import shutil

import copy

sys 系统相关

[codecs](http://san-yun.iteye.com/blog/1544123) [自然语言编码转换](http://san-yun.iteye.com/blog/1544123)

csv [csv文件](http://blog.csdn.net/lixiang0522/article/details/7755059)

statsmodels 统计建模和计量经济学工具包，包括一些描述统计、统计模型估计和推断

shutil 高级的文件，文件夹，压缩包处理模块：复制、移动、新建等……

copy 复制

三、机器学习相关常用库

<http://qxde01.blog.163.com/blog/static/67335744201368101922991/>

1 numpy

Numpy是一个用python实现的科学计算包。包括：

一个强大的N维数组对象Array；

比较成熟的（广播）函数库；

用于整合C/C++和Fortran代码的工具包；

实用的线性代数、傅里叶变换和随机数生成函数。

2 scipy

SciPy是一个开源的Python算法库和数学工具包，SciPy包含的模块有最优化、线性代数、积分、插值、特殊函数、快速傅里叶变换、信号处理和图像处理、常微分方程求解和其他科学与工程中常用的计算。其功能与软件MATLAB、Scilab和GNU Octave类似。

3 scikit-learn

scikit-learn 是一个基于SciPy和Numpy的开源机器学习模块，包括分类、回归、聚类系列算法，主要算法有SVM、逻辑回归、朴素贝叶斯、Kmeans、DBSCAN等，目前由INRI 资助，偶尔Google也资助一点。

项目主页：

<https://pypi.python.org/pypi/scikit-learn/>

<http://scikit-learn.org/>

<https://github.com/scikit-learn/scikit-learn>

4 NLTK

NLTK(Natural Language Toolkit)是Python的自然语言处理模块，包括一系列的字符处理和语言统计模型。

NLTK 常用于学术研究和教学，应用的领域有语言学、认知科学、人工智能、信息检索、机器学习等。 NLTK提供超过50个语料库和词典资源，文本处理库包括分类、分词、词干提取、解析、语义推理。可稳定运行在Windows, Mac OS X和Linux平台上.

项目主页：

<http://sourceforge.net/projects/nltk/>

<https://pypi.python.org/pypi/nltk/>

<http://nltk.org/>

5 Mlpy

Mlpy是基于NumPy/SciPy的Python机器学习模块，它是Cython的扩展应用。包含的机器学习算法有：

l  回归

[least squares](http://en.wikipedia.org/wiki/Least_squares), [ridge regression](http://en.wikipedia.org/wiki/Ridge_regression), least angle regression, [elastic net](http://en.wikipedia.org/wiki/Elastic_net_regularization), kernel ridge regression, [support vector machines](http://en.wikipedia.org/wiki/Support_vector_machines) (SVM), partial least squares (PLS)

l  分类

linear discriminant analysis (LDA), Basic perceptron, Elastic Net, logistic regression, (Kernel) Support Vector Machines (SVM), Diagonal Linear Discriminant Analysis (DLDA), Golub Classifier, Parzen-based, (kernel) Fisher Discriminant Classifier, k-nearest neighbor, Iterative RELIEF, Classification Tree, Maximum Likelihood Classifier

l  聚类

hierarchical clustering, Memory-saving Hierarchical Clustering, [k-means](http://en.wikipedia.org/wiki/K-means)

l  维度约减

(Kernel) [Fisher discriminant analysis](http://en.wikipedia.org/wiki/Fisher_discriminant_analysis) (FDA), Spectral Regression Discriminant Analysis (SRDA), (kernel) [Principal component analysis](http://en.wikipedia.org/wiki/Principal_component_analysis) (PCA)

项目主页：

<http://sourceforge.net/projects/mlpy>

<https://mlpy.fbk.eu/>

6 Shogun

Shogun是一个开源的大规模机器学习工具箱。

目前Shogun的机器学习功能分为几个部分：feature表示，feature预处理， 核函数表示,核函数标准化，距离表示，分类器表示，聚类方法，分布， 性能评价方法，回归方法，结构化输出学习器。

SHOGUN 的核心由C++实现，提供 Matlab、 R、 Octave、 Python接口。主要应用在linux平台上。

项目主页：

<http://www.shogun-toolbox.org/>

7 MDP

The Modular toolkit for Data Processing (MDP) ，用于数据处理的模块化工具包，一个Python数据处理框架。

从用户的观点，MDP是能够被整合到数据处理序列和更复杂的前馈网络结构的一批监督学习和非监督学习算法和其他数据处理单元。计算依照速度和内存需求而高效的执行。从科学开发者的观点，MDP是一个模块框架，它能够被容易地扩展。新算法的实现是容易且直观的。新实现的单元然后被自动地与程序库的其余部件进行整合。

MDP在神经科学的理论研究背景下被编写，但是它已经被设计为在使用可训练数据处理算法的任何情况中都是有用的。其站在用户一边的简单性，各种不同的随时可用的算法，及应用单元的可重用性，使得它也是一个有用的教学工具。

项目主页：

<http://mdp-toolkit.sourceforge.net/>

<https://pypi.python.org/pypi/MDP/>

8 PyBrain

PyBrain(Python-Based Reinforcement Learning, Artificial Intelligence and Neural Network)是Python的一个机器学习模块，它的目标是为机器学习任务提供灵活、易应、强大的机器学习算法。（这名字很霸气）

PyBrain正如其名，包括神经网络、强化学习(及二者结合)、无监督学习、进化算法。因为目前的许多问题需要处理连续态和行为空间，必须使用函数逼近(如神经网络)以应对高维数据。

PyBrain以神经网络为核心，所有的训练方法都以神经网络为一个实例。

项目主页：

<http://www.pybrain.org/>

<https://github.com/pybrain/pybrain/>

9 BigML

BigML 使得机器学习为数据驱动决策和预测变得容易，BigML使用容易理解的交互式操作创建优雅的预测模型。BigML使用BigML.io,捆绑Python。

项目主页：

<https://bigml.com/>

<https://pypi.python.org/pypi/bigml>

<http://bigml.readthedocs.org/>

10 PyML

PyML是一个Python机器学习工具包， 为各分类和回归方法提供灵活的架构。它主要提供特征选择、模型选择、组合分类器、分类评估等功能。

项目主页：

<http://cmgm.stanford.edu/~asab/pyml/tutorial/>

<http://pyml.sourceforge.net/>

11 Milk

Milk是Python的一个机器学习工具箱，其重点是提供监督分类法与几种有效的分类分析：SVMs(基于libsvm)，K-NN，随机森林经济和决策树。它还可以进行特征选择。这些分类可以在许多方面相结合，形成不同的分类系统。

对于无监督学习，它提供K-means和affinity propagation聚类算法。

项目主页：

<https://pypi.python.org/pypi/milk/>

<http://luispedro.org/software/milk>

12 PyMVPA

PyMVPA(Multivariate Pattern Analysis in Python)是为大数据集提供统计学习分析的Python工具包，它提供了一个灵活可扩展的框架。它提供的功能有分类、回归、特征选择、数据导入导出、可视化等

项目主页：

<http://www.pymvpa.org/>

<https://github.com/PyMVPA/PyMVPA>

13 Pattern

Pattern是Python的web挖掘模块，它绑定了  Google、Twitter 、Wikipedia API，提供网络爬虫、HTML解析功能，文本分析包括浅层规则解析、WordNet接口、句法与语义分析、TF-IDF、LSA等，还提供聚类、分类和图网络可视化的功能。

项目主页：

<http://www.clips.ua.ac.be/pages/pattern>

<https://pypi.python.org/pypi/Pattern>

14 pyrallel

Pyrallel(Parallel Data Analytics in Python)基于分布式计算模式的机器学习和半交互式的试验项目，可在小型集群上运行，适用范围：

l  focus on small to medium dataset that fits in memory on a small (10+ nodes) to medium cluster (100+ nodes).

l  focus on small to medium data (with data locality when possible).

l  focus on CPU bound tasks (e.g. training Random Forests) while trying to limit disk / network access to a minimum.

l  do not focus on HA / Fault Tolerance (yet).

l  do not try to invent new set of high level programming abstractions (yet): use a low level programming model (IPython.parallel) to finely control the cluster elements and messages transfered and help identify what are the practical underlying constraints in distributed machine learning setting.

项目主页：

https://pypi.python.org/pypi/pyrallel

<http://github.com/pydata/pyrallel>

15 Monte

Monte ( machine learning in pure Python)是一个纯Python机器学习库。它可以迅速构建神经网络、条件随机场、逻辑回归等模型，使用inline-C优化，极易使用和扩展。

项目主页：

<https://pypi.python.org/pypi/Monte>

[http://montepython.sourceforge.net](http://montepython.sourceforge.net/)

16 Orange

Orange 是一个基于组件的数据挖掘和机器学习软件套装，它的功能即友好，又很强大，快速而又多功能的可视化编程前端，以便浏览数据分析和可视化，基绑定了 Python以进行脚本开发。它包含了完整的一系列的组件以进行数据预处理，并提供了数据帐目，过渡，建模，模式评估和勘探的功能。其由C++ 和 Python开发，它的图形库是由跨平台的Qt框架开发。

项目主页：

<https://pypi.python.org/pypi/Orange/>

<http://orange.biolab.si/>

17 Theano

Theano 是一个 Python 库，用来定义、优化和模拟数学表达式计算，用于高效的解决多维数组的计算问题。Theano的特点：

l  紧密集成Numpy

l  高效的数据密集型GPU计算

l  高效的符号微分运算

l  高速和稳定的优化

l  动态生成c代码

l  广泛的单元测试和自我验证

自2007年以来，Theano已被广泛应用于科学运算。theano使得构建深度学习模型更加容易，可以快速实现下列模型：

l  Logistic Regression

l  Multilayer perceptron

l  Deep Convolutional Network

l  Auto Encoders, Denoising Autoencoders

l  Stacked Denoising Auto-Encoders

l  Restricted Boltzmann Machines

l  Deep Belief Networks

l  HMC Sampling

l  Contractive auto-encoders

Theano，一位希腊美女，Croton最有权势的Milo的女儿，后来成为了毕达哥拉斯的老婆。

项目主页：

<http://deeplearning.net/tutorial/>

<https://pypi.python.org/pypi/Theano>

18 Pylearn2

Pylearn2建立在theano上，部分依赖scikit-learn上，目前Pylearn2正处于开发中，将可以处理向量、图像、视频等数据，提供MLP、RBM、SDA等深度学习模型。Pylearn2的目标是：

Researchers add features as they need them. We avoid getting bogged down by too much top-down planning in advance.

A machine learning toolbox for easy scientific experimentation.

All models/algorithms published by the LISA lab should have reference implementations in Pylearn2.

Pylearn2 may wrap other libraries such as scikits.learn when this is practical

Pylearn2 differs from scikits.learn in that Pylearn2 aims to provide great flexibility and make it possible for a researcher to do almost anything, while scikits.learn aims to work as a “black box” that can produce good results even if the user does not understand the implementation

Dataset interface for vector, images, video, ...

Small framework for all what is needed for one normal MLP/RBM/SDA/Convolution experiments.

Easy reuse of sub-component of Pylearn2.

Using one sub-component of the library does not force you to use / learn to use all of the other sub-components if you choose not to.

Support cross-platform serialization of learned models.

Remain approachable enough to be used in the classroom (IFT6266 at the University of Montreal).

项目主页：

<http://deeplearning.net/software/pylearn2/>

<https://github.com/lisa-lab/pylearn2>

19 其他Python机器学习库

pmll(<https://github.com/pavlov99/pmll>)

[pymining](https://pypi.python.org/pypi/pymining/0.1)(<https://github.com/bartdag/pymining>)

ease (https://github.com/edx/ease)

textmining(<http://www.christianpeccei.com/textmining/>)