

Python

CSCI 347,
Data Mining

Python Libraries

- Pandas – data analysis library that also provides data structures
- Xldr – extracts data from an Excel spreadsheet
- Pyforms – provides GUI support (also terminal and web support)
- Numpy – math functions
- BeautifulSoup – xml and html parsing library
- SciPy – algorithms and mathematic tools
- matplotlib – a numerical plotting library
- Scikit-learn - efficient tools for machine learning and statistical modeling

Access to Libraries

Python uses lots of libraries:

- Libraries must be “installed” on the machine before they can be “imported” into a Python environment.
- pip :
 - Package management system
 - Stands for “Pip installs Packages” or “Pip installs Python”
 - Within MSPowerShell type:
 - > pip install libraryName
 - > pip3 list // Displays a list of installed libraries
// and their versions
- Within Python program type |
 - import module from libraryName

Jupyter Notebook

- Interactive computing environment
- Can create documents that include live code, interactive widgets, plots, narrative text, equations, images and video
- Complete and self-contained record of a computation that can be shared via email, Dropbox, Git, or nbviewer.jupyter.org
- Code options: Python (default), Julia, R, Ruby, Haskell, Scala, node.js or Go
- Edit and run code in a browser and see results
- Notebooks are stored in JSON format, allowing mixing formatted text, Python code and code output. It requires the IPython notebook server to run it.

Jupyter Notebook Installation

Requires Python 3.3 or greater

Directions:

<https://jupyter.readthedocs.io/en/latest/install.html>

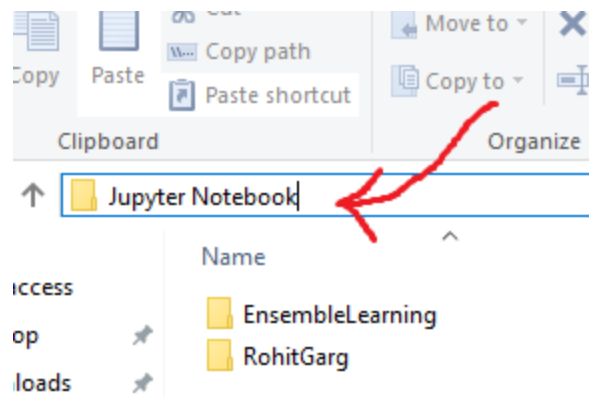
Jupyter Notebook - continued

Notebook documents:

- Contain inputs, outputs and narrative
- .ipynb extension (interactive python notebook)
- 3 cells types: code (input and output), markdown and raw (unformatted text) cells
- Run a cell via Shift-Enter

Jupyter Notebook Startup

When inside the folder, type “Jupyter Notebook in the address bar



Jupyter Notebook

Good tutorial for Python and Jupyter Notebook:

<https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/>