Env setup instructions for tutorial #470: “A beginner's introduction to Pydata: how to build a minimal recommendation engine”

There are a few options available to set up your Python environment for this tutorial, all of which will be outlined below. The easiest option is to download the EPDFree package, which includes Python, IPython and several libraries that we’ll use (Numpy, pandas, etc). The alternative is to install everything from scratch, if you feel more adventurous. Also, if you are familiar with virtualenv, you can just ‘pip install’ each library individually, but we will not cover that in this document.

Windows

The following instructions should apply to Windows 7 & 8, 32/64-bit. If you are working with Windows 8, it is recommended to use the “Desktop” mode as you’ll be using the command prompt quite often. It is recommended to run the Python extension installers via the command line. Also, ensure that you run the command prompt “as an administrator” to avoid permission related issues.

EPDFree method (32-bit version of Python only):

1. Select the first option (EPD Free - Windows) from the following link: http://epd-free.enthought.com/?Download=Download+EPD+Free+7.3-2

2. Double-click to run the download MSI. Select the desired path for the installation, as well as the Desktop/Quickbar shortcuts.

3. Once the installation is complete, you should see a “Pylab” icon in your Start Menu if in Windows 8, and an entry for “Enthought” with “Pylab” in Windows 7.

4. You should also double-check to ensure that Python appears in your PATH environment variable by opening up a command prompt and running: “echo %path%”. Ensure the “c:\Python27;c:\Python27\scripts” appear in the path (if you did change the default installation path, ensure that it appears in the PATH environment variable). If they do not appear in the path, then add it manually: Windows 7 - http://pythoncentral.org/how-to-install-python-2-7-on-windows-7-python-is-not-recognized-as-an-internal-or-external-command/, Windows 8 - http://install-climber.blogspot.ca/2013/02/howto-setup-path-environmentvariables-windows8.html

5. If you installed other versions of Python, be sure to delete any other Python-related directories from both the system and user Path variables. After making a path alteration, you have to restart the command prompt for the changes to take effect.

6 (Optional). In order to install pytables, you'll need to first install HDF5 and then the 'numexpr' library. To install HDF5, navigate to:
http://www.hdfgroup.org/HDF5/release/obtain5.html

and download 'HDF51810-win32.zip'. Unzip and run the installation executable (you may have to do this in 'Administrator' mode depending on your Windows user). Accept all the defaults. You may receive a warning about the PATH being too long, which should not be an issue for the purpose of this tutorial.

Then, download 'numexpr-2.0.1.win32-py2.7.exe' from http://www.lfd.uci.edu/~gohlke/pythonlibs/#numexpr and install the binary by either double-clicking the exe file or by running it from the command prompt.

Then, download 'tables-2.4.0.win32-py2.7.exe' from http://www.lfd.uci.edu/~gohlke/pythonlibs/#pytables and install the binary by either double-clicking the exe file or by running it from the command prompt.

7. Download 'pandas-0.10.1.win32-py2.7.exe' from http://www.lfd.uci.edu/~gohlke/pythonlibs/#pandas and install the binary by either double-clicking the exe file or by running it from the command prompt.

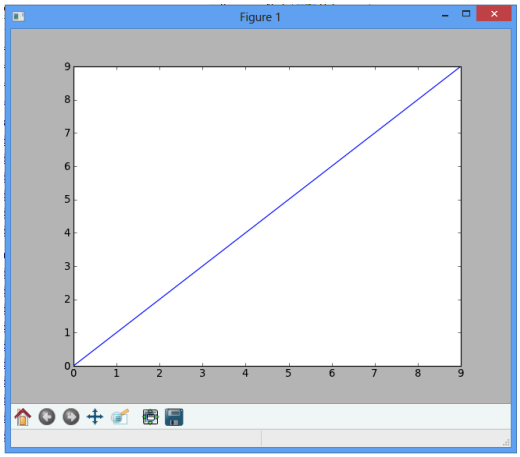
8. Download 'numpy-MKL-1.7.0.win32-py2.7.exe' from http://www.lfd.uci.edu/~gohlke/pythonlibs/#numpy and install the binary by either double-clicking the exe file or by running it from the command prompt.

9 (Optional). Download 'pip-1.2.1.win32-py2.7.exe' from http://www.lfd.uci.edu/~gohlke/pythonlibs/#pip, and install the binary by either double-clicking the exe file or by running it from the command prompt.

10. Your Python environment should now be ready. To test, either double-click on the "PyLab" icon on your desktop (from step 3), or run "ipython --pylab" in your command prompt. You should see the following:

Python 2.7.3 |EPD_free 7.3-2 (32-bit)|
Type "copyright", "credits" or "license" for more information.
IPython 0.13 -- An enhanced Interactive Python.
? -> Introduction and overview of IPython's features. %quickref -> Quick reference.
help -> Python's own help system.

object? -> Details about 'object', use 'object??' for extra details.
Welcome to pylab, a matplotlib-based Python environment [backend: WXAgg]. For more information, type 'help(pylab)'.
In [1]:
You can run a quick test by entering in the following commands:
In [1]: import pandas
In [2]: plot(arange(10))
This should result in the following graph in a new window:



Installing from scratch (32 & 64-bit):

1. Download Python 2.7.3 from the following link:
http://www.python.org/download/releases/2.7.3/

Windows x86 MSI - 32-bit
Windows X86-64 MSI Installer - 64 bit

2. Double-click to run the download MSI. Select the desired path for the installation. You can use the default installation settings, or de-select any options that are not required (such as documentation).

3. Add the following entries to your Windows PATH environment variable: "c:\Python27;c:\Python27\scripts". If you changed the default installation path from step 2, then ensure that the proper path is reflect in the PATH.
Windows 7 -
http://pythoncentral.org/how-to-install-python-2-7-on-windows-7-python-is-not-recognized-as-an-internal-or-external-command/,
Windows 8 -
http://install-climber.blogspot.ca/2013/02/howto-setup-path-environmentvariables-windows8.html

4. Download 'distribute-0.6.33.win32-py2.7.exe' (or the 64-bit version if you're using 64-bit python) from http://www.lfd.uci.edu/~gohlke/pythonlibs/#distribute and install the binary by either double-clicking the exe file or by running it from the command prompt.

5. Download 'ipython-0.13.1.win32-py2.7.exe' (or the 64-bit version if you're using 64-bit python) from http://www.lfd.uci.edu/~gohlke/pythonlibs/#ipython and install the binary by either double-clicking the exe file or by running it from the command prompt.

6. Download 'numpy-MKL-1.7.0.win32-py2.7.exe' (or the 64-bit version if you're using 64-bit python) from http://www.lfd.uci.edu/~gohlke/pythonlibs/#numpy and install the binary by either double-clicking the exe file or by running it from the command prompt.

7. Download 'scipy-0.11.0.win32-py2.7.exe' (or the 64-bit version if you're using 64-bit python) from http://www.lfd.uci.edu/~gohlke/pythonlibs/#scipy and install the binary by either double-clicking the exe file or by running it from the command prompt.

8. Download 'pandas-0.10.1.win32-py2.7.exe' (or the 64-bit version if you're using 64-bit python) from http://www.lfd.uci.edu/~gohlke/pythonlibs/#pandas and install the binary by either double-clicking the exe file or by running it from the command prompt.

9(Optional). In order to install pytables, you'll need to first install HDF5 and then the 'numexpr'

library. To install HDF5, navigate to:
http://www.hdfgroup.org/HDF5/release/obtain5.html

and download 'HDF51810-win32.zip'. Unzip and run the installation executable (you may have to do this in 'Administrator' mode depending on your Windows user). Accept all the defaults. You may receive a warning about the PATH being too long, which should not be an issue for the purpose of this tutorial.

Then, download 'numexpr-2.0.1.win32-py2.7.exe' from http://www.lfd.uci.edu/~gohlke/pythonlibs/#numexpr and install the binary by either double-clicking the exe file or by running it from the command prompt.

Then, download 'tables-2.4.0.win32-py2.7.exe' from http://www.lfd.uci.edu/~gohlke/pythonlibs/#pytables and install the binary by either double-clicking the exe file or by running it from the command prompt.

10. Download 'matplotlib-1.2.0.win32-py2.7.exe' (or the 64-bit version if you're using 64-bit python) from http://www.lfd.uci.edu/~gohlke/pythonlibs/#matplotlib and install the binary by either double-clicking the exe file or by running it from the command prompt.

11 (Optional). Download 'pip-1.2.1.win32-py2.7.exe' from http://www.lfd.uci.edu/~gohlke/pythonlibs/#pip, and install the binary by either double-clicking the exe file or by running it from the command prompt.

12. Your Python environment should now be ready. To test, either double-click on the "PyLab" icon on your desktop (from step 3), or run "ipython --pylab" in your command prompt. You should see the following:

You can run a quick test by entering in the following commands:

In [1]: import pandas

In [2]: from matplotlib.pyplot import plot

In [3]: import numpy as np

In [4]: plot(np.arange(10))
Out[4]: [<matplotlib.lines.Line2D at 0xa57126c>

Note: the graph window will not appear as the extension required to display it is not installed, but this is not an issue as we will not be using that extension during the tutorial.

Linux

The following instructions should apply to Ubuntu 12.x 32/64-bit.

EPDFree method:

1. Select the EPD Free - Linux 32/64-bit version from the following link:
http://epd-free.enthought.com/?Download=Download+EPD+Free+7.3-2

2. In terminal, navigate to your terminal and run:
ahiram@ubuntu:~/Downloads$ bash epd_free-7.3-2-rh5-x86.sh

3. Once you've accept the license, you'll be prompted to enter a default installation path. It's recommended to install it directly within your home folder. The default provided places it within the Download folder, so you may need to modify it:

EPD_free will be installed to this location:
/home/ahiram/Downloads/epd_free-7.3-2-rh5-x86

* *Press Enter to accept this location*
* *Press CTRL-C to abort*
* *or specify an alternate location. Please ensure that your location contains only ASCII letters, numbers and the following punctuation chars: '.', '_', '-'*

[/home/ahiram/Downloads/epd_free-7.3-2-rh5-x86] >>> /home/ahiram/epd

4. Once the installation is complete, you'll need to add the Enthought bin folder to you PATH environment variable. To do so, open up .bashrc in your home directory and the following:

export PATH=/home/ahiram/epd/bin:$PATH
(replacing the path with the location where you installed EPD)

Save the file and exit your editor.

5. To activate the change, run the following:
~$ source ~/.bashrc

6. Due to a missing library in the EPD-Free package, we also need to manually install libjpeg62. To do so, run:
~$ sudo apt-get install libjpeg62

7. Install ‘pandas’ using easy_install:
~$ easy_install pandas==0.10.1

8. Upgrade ‘numpy’ using easy_install:
~$ easy_install numpy==1.7.0

9. As a pre-requisite to pytables, install HDF5 by running:
~$ sudo apt-get install libhdf5-serial-dev

10(Optional). Install a couple libraries related to pytables, then pytables:
~$ easy_install numexpr==1.4.1
~$ easy_install cython==0.13
~$ easy_install tables==2.4.0

11. Your Python environment should now be ready. To test, run ‘ipython -- pylab’ in terminal:

~$ ipython --pylab
Enthought Python Distribution (free version) -- www.enthought.com
(type 'upgrade' or see www.enthought.com/epd/upgrade to get the full EPD)

Python 2.7.3 |EPD_free 7.3-2 (32-bit)| (default, Apr 11 2012, 18:02:54)
Type "copyright", "credits" or "license" for more information.

IPython 0.12.1 -- An enhanced Interactive Python.
?   -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help   -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.

Welcome to pylab, a matplotlib-based Python environment [backend: WXAgg].
For more information, type 'help(pylab)'.

In [1]:

You can run a quick test by entering in the following commands:

In [1]: import pandas

In [2]: plot(arange(10))

This should display a graph in a new window.

Installing from scratch (32 & 64-bit):

1. Ubuntu Linux comes with Python 2.7.3 pre-installed. The remaining steps cover the installation of the required python extensions.

2. Use the built-in package installed ‘apt’ to download the extensions. From terminal, run:
~$ sudo apt-get install ipython
~$ sudo apt-get install python-numpy
~$ sudo apt-get install python-scipy
~$ sudo apt-get install python-pandas
(Optional)~$ sudo apt-get install libhdf5-serial-dev
(Optional)~$ sudo apt-get install python-tables
~$ sudo apt-get install python-matplotlib

3. Your Python environment should now be ready. Your Python environment should now be ready. To test, run ‘ipython’ in terminal:

~$ ipython
Python 2.7.3 (default, Sep 26 2012, 21:53:58)
Type "copyright", "credits" or "license" for more information.

IPython 0.13 -- An enhanced Interactive Python.
? -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help -> Python's own help system.
object? -> Details about 'object', use 'object??' for extra details.

In [1]:

You can run a quick test by entering in the following commands:

In [1]: import pandas

In [2]: from matplotlib.pyplot import plot

In [3]: import numpy as np

In [4]: plot(np.arange(10))
Out[4]: [<matplotlib.lines.Line2D at 0xa57126c>]

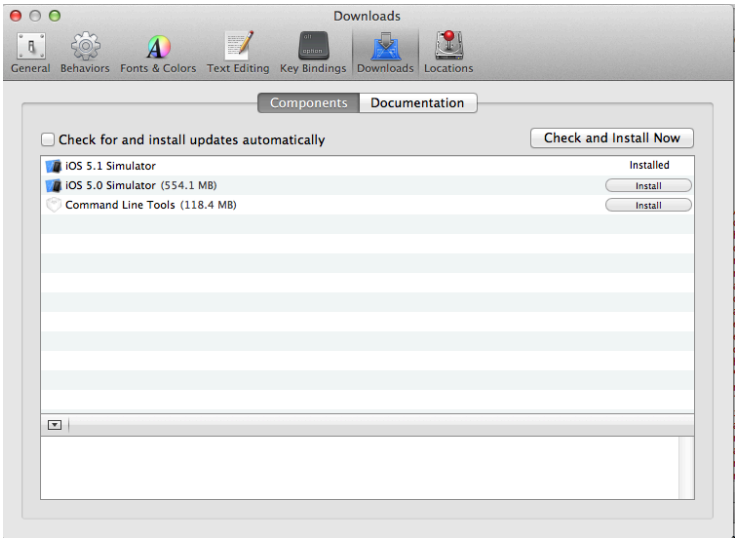
Note: the graph window will not appear as the extension required to display it is not installed, but this is not an issue as we will not be using that extension during the tutorial.

Mac

The following instructions should apply to Mac OSX.

Before proceeding with the instructions based on the options below, there are a couple of prerequisites: you must have Xcode installed with “Command Line Tools” also installed. To do this:

i. Install Xcode from the Mac App Store. You do not need to install any additional SDK’s.
ii. Once the Xcode installation is complete, install the “Command Line” tools by navigating to Xcode menu option -> Preferences -> Downloads -> Command Line Tools:



iii. Once “Command Line Tools” have been installed, then proceed to the next steps.

EPDFree method:

1. Select the EPD Free - Mac version from the following link:
http://epd-free.enthought.com/?Download=Download+EPD+Free+7.3-2

2. Double-click the .dmg file from your default download folder, then double-click the .mpkg file.

3. After the installation is complete, double-check the .bash_profile file located in your home directory located at /Users/<your user>/.bash_profile. You should see the following line:

```
# Setting PATH for EPD_free-7.3-1
PATH="/Library/Frameworks/Python.framework/Versions/Current/bin:${PATH}"
export PATH
```

If those lines do not appear in your .bash_profile, add them, save and exit. Then run 'source ~/.bash_profile' to activate your modification.

4. Install pandas by running the following command in your favourite terminal:
$ sudo easy_install pandas==0.10.1

5. Install numpy by running the following command in your favourite terminal:
$ sudo easy_install numpy==1.7.0

6(Optional). Before installing pytables, you'll need to install HDF5 as well as a couple of additional Python libraries. To install HDF5, go to:

http://www.hdfgroup.org/ftp/HDF5/current/src/

And download 'hdf5-1.8.10-patch1.tar.gz'. Uncompress the file and install by running:

$ tar -xvzf hdf5-1.8.10-patch1.tar.gz
$ cd hdf5-1.8.10-patch1
$ ./configure --prefix=/usr/local/
$ make
$ make install

Next, install numexpr/cython:
$ sudo easy_install numexpr
$ sudo easy_install cython

Finally, install pytables:

$ sudo easy_install tables==2.4.0

8. Your Python environment should now be ready. To test, run 'ipython -- pylab' in terminal:

~$ ipython --pylab
Enthought Python Distribution (free version) -- www.enthought.com
(type 'upgrade' or see www.enthought.com/epd/upgrade to get the full EPD)

Python 2.7.3 |EPD_free 7.3-2 (32-bit)| (default, Apr 11 2012, 18:02:54)
Type "copyright", "credits" or "license" for more information.

IPython 0.12.1 -- An enhanced Interactive Python.
? -> Introduction and overview of IPython's features.
%quickref -> Quick reference.

help   -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.

Welcome to pylab, a matplotlib-based Python environment [backend: WXAgg].
For more information, type 'help(pylab)'.

In [1]:

You can run a quick test by entering in the following commands:
In [1]: import pandas

In [2]: plot(arange(10))

This should display a graph in a new window.

Installing from scratch:

1. Python is bundled with Mac OSX, therefore you only need to install the extensions using 'easy_install'. To install the extensions, run the following commands in terminal:

$ sudo easy_install "ipython==0.13"
$ sudo easy_install "numpy==1.7.0"
$ sudo easy_install "scipy==0.11.0"
$ sudo easy_install "pandas==0.10.1"
$ sudo easy_install "matplotlib==1.1.0"

2(Optional). Before installing pytables, you'll need to install HDF5 as well as a couple of additional Python libraries. To install HDF5, go to:

http://www.hdfgroup.org/ftp/HDF5/current/src/

And download 'hdf5-1.8.10-patch1.tar.gz'. Uncompress the file and install by running:

$ tar -xvzf hdf5-1.8.10-patch1.tar.gz
$ cd hdf5-1.8.10-patch1
$ ./configure --prefix=/usr/local/
$ make
$ make install

Next, install numexpr/cython:
$ sudo easy_install numexpr
$ sudo easy_install cython

Finally, install pytables:

$ sudo easy_install tables==2.4.0

3. Your Python environment should now be ready. Your Python environment should now be ready. To test, run 'ipython' in terminal:

~$ ipython
Python 2.7.3 (default, Sep 26 2012, 21:53:58)
Type "copyright", "credits" or "license" for more information.

IPython 0.13 -- An enhanced Interactive Python.
?   -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help   -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.

In [1]:

You can run a quick test by entering in the following commands:

In [1]: import pandas

In [2]: from matplotlib.pyplot import plot

In [3]: import numpy as np

In [4]: plot(np.arange(10))
Out[4]: [<matplotlib.lines.Line2D at 0xa57126c>]