How to Setup a Workstation for Python-Tutorial

It can be difficult to install a Python machine learning environment on some platforms. Python itself must be installed first and then there are many packages to install, and it can be confusing for beginners. In this tutorial, you will learn how to setup a Python machine learning development environment using Anaconda.

After completing this tutorial, you will have a working Python environment to begin learning, practicing, and developing machine learning software. These instructions are suitable for Windows, Mac OS X, and Linux platforms. I will demonstrate them on OS X, so you may see some mac dialogs and file extensions.

Overview

In this tutorial, we will cover the following steps:

1. Download Anaconda
2. Install Anaconda
3. Start and Update Anaconda

Note: The specific versions may differ as the software and libraries are updated frequently.

Download Anaconda

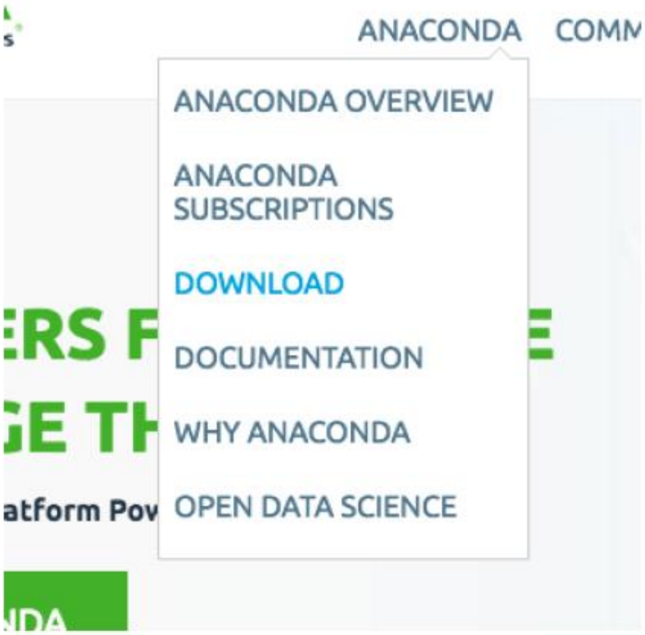
In this step, we will download the Anaconda Python package for your platform. Anaconda is a free and easy-to-use environment for scientific Python.

1. Visit the Anaconda homepage.

   https://www.continuum.io/

2. Click Anaconda from the menu and click Download to go to the download page.

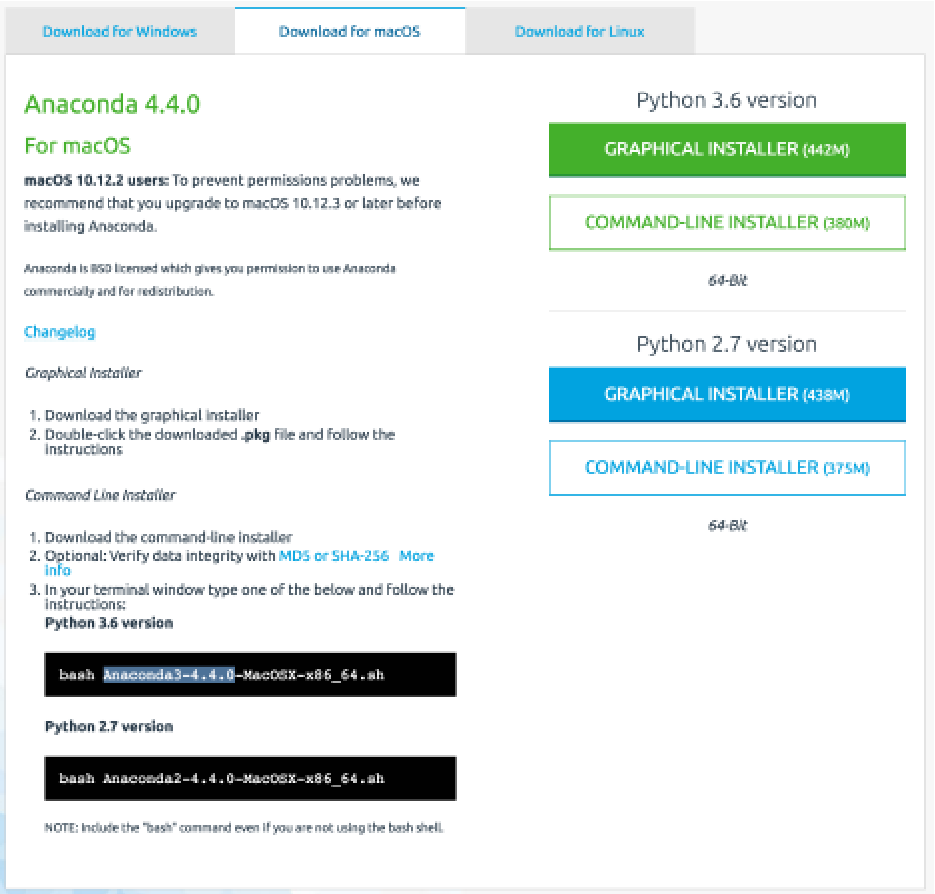
   https://www.continuum.io/downloads



3. Choose the download suitable for your platform (Windows, OSX, or Linux):

-Choose Python 3.6

-Choose the Graphical Installer


Download for Windows
Download for macOS
Download for Linux
Python 3.6 version
GRAPHICAL INSTALLER (442M)
COMMAND-LINE INSTALLER (380M)
64-Bit
Python 2.7 version
GRAPHICAL INSTALLER (438M)
COMMAND-LINE INSTALLER (375M)
64-Bit

Anaconda 4.4.0

For macOS

macOS 10.12.2 users: To prevent permissions problems, we recommend that you upgrade to macOS 10.12.3 or later before installing Anaconda.

Anaconda is BSD licensed which gives you permission to use Anaconda commercially and for redistribution.

Changelog

Graphical Installer

1. Download the graphical installer
2. Double-click the downloaded **.pkg** file and follow the instructions

Command Line Installer

1. Download the command-line installer
2. Optional: Verify data integrity with MD5 or SHA-256 More info
3. In your terminal window type one of the below and follow the instructions:

Python 3.6 version

```
bash Anaconda3-4.4.0-MacOSX-x86_64.sh
```

Python 2.7 version

```
bash Anaconda2-4.4.0-MacOSX-x86_64.sh
```

NOTE: Include the "bash" command even if you are not using the bash shell.

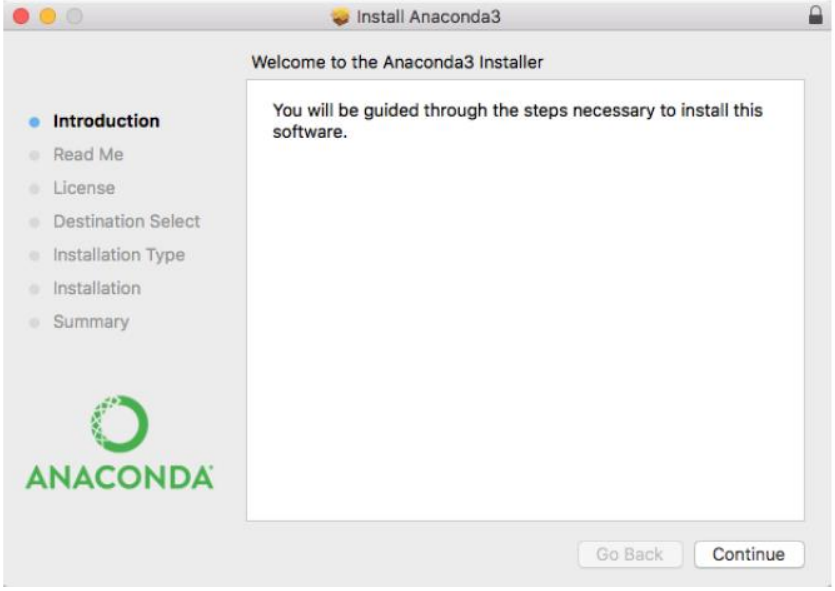
This will download the Anaconda Python package to your workstation. I'm on OS X, so I chose the OS X version. The file is about 426 MB. You should have a file with a name like:

Anaconda3-4.4.0-MacOSX-x86_64.pkg

Install Anaconda

In this step, we will install the Anaconda Python software on your system. This step assumes you have sufficient administrative privileges to install software on your system.

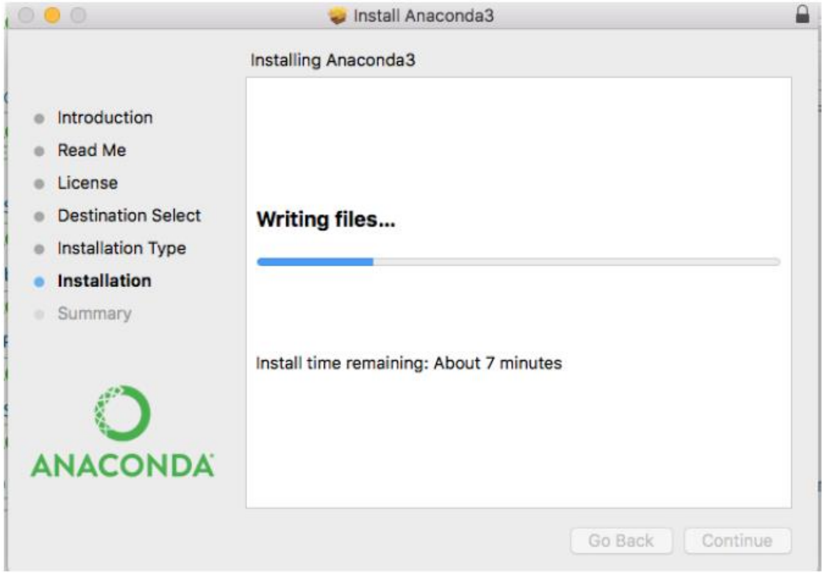
1. Double click the downloaded file.
2. Follow the installation wizard.



Welcome to the Anaconda3 Installer

You will be guided through the steps necessary to install this software.

Installation is quick and painless. There should be no tricky questions or sticking points.



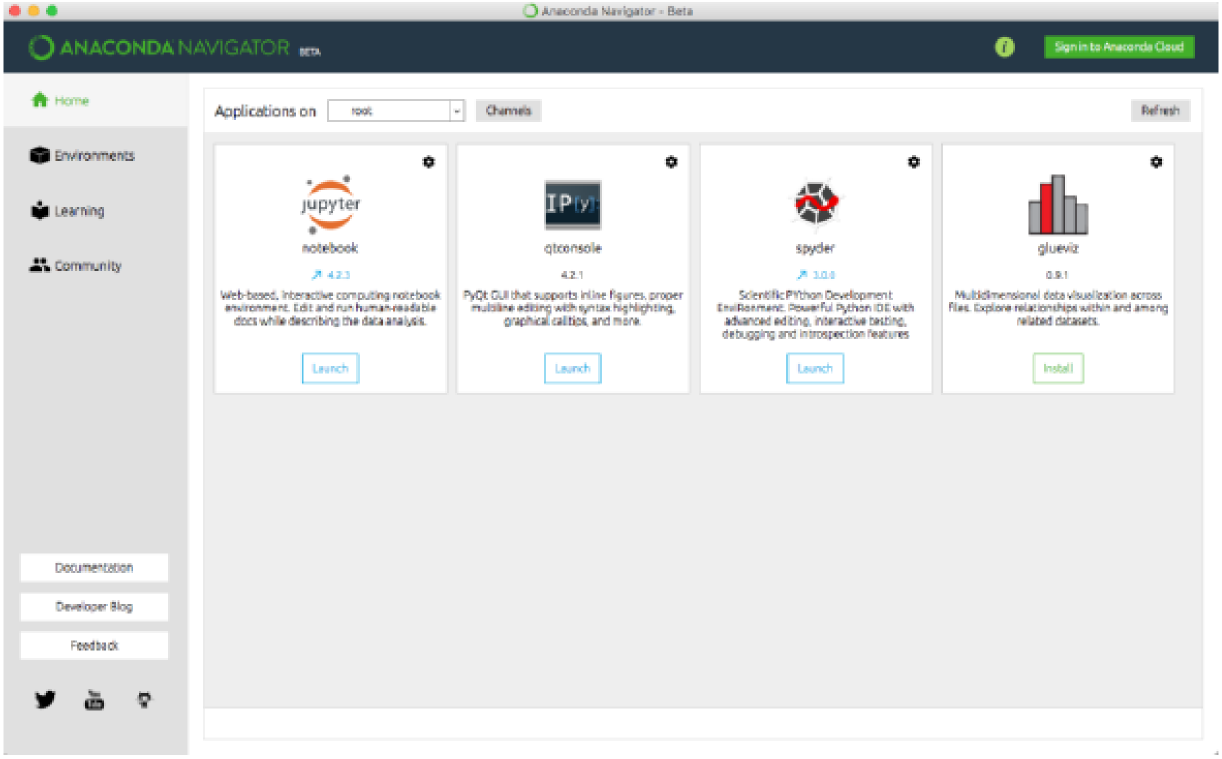
Writing files...

Install time remaining: About 7 minutes

The installation should take less than 10 minutes and take up a little more than 1 GB of space on your hard drive.

Start and Update Anaconda

In this step, we will confirm that your Anaconda Python environment is up to date. Anaconda comes with a suite of graphical tools called Anaconda Navigator. You can start Anaconda Navigator by opening it from your application launcher.



Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.

Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features

Multidimensional data visualization across files. Explore relationships within and among related datasets.

You can use the Anaconda Navigator and graphical development environments later; for now, I recommend starting with the Anaconda command line environment called conda. Conda is fast, simple, it's hard for error messages to hide, and you can quickly confirm your environment is installed and working correctly.

1. Open a terminal (command line window).
2. Confirm conda is installed correctly, by typing:

```
conda -V
```

You should see the following (or something similar):

```
conda 4.3.21
```

3. Confirm Python is installed correctly by typing:

```
python –V
```

You should see the following (or something similar):

```
Python 3.6.1 :: Anaconda 4.4.0 (x86_64)
```

If the commands do not work or have an error, please check the documentation for help for your platform.

4. Confirm your conda environment is up-to-date, type:

```
conda update conda
conda update anaconda
```

You may need to install some packages and confirm the updates.

5. Confirm your SciPy environment.

The script below will print the version number of the key SciPy libraries you require for machine learning development, specifically: SciPy, NumPy, Matplotlib, Pandas, Statsmodels, and Scikit-learn. You can type python and type the commands in directly. Alternatively, I recommend opening a text editor and copy-pasting the script into your editor.

```
# scipy
import scipy
print('scipy: %s' % scipy._ _version_ _)
# numpy
import numpy
print('numpy: %s' % numpy._ _version_ _)
# matplotlib
import matplotlib
print('matplotlib: %s' % matplotlib._ _version_ _)
# pandas
import pandas
print('pandas: %s' % pandas._ _version_ _)
# statsmodels
import statsmodels
print('statsmodels: %s' % statsmodels._ _version_ _)
# scikit-learn
import sklearn
print('sklearn: %s' % sklearn._ _version_ _)
```

Save the script as a file with the name: versions.py. On the command line, change your directory to where you saved the script and type:

```
python versions.py
```

You should see output like the following:

```
scipy: 1.0.0
numpy: 1.14.0
matplotlib: 2.1.1
pandas: 0.22.0
statsmodels: 0.8.0
sklearn: 0.19.1
```

Congratulations, you now have a working Python development environment for machine learning. You can now learn and practice machine learning on your workstation.

Python Ecosystem for Machine Learning

Python is a general purpose interpreted programming language. It is easy to learn and use primarily because the language focuses on readability. The philosophy of Python is captured in the Zen of Python which includes phrases like:

Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.

It is a popular language in general, consistently appearing in the top 10 programming languages in surveys on StackOverow[1]. It's a dynamic language and very suited to interactive development and quick prototyping with the power to support the development of large applications. It is also widely used for machine learning and data science because of the excellent library support and because it is a general purpose programming language (unlike R or Matlab). For example, see the results of the Kaggle platform survey results in 2011[2] and the KDD Nuggets 2015 tool survey results[3]. This is a simple and very important consideration. It means that you can perform your research and development (figuring out what models to use) in the same programming language that you use for your production systems. Greatly simplifying the transition from development to production.

SciPy

SciPy is an ecosystem of Python libraries for mathematics, science and engineering. It is an add-on to Python that you will need for machine learning. The SciPy ecosystem is comprised of the following core modules relevant to machine learning:

- NumPy: A foundation for SciPy that allows you to efficiently work with data in arrays.
- Matplotlib: Allows you to create 2D charts and plots from data.
- Pandas: Tools and data structures to organize and analyze your data.

To be effective at machine learning in Python you must install and become familiar with SciPy. Specifically:

- You will prepare your data as NumPy arrays for modeling in machine learning algorithms.
- You will use Matplotlib (and wrappers of Matplotlib in other frameworks) to create plots and charts of your data.
- You will use Pandas to load explore and better understand your data.

scikit-learn

The scikit-learn library is how you can develop and practice machine learning in Python. It is built upon and requires the SciPy ecosystem. The name scikit suggests that it is a SciPy plug-in or toolkit. The focus of the library is machine learning algorithms for classification, regression, clustering and more. It also provides tools for related tasks such as evaluating models, tuning parameters and pre-processing data. Like Python and SciPy, scikit-learn is open source and is usable commercially under the BSD license. This means that you can learn about machine learning, develop models and put them into operations all with the same ecosystem and code. A powerful reason to use scikit-learn.

[1] http://stackoverflow.com/research/developer-survey-2015
[2] http://blog.kaggle.com/2011/11/27/kagglers-favorite-tools/
[3] http://www.kdnuggets.com/polls/2015/analytics-data-mining-data-science-software-used.html

Python Ecosystem Installation

There are multiple ways to install the Python ecosystem for machine learning. In this section we cover how to install the Python ecosystem for machine learning.

How To Install Python

The first step is to install Python. I prefer to use and recommend Python 3.6. The instructions for installing Python will be specific to your platform. For instructions see Downloading Python[4] in the Python Beginners Guide. Once installed you can confirm the installation was successful.

Open a command line and type: *python –version*
You should see a response like the following: *Python 3.6.3*

How To Install SciPy

There are many ways to install SciPy. For example two popular ways are to use package management on your platform (e.g. yum on RedHat or macports on OS X) or use a Python package management tool like pip. The SciPy documentation is excellent and covers how-to instructions for many different platforms on the page Installing the SciPy Stack[5]. When installing SciPy, ensure that you install the following packages as a minimum:

- scipy
- numpy
- matplotlib
- pandas

Once installed, you can confirm that the installation was successful. Open the Python interactive environment by typing python at the command line, then type in and run the following Python code to print the versions of the installed libraries.

```
# scipy
import scipy
print('scipy: %s' % scipy._ _version_ _)
# numpy
import numpy
print('numpy: %s' % numpy._ _version_ _)
# matplotlib
import matplotlib
print('matplotlib: %s' % matplotlib._ _version_ _)
# pandas
import pandas
print('pandas: %s' % pandas._ _version_ _)
```

You may see the following output (or different versions)

```
scipy: 1.0.0
numpy: 1.13.3
matplotlib: 2.1.0
pandas: 0.21.0
```

How To Install scikit-learn

I would suggest that you use the same method to install scikit-learn as you used to install SciPy.

[4] https://wiki.python.org/moin/BeginnersGuide/Download
[5] http://scipy.org/install.html

There are instructions for installing scikit-learn[6], but they are limited to using the Python *pip* and *conda* package managers. Like SciPy, you can confirm that scikit-learn was installed successfully. Start your Python interactive environment and type and run the following code.

```
# scikit-learn
import sklearn
print('sklearn: %s' % sklearn._ _version_ _)
```

It will print the version of the scikit-learn library installed which maybe something like:

sklearn: 0.19.1

How To Install The Ecosystem: An Easier Way

If you are not confident at installing software on your machine, there is an easier option for you. There is a distribution called Anaconda that you can download and install for free[7]. It supports the three main platforms of Microsoft Windows, Mac OS X and Linux. It includes Python, SciPy and scikit-learn. Everything you need to learn, practice and use machine learning with the Python Environment.

Installing Notebook Jupyter

https://365datascience.com/how-to-install-python/

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

https://medium.com/@neuralnets/beginners-quick-guide-for-handling-issues-launching-jupyter-notebook-for-python-using-anaconda-8be3d57a209b

https://medium.com/@margaretmz/anaconda-jupyter-notebook-tensorflow-and-keras-b91f381405f8

https://docs.anaconda.com/ae-notebooks/user-guide/basic-tasks/apps/jupyter/

[6] http://scikit-learn.org/stable/install.html

[7] https://www.continuum.io/downloads