Python Anaconda distribution, PyCharm IDE, Conda & PyPI Packages and Theano Installation Guides

Machine Learning (41204)

Robert E. McCulloch / Mladen Kolar
Vinh Luong / Juan Yrigoyen

Fall 2015

IMPORTANT: Computer Hardware and Operating System Assumptions

In the following installation guides, we assume that your machine runs on either **Mac** or **Windows**.

If you happen to use Linux or another operating system, certain setups can get hairier and there may be a need to consult Dr. Google ad Prof. Stack Overflow heavily.

You may contact our kind-hearted Teaching Assistants for individualized assistance, although there is no guarantee we can solve non-Mac, non-Windows problems.

Python Installation

We recommend Continuum Analytics' Anaconda distribution of Python **v2.7** and compatible popular scientific computation packages.

Go **here** to download a graphical installer appropriate for your machine's

- processor (32-bit / 64-bit) and
- operating system (Mac / Windows).

Run the installer and follow its instructions to install the software into a folder you prefer.

Note that the Anaconda installation folder path MUST NOT CONTAIN BLANK SPACES, because spaces occasionally cause bugs in dependent open-source software packages.

Anaconda Academic License

Continuum offers free Anaconda advanced computation optimization add-ons for academic use

- ▶ You may obtain an Academic License **here** and follow the instructions in the company's email after you submit your request

Among all things, this Academic License gives you a free upgrade of the default BLAS (Basic Linear Algebra Subprograms) libraries of Anaconda Python to Intel's Math Kernel Library (MKL), which is among the fastest BLAS'es around

- ▶ This enables faster math, saving time for you and your machine – and salvaging a bit more of your machine's resale value :)
- ▶ The commands to install MKL are already included in the pre-included installation script file discussed in later slide **Additional Conda & PyPI Packages Installation**

JetBrains PyCharm IDE Installation *[Recommended but Optional]*

JetBrains, a vendor of popular integrated development environments (IDEs) for professional software developers, has an excellent IDE named PyCharm for Python.

We recommend PyCharm, but you may feel free to go with any ther Python IDE(s) of your own choice.

If you do opt for PyCharm, go **here** to download its free Community Edition and install it onto your machine.

Note that you typically need to manually configure your installed IDE (PyCharm or other) to link it to the a Python backend ("Python interpreter"). There should be easy-to-follow online guides on how to do this for each IDE.

Additional Conda & PyPI Packages Installation

We need to install a number of packages from both Continuum Analytics' Anaconda and the Python Package Index (PyPI).

After you have cloned and synchronized the **course GitHub repository** onto your machine, you may find script file named **PythonInstallPackages.sh** under the **[Your Course Repository Folder]/Software Guides/Python** folder.

Mac:

- ▶ Open your iTerm terminal via the **iTerm app**
- ▶ Navigate to the folder mentioned above
- ▶ Run the folllowing command: **./PythonInstallPackages.sh**

Windows:

- ▶ Open your **CygWin** terminal
- ▶ Navigate to the folder mentioned above
- ▶ Run the folllowing command: **sh PythonInstallPackages.sh**

Theano: Overview

Theano is a Python library that enables using a compatible GPU (Graphical Processing Unit) of the computer for numerical computation, which is far superior in performance terms to computation by the computer's CPU (Central Processing Unit).

- **Currently, the most, if not only, effectively Theano-supported GPUs are those by NVIDIA**
- Even if your machine do not have an NVIDIA graphics card, Theano is still useful in compiling complex calculations down to highly efficient C / C++ or machine code that can execute very fast

Theano: Overview (cont'd.)

Because of its huge performance-boosting benefits, Theano is prominent in extremely data-intensive, large-scale Machine Learning applications such as those in cutting-edge Neural Networks-based Deep Learning.

Having said that, Theano is still very young *(read: buggy)*, and its setup, configuration and usage demand a high degree of risk-taking and perseverance. :(*(It helps if you are good with your friends and hence enjoy good karma)*

We will hence use Theano only on a **very limited, experimental basis** for some exercises in Neural Networks, and we will install it in such a way that if your computer does not have an NVIDIA GPU, Theano will fall back to NumPy, which is Python's default numerical computation library.

In case you are unsuccessful in struggling through the Theano installation, we will use a non-Theano-dependent library for the Neural Networks exercises.

Theano Installation and Configuration on Mac

You may refer to the Theano installation and configuration guide for Mac **here**.

Appropriate steps vary from machine to machine. Consult Dr. Google and Prof. Stack Overflow whenever you get stuck.

In the Teaching Assistants' own experience, the recommended key steps include those detailed in the next slide.

Theano Installation and Configuration on Mac: Key Steps

1. Install the **Anaconda Python v2.7 distribution** (this should be already done, per instructions in previous slides)
2. Install the **additional Conda & PyPI packages** per instructions in previous slides
3. Install the **Clang compiler** through installing the **XCode app** from the Apple App Store and **running XCode once** to install the command-line tools
4. Install the **CUDA graphics driver and toolkit v7** from NDIVIA in order to use the GPU to perform numerical computation
5. Copy the **.theanorc file** in the **[Your Course Repository Folder]/Software Guides/Python/Theano/Mac** folder to the **/[Your Username]/** home folder, customize the **copied file** (NOT the original file) according to the comments, and save it

Theano Installation and Configuration on Mac (cont'd.)

One common source of error in setting up Theano on Mac concerns the BLAS (Basic Linear Algebra Subprograms) libraries

- Refer **here** for some troubleshooting guidance on this issue

Verification steps:

- Run the **TheanoTestScript.py** file in the ***[Your Course Repository Folder\/Software Guides/Python/Theano*** folder, either via your Python IDE or through the terminal, and verify that it completes successfully, giving your a comparison of speeds between NumPy and Theano

Theano Installation and Configuration on Windows

You may refer to the Theano installation and configuration guide for Windows **here**.

Appropriate steps vary from machine to machine. Consult Dr. Google and Prof. Stack Overflow whenever you get stuck.

In the Teaching Assistants' own experience, the recommended key steps include those detailed in the next slide.

Theano Installation and Configuration on Windows: Key Steps

1. Install the **Anaconda Python v2.7 distribution** (this should be already done, per instructions in previous slides)
2. Install the **additional Conda & PyPI packages** per instructions in previous slides
3. Install **Visual Studio Community 2013**
   - do **NOT** install the 2015 version, which is not compatible with the CUDA software below)
4. Install the **CUDA graphics driver and toolkit v7** from NDIVIA in order to use the GPU to perform numerical computation
5. Install the **Microsoft Visual C++ Compiler for Python v2.7**
6. Install a **GNU C Compiler (GCC)** such as **this**

Theano Installation and Configuration on Windows: Key Steps (cont'd.)

7. Install **GraphViz** into a *folder whose path has no blank spaces*
8. Open the **pydot.py** file in the *[Your Anaconda Installation Folder]/lib/site-packages* folder. Locate the **_find_graphviz** function, comment its whole content out and replace with the following: **return _find_executables('[Your GraphViz Installation Folder]/bin')**
9. Copy the **.theanorc** file under *[Your Course Repository Folder]/Software Guides/Python/Theano/Windows* folder to *C:/Users/[Your Windows User Name]/.theano/* folder, customize the **copied file** (NOT the original file) according to the comments, and save it

Theano Installation and Configuration on Windows (cont'd.)

Verification steps: open your CygWin terminal and:

- ▶ run **where gcc** and verify that the path to your GCC installation's gcc.exe file appears
- ▶ run **where nvcc** and verify that the path to your CUDA installation's nvcc.exe file appears
- ▶ run the **TheanoTestScript.py** file under ***[Your Course Repository Folder\/Software Guides/Python/Theano***, either via your Python IDE or through the terminal, and verify that it completes successfully, giving your a comparison of speeds between NumPy and Theano

Theano Configurations *[for info only]*

Theano has a large number of configuration levers, the most common of which are discussed **here**.

You do not need to pay attention to those for this course.

Theano Tutorials *[for info only]*

DeepLearning.net offers a comprehensive series of tutorials on Theano **here**.

Again, these are not required for this course. You can look at these at your own time if you are keen.