

# Machine Learning Installation Cheat Sheet

**1)Installation:** Please make sure you have the following Installed.

**Operating System:** I prefer Linux (any flavor) is good. However these instructions should work with some tweaking for Windows as well. If you do not have access to a Linux Machine, you can also use VirtualBox and spin off a Ubuntu VM. There are plenty of tutorials on Youtube that will get you there.

<https://www.youtube.com/watch?v=ncA85gRAJxk>

**Python /Anaconda:** We will be using Python all along in the assignments/exercises . Those if you not familiar with Python , there are plenty of tutorials on the internet that will help you get started. Python is easy to learn and Python has full fledged libraries to support every conceivable Machine Learning models that are out there. Besides Python is very useful for rapid development and quick prototyping and the best part is that it is very easy to learn.

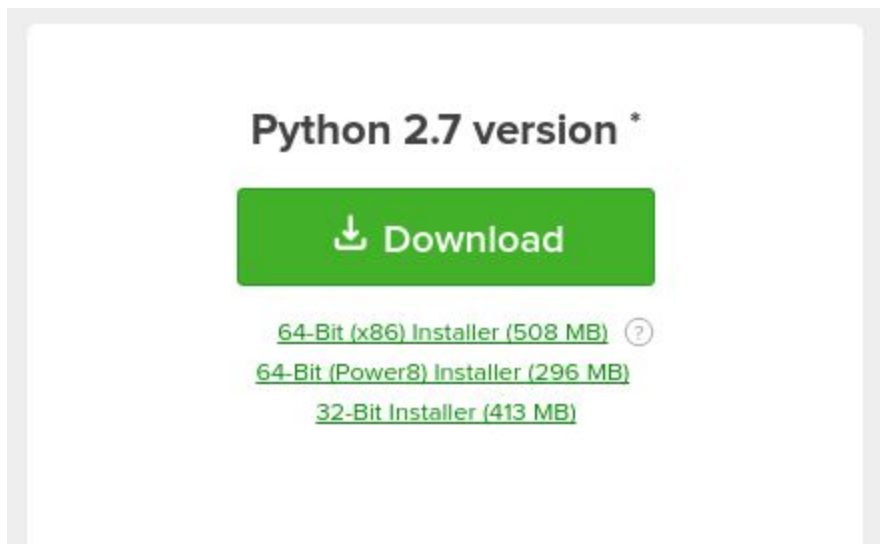
Here is a good Python Tutorial that will get you started.

<https://www.tutorialspoint.com/python/>

For our tutorials we will be using Python with Anaconda Packages . Please make sure you install Anaconda with Python 2.7 (not 3+ or above). The reason we are using Python 2.7 is that the GraphLab package which we will install subsequently works only on Python 2.7. **DO NOT** install Python 3.5 or any of the 3.0 series.

Here is the URL.

<https://www.anaconda.com/download/#linux>



If you already have a Python installation on your machine , I suggest you remove/uninstall it and install Python from Anaconda.

When you install Anaconda , it installs a bunch a Python packages that will be useful for Machine Learning , which you need to install explicitly if you choose other Python installations. DURING installation you will be prompted for the path to install (ex: /home/rbhupathi/anaconda2 ) Towards the end of the installation , you will be prompted to update your BASH file (equivalent one in Windows). Please choose to update. This will set the correct PATH settings for your environment.

Once your installation is done , go to the installation directory and type “conda list” . This will list all the packages that have been installed. If it is not able to find “conda”.. ,it means that the PATH variable has not been set correctly. Please fix it before proceeding further.

### Install GraphLab

We will be using GraphLab package from Turi. This is the company that developed this package and this company has since been acquired by Apple. To use this GraphLab , you need to register and then get an **ACCESS** code or the product-key that will be valid for an Year or so. Register yourself with your email address and make sure you get the access code. To fill up the attributes just make up something logical.

<https://turi.com/download/academic.html>

Next access this URL to download the GraphLab package . Note that we will not be using the Turi site to get the package. Instead we will get it from the regular Python Library repository.

<https://pypi.python.org/pypi/GraphLab-Create>

For Linux choose the \*.gz and Windows choose .whl

The next step now is to install it. To install the package we will be using “pip” , which comes standard with Anaconda. Go to the directory where your Anaconda is installed (ex:/home/rbhupathi/anaconda2) and type

**pip install GraphLab\_Create-2.1-py2.7.tar.gz** (for Linux)

**pip install [GraphLab\\_Create-2.1-cp27-none-win\\_amd64.whl](#)** (for Windows)

Assuming no errors , your installing of the package should be complete.  
Now we need to activate the package that has been installed.

To Activate (THIS IS ONE TIME ACTIVATION ONLY)  
Run python and then

```
import graphlab as gl
gl.product_key.set_product_key(<your-product-key>)
```

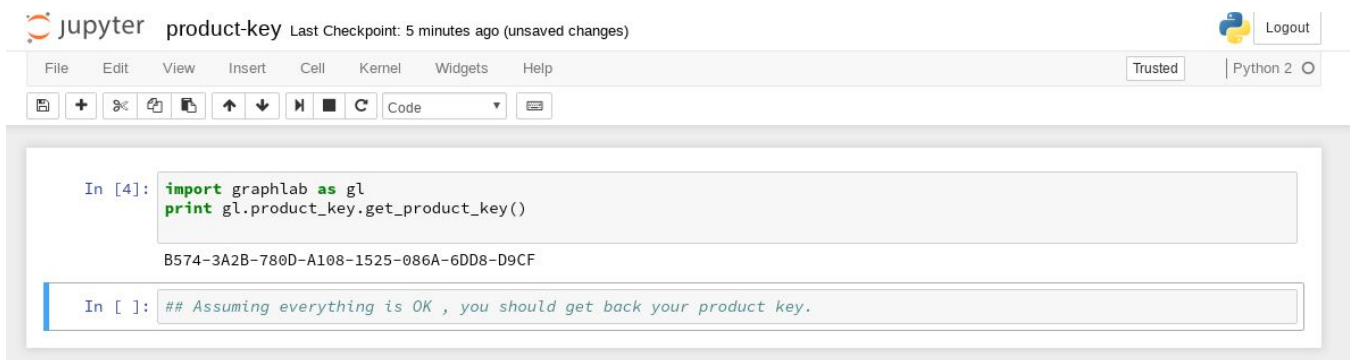
If this throws any error try putting the product-key in quotes (like strings)

This completes your installation .

Now go to your Anaconda2 installation directory and run the following:

*jupyter notebook*

It opens a page in your default browser . Type the following and press <Shift+Return> which executes the python code in the cell block. Just returns the key that you previously set.



**GitHub Account:** Access <https://github.com/> and create an account . Also install git. Git is a source code control system that comes very handy. I will be posting all my code on my github account and you will need git installed locally. This will help you pull in the latest exercise code that I post on to my account.

The account that you create will be used to safeguard your own assignments and tutorials.

All the tutorial information and exercises will be available at  
<https://github.com/ramakris/gentfg-machine-learning.git>

So to get all the assignments

Run “git clone <https://github.com/ramakris/gentfg-machine-learning.git>” on a Terminal window of your machine.

Then change directory to gentfg-machine-learning

## Other Useful Tools

**Git Cheat Sheet :** <https://www.git-tower.com/blog/git-cheat-sheet/>

**Jupyter Notebook Cheat Sheet :**

[https://s3.amazonaws.com/assets.datacamp.com/blog\\_assets/Jupyter\\_Notebook\\_Cheat\\_Sheet.pdf](https://s3.amazonaws.com/assets.datacamp.com/blog_assets/Jupyter_Notebook_Cheat_Sheet.pdf)

**DataScience CheatSheet :**

[https://s3.amazonaws.com/assets.datacamp.com/blog\\_assets/PythonForDataScience.pdf](https://s3.amazonaws.com/assets.datacamp.com/blog_assets/PythonForDataScience.pdf)

**Python CheatSheet :**

[https://github.com/ehmatthes/pcc/releases/download/v1.0.0/beginners\\_python\\_cheat\\_sheet\\_pcc\\_all.pdf](https://github.com/ehmatthes/pcc/releases/download/v1.0.0/beginners_python_cheat_sheet_pcc_all.pdf)

**Numpy CheatSheet:**

[https://s3.amazonaws.com/assets.datacamp.com/blog\\_assets/Numpy\\_Python\\_Cheat\\_Sheet.pdf](https://s3.amazonaws.com/assets.datacamp.com/blog_assets/Numpy_Python_Cheat_Sheet.pdf)

[https://www.dataquest.io/blog/large\\_files/numpy-cheat-sheet.pdf](https://www.dataquest.io/blog/large_files/numpy-cheat-sheet.pdf)

**Scikit Learn**

This is a great tutorial on the various Machine Learning Models

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

## **Sessions and Agenda**

**Session 1:** Introduction to ML, installation, setup.

**Session 2:** Linear Regression , Multivariant Regression , Predicting House Prices

**Session 3:** Classification : Analyzing Sentiment

**Session 4:** Clustering and Similarities

**Session 5:** Recommending Products

**Session 6:** Linear Classifiers and Logistic Regression

**Session 7:** -- Same as Above Continued --

**Session 8:** Regulations Techniques to deal with Overfitting(Lasso etc)