Aadith Lasar

LY CORE-2

2203262

MACHINE LEARNING LAB-1

Title: A. Installation and Configuration of machine learning environment with Anaconda on windows or Ubuntu (Jupyter notebook)

B. Installation, Confguration and checking the current Version of numpy, scipy, scikit, pandas and matplot lib using anaconda prompt and list their uses in machine learning.

After completion of this experiment students will be able to:

- Use Anaconda environment for the development of machine learning application.

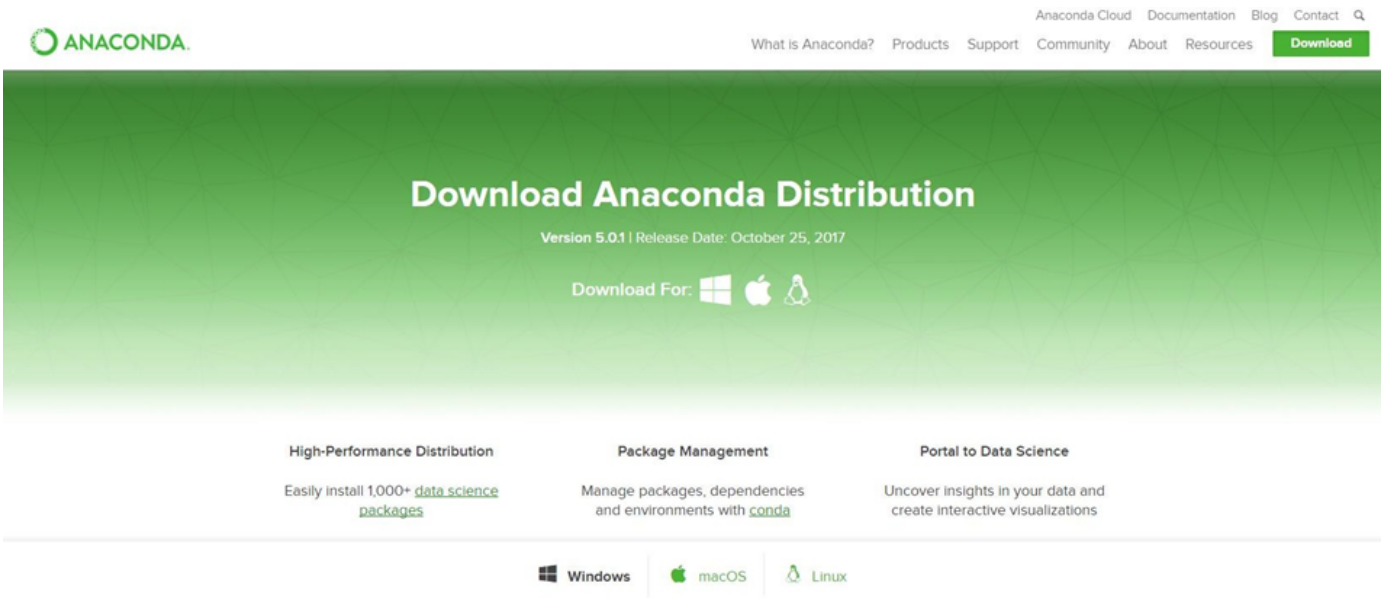
Aim: perform Installation and Configuration of machine learning environment with Anaconda on windows or Ubuntu (Jupyter notebook)

Theory:

Installation Of Anaconda in Windows

1. Visit the Anaconda downloads page

Go to the following link: Anaconda.com/downloads

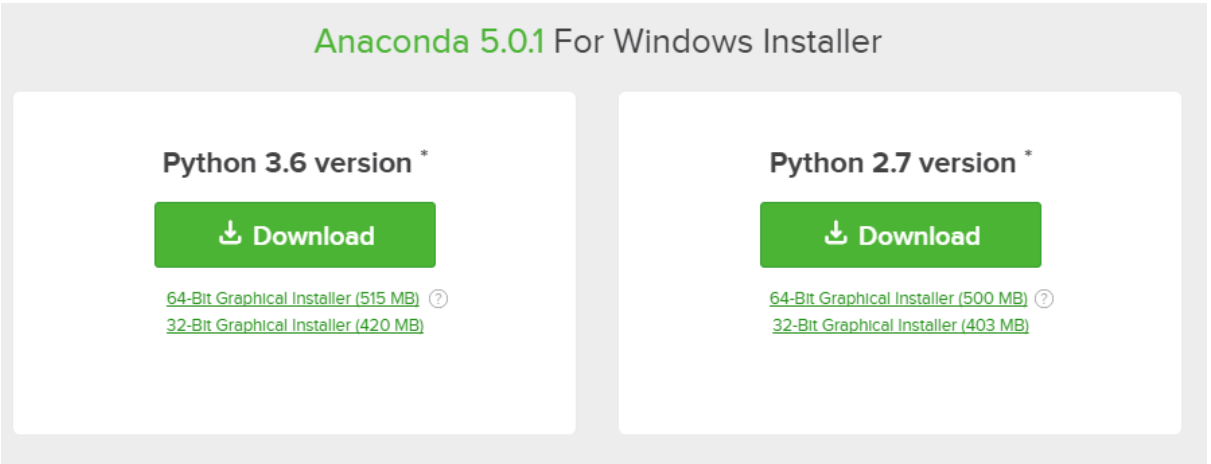


2. Select Windows

Select Windows where the three operating systems are listed.

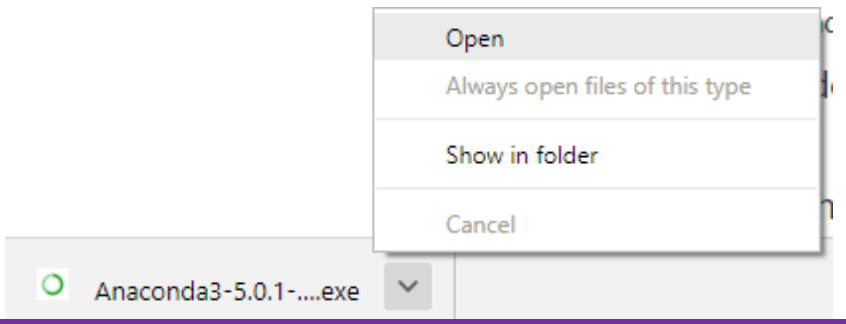
3. Download

Download the most recent Python 3 release. At the time of writing, the most recent release was the Python 3.6 Version. Python 2.7 is legacy Python. For problem solvers, select the Python 3.6 version. If you are unsure if your computer is running a 64-bit or 32-bit version of Windows, select 64-bit as 64- bit Windows is most common.



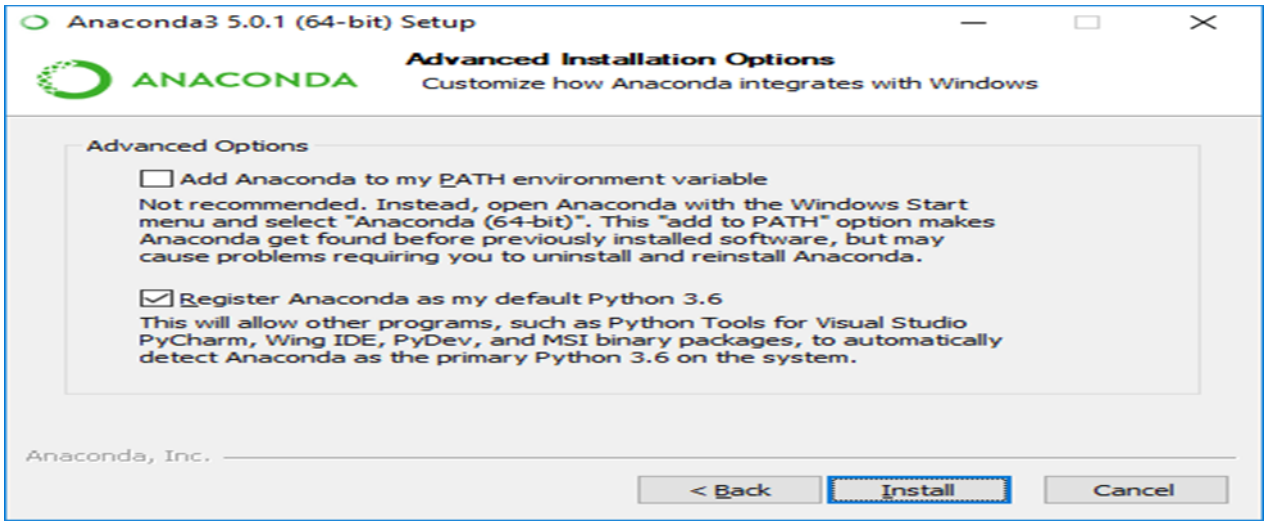
4. Open and run the installer

Once the download completes, open and run the ***.exe*** installer



At the beginning of the install, you need to click Next to confirm the installation

At the Advanced Installation Options screen, I recommend that you **do not check** "Add Anaconda to my PATH environment variable"



Installation Of Anaconda in Ubuntu

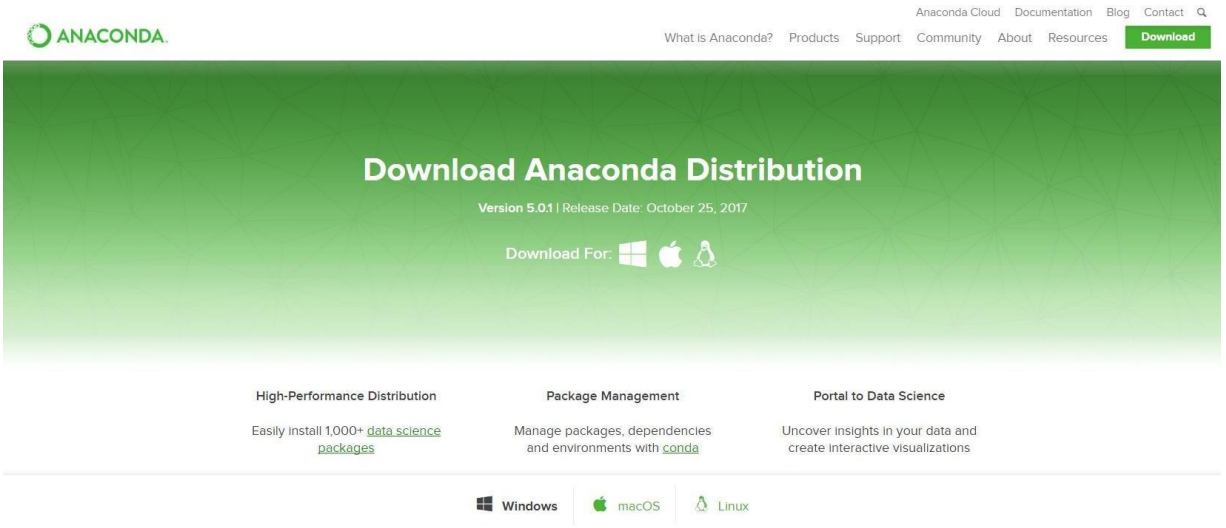
This section details the installation of the Anaconda distribution of Python on Linux, specifically Ubuntu 18.04, but the instructions should work for other Debian-based Linux distributions as well.

Ubuntu 18.04 comes pre-installed with Python (Version 3.6) and legacy Python (Version 2.7).

You can confirm the legacy version of Python is installed by opening up a terminal.

1. Visit the Anaconda downloads page

Go to the following link: Anaconda.com/downloads

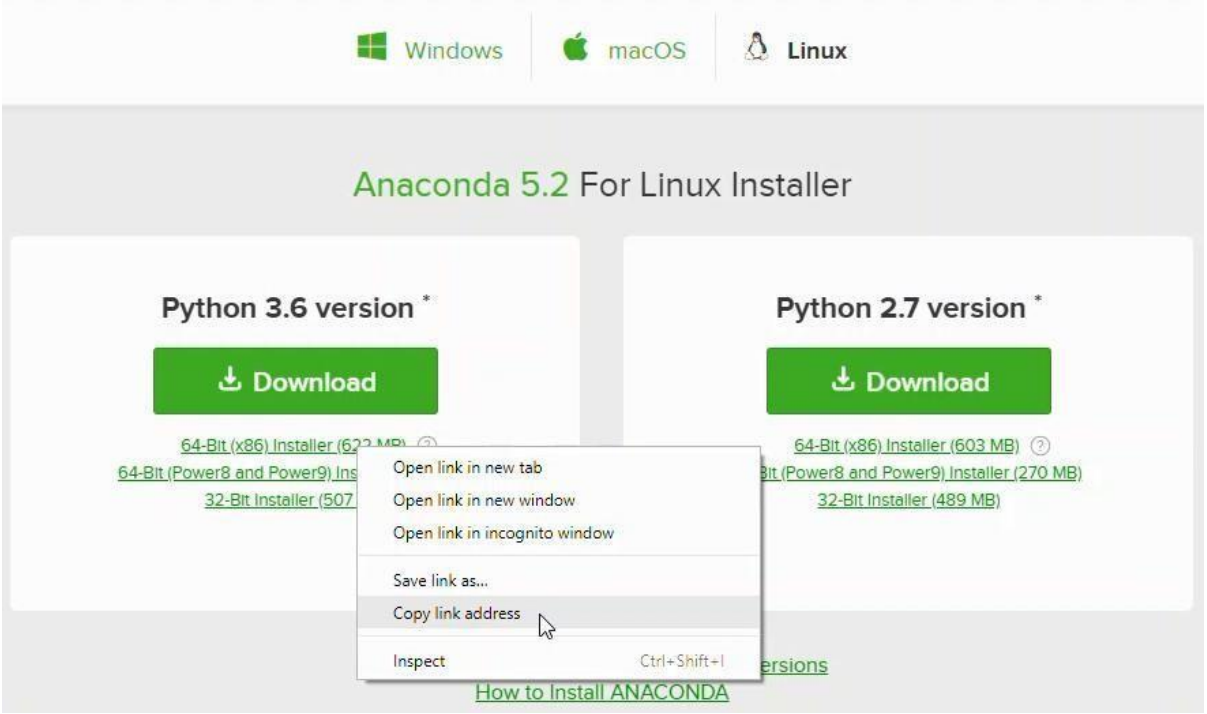


2. Select Linux

On the downloads page, select the Linux operating system.

3. Copy the bash (.sh file) installer link

In the **Python 3.6 Version*** box, right-click on the [64-Bit(x86) Installer] link. Select [copy link address].



4. Use wget to download the bash installer

Now that the bash installer (.sh file) link is stored on the clipboard, use wget to download the installer script. In a terminal, cd into the home directory and make a new directory called tmp. cd into tmp and use wget to download the installer. Although the installer is a bash script, it is still quite large and the download will not be immediate (Note the link below includes <release>. the specific release depends on when you download the installer)

$ cd ~
$ mkdir tmp
$ cd tmp
$ https://repo.continuum.io/archive/Anaconda3<release>.sh

5. Run the bash script to install **Anaconda3**

With the bash installer script downloaded, run the ***.sh*** script to install **Anaconda3**. Ensure you are in the directory where the installer script downloaded:

$ ls
Anaconda3-5.2.0-Linux-x86_64.sh Run

the installer script with bash.

$ bash Anaconda3-5.2.0-Linux-x86_64.sh
Accept the Licence Agreement and allow Anaconda to be added to your PATH. By adding Anaconda to your PATH, the Anaconda distribution of Python will be called when you type $ python in a terminal.

6. source the .bash-rc file to add Anaconda to your PATH

Now that **Anaconda3** is installed and **Anaconda3** is added to our PATH, source the .bashrc file to load the new PATH environment variable into the current terminal session. Note the .bashrc file is in the home directory. You can see it with $ ls -a.
$ cd ~

$ source .bashrc

Installation Of NumPy Command –

pip3 install numpy To check the

version

Import numpy as np

Installation of Scipy-stack Command –

pip3 install scipy-stack

NUMPY/SCIPY:-
NumPy is a Python library, which stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n-dimensional array object, provide tools for integrating C, C++ etc.

- NumPy array can also be used as an efficient multi-dimensional container for generic .data.
- The ndarray (NumPy Array) is a multidimensional array used to store values of same datatype. These arrays are indexed just like Sequences, starts with zero.
- The ndarrays are better than regular arrays in terms of faster computation and ease of manipulation.
- In different algorithms of Machine Learning like K-means Clustering, Random Forest etc. we have to store the values in an array. So, instead of using regular array, ndarray helps us to manipulate and execute easily.

Installation of scikit

Command pip3 install -u scikit-learn

SCIKIT:-
The functionality that scikit-learn provides include:

- **Regression**, including Linear and Logistic Regression
- **Classification**, including K-Nearest Neighbors
- **Clustering**, including K-Means and K-Means++
- Model selection
- **Preprocessing**, including Min-Max Normalization.

Installation of Pandas

Command- pip3 install pandas

PANDAS:-

- Merging and Joining Data Sets.
- Reshaping & pivoting Data Sets.
- Inserting & deleting columns in Data Structure.
- Aligning data & dealing with missing data.
- Iterating over a Data set.
- Analyzing Time Series.
- Filtering Data around a condition.
- Arranging Data in an ascending & descending.
- Reading from flies with CSV, TXT, XLSX, other formats.
- Manipulating Data using integrated indexing for DataFrame objects.
- Generating Data range, date shifting, lagging, converting frequency, and other other Time Series functionality.
- Subsettting fancy indexing, & label based slicing Data Sets that are large in size.
- Performing split apply combine on Data Sets using the group by engine. With Python Pandas, it is easier to clean & wrangle with your Data. features of Pandas make it a great choice for Data Science and Analysis.

Installation of Matplotlib Command-

pip3 install matplotlib

MATPLOTLIB:-

- Matplotlib is a visualization library in Python for 2D plots of arrays. It consists of several plots like line, bar, scatter, histogram etc.
- Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack.It can also be used with graphics toolkits like PyQt and wxPython.
- One of the advantage of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals.

Conclusion

Thus we have successfully installed Anaconda on windows and Ubuntu; We also have

described the libraries and the installation of the libraries.