

Artificial Intelligence Environment Setup

Authors: Group 1C1

Neba Roland, Okoye-Nobert C, Christopher K, Natnael A

21 January, 2021

African Leadership University

graphicx

1 Introduction

This section explains how to download and install some the different tools which would be used in the AI class. The tools include, Anaconda, Jupyter Notebook, Python, Tensorflow, Keras, Numpy, SCipy,Matplotlib,Pandas,Scikit-learn.

1.1 Downloads and Installations

- **Downloading and Installing Anaconda on Linux**

Anaconda is a distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytic, etc.), that aims to simplify package management and deployment. To install Anaconda we took a look at the Anaconda Official documentation. Here's a list of preferred instructions which could be follow to install this tool.

1. Firstly, we would want to download the Anaconda Installer for Linux
2. Once the program is downloaded, the second step would be to run some commands on the Linux terminal. The commands to be entered on the terminal is "sha256sum /path/filename"
3. To install Anaconda for Python 3.7 and 2.7 we assume the download is made on the Downloads directory, then use these commands
`bash /Downloads/Anaconda3-2020.02-Linux-x86_64.sh - Python 3.7`
`bash /Downloads/Anaconda2-2019.10-Linux-x86_64.sh-Python2.7`
if not, replace /Downloads/ with the path to the file you downloaded.
4. The installer prompts "In order to continue the installation process, please review the license agreement." Click Enter to view license terms.
5. Scroll to the bottom of the license terms and enter "Yes" to agree.

6. The installer then prompts you to click Enter to accept the default install location, CTRL-C to cancel the installation, or you can specify an alternate installation directory. If you accept the default install location, the installer displays “PREFIX=/home/username/anaconda2 or 3!” and continues the installation. It may take a few minutes to complete.
7. The installer prompts “Do you wish the installer to initialize Anaconda3 by running conda init?” We recommend “yes”.
8. The installer finishes and displays “Thank you for installing Anaconda 2 or 3!”
9. The installer provides a link to install PyCharm for Anaconda at <https://www.anaconda.com/pycharm>.
10. Close and open your terminal window for the installation to take effect, or you can enter the command `source ~/.bashrc`.
11. To control whether or not each shell session has the base environment activated or not, run `conda config --set auto_activate_base False` or `True`. To run conda from anywhere without having the base environment activated by default, use `conda config --set auto_activate_base False`. This only works if you have run `conda init` first.

Further documentations can be found at:

<https://docs.anaconda.com/anaconda/install/linux/>

- **Verifying Jupyter Notebook**

Jupyter Notebook is an open-source web application that allows one to create and share documents that contain live code, equations, data visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

Since Anaconda comes in-built with Jupyter Notebook it makes it easier to install this application. To check if Jupyter Notebook is pre-installed

1. We can run “conda activate” on the terminal to transition to the conda environment
2. Once conda is activated, then we exhibit Jupyter Notebook by running “Jupyter notebook” in the terminal.

Further documentations can be found at:

<https://jupyter.org/documentation>

- **Activating Python**

Python is a general-purpose coding language—which means that, unlike HTML, CSS, and JavaScript, it can be used for other types of programming and software development besides web development. That includes back end development, software development, data science and writing

system scripts among other things.

Since Anaconda comes in-built with Python it makes it easier to install this application. To check if Python is pre-installed

1. We can run "conda activate" on the terminal to transition to the conda environment
2. Once conda is activated, then we exhibit Python by running "Python" in the terminal.

This shows the version of python installed and now the environment is set to python environment which could be used performing any python function

- **Installing Tensorflow**

Created by the Google Brain team, TensorFlow is an open source library for numerical computation and large-scale machine learning. TensorFlow bundles together a slew of machine learning and deep learning (aka neural networking) models and algorithms and makes them useful by way of a common metaphor.

To install Tensorflow, the following steps should be followed:

- On the Linux terminal, we create a Tensorflow environment. To do this, we choose a name for the Tensorflow environment such as "tf".
- Now we'd want to install the current release of the CPU-only Tensorflow. To achieve this, the following commands were entered:

1. "conda create -n tf tensorflow"
2. After that command, Tensorflow can be activated by entering the following
"conda activate tf"

- Tensorflow is now installed and ready to use.

Further documentations can be found at:

<https://docs.anaconda.com/anaconda/user-guide/tasks/tensorflow/>

- **Installing Keras, Numpy, SCipy,Matplotlib,Pandas,Scikit-learn**

Keras
Keras is a neural networks library written in Python that is high-level in nature – which makes it extremely simple and intuitive to use. It works as a wrapper to low-level libraries like TensorFlow or Theano high-level neural networks library, written in Python that works as a wrapper to TensorFlow or Theano.

To install Keras, the following command should be run on the terminal:
"conda install -c conda-forge keras"

Numpy

NumPy is an open-source numerical Python library. NumPy contains a

multi-dimensional array and matrix data structures. It can be utilised to perform a number of mathematical operations on arrays such as trigonometric, statistical, and algebraic routines.

To install NumPy, the following command should be run on the terminal:
"conda install -c anaconda numpy

SCiPy

SCiPy is an open-source Python library which is used to solve scientific and mathematical problems. It is built on the NumPy extension and allows the user to manipulate and visualize data with a wide range of high-level commands.

To install SCiPy, the following command should be run on the terminal:
"conda install -c anaconda scipy

Matplotlib

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+

To install matplotlib, the following command should be run on the terminal: "conda install -c anaconda matplotlib

Pandas

Pandas is mainly used for data analysis. Pandas allows importing data from various file formats such as comma-separated values, JSON, SQL, Microsoft Excel. Pandas allows various data manipulation operations such as merging, reshaping, selecting, as well as data cleaning, and data wrangling features.

To install matplotlib, the following command should be run on the terminal: "conda install -c anaconda pandas

Scikit-learn

Scikit-learn is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

To install Scikit-learn, the following command should be run on the terminal: "conda install -c anaconda scikit-learn

1.2 Testing Data with Jupyter Notebook

In order to test the environment and the different libraries that have been installed, we used a covid 19 data set from Ghana

Testing the Pandas Library In our Environment

Firstly we had to import the library and used pd as an alias for it in the notebook. We further used this library to explore the data set by using the columns, describe, info, head, mean and median objects on our data frame. We used the data directly hosted from the cloud by hosting in on a public github repository. We used pandas library to read the data into a data frame using the read_csv object as

```
df = pd.read_csv('https://raw.githubusercontent.com/Kuzagbe/AI-Group-Assignment/main/Ghana_Covid19_DailyActive.csv')
```

Testing the Matplotlib Library In our Environment

In order to test matplotlib, we created various kinds of histograms for all the columns in the dataset by running the following command "df.hist(bins= 30, figsize=(20,20), color='blue')"

Testing the numpy Library In our Environment

To test the numpy library, we created an identity matrix of 8 rows and as well divided the dataset into two and converted each part into a numpy array, this was so that we can be able to pass these two arrays into an sklearn model

Testing the sklearn Library In our Environment

We imported the test train split object from the sklearn library and used it to divide the dataset into train and test portions using the two numpy arrays that were created above, we further tested the library by importing the linear regression object from it and fitting the train dataset into it and calculating the accuracy using the score method and further checking the coefficient and intercept of the newly created model. The dataset was not fully fit for a linear regression model or problem but then we just used it this way for the sake of checking if the library works properly

Testing the Tensorflow and Keras Library In our Environment Since keras is a component of the tensorflow library, we just tested them together by using an artificial neural network on the dataset to create a linear regression model, based on previous information we had about this library, we were able to implement a simple artificial neural network on the dataframe

Testing the Scipy Library In our Environment In order to test the scipy library, we realized that it was not really possible to do it directly on a dataset so we just tried out a few mathematical formulas that come with scipy in order to check if the library has been properly installed, some of the mathematical concepts we tried are permutation and also the determinant of a matrix. An image for this representation can be found below: C:/Users/EBUKA/Downloads/Scipy

2 Conclusion

In conclusion, the major installation required to setup a suitable environment for machine learning and/or data science is Anaconda. By installing Anaconda,

it becomes much easier and convenient to install other scientific computing libraries like sci-kit learn and numpy as anaconda comes with a package manager, Conda.

Also, to test the data sets we looked for a publicly available dataset on Kaggle that was focusing on African context. And as such, testing the various environments and libraries became much easier.

3 GitHub Link

A Github link that contains the Jupyter notebook file, along with a copy of this latex file can be found at https://github.com/NatnaelAlemayehu/AI_G1C1_Assignment1.git

4 References

- Installing on Linux — Anaconda documentation. Docs.anaconda.com. (2021). Retrieved 24 January 2021, from <https://docs.anaconda.com/anaconda/install/linux/>
- Anaconda (Python distribution). En.wikipedia.org. (2021). Retrieved 24 January 2021, from [https://en.wikipedia.org/wiki/Anaconda_\(Python_distribution\)](https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)).
- Yegulalp, S. (2021). What is TensorFlow? The machine learning library explained. InfoWorld. Retrieved 24 January 2021, from <https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html>.
- Jain, K. (2021). SKLearn — Scikit-Learn In Python — SciKit Learn Tutorial. Analytics Vidhya. Retrieved 24 January 2021, from <https://www.analyticsvidhya.com/blog/2015/01/scikit-learn-python-machine-learning-tool/>.
- Pandas (software). En.wikipedia.org. (2021). Retrieved 24 January 2021, from [https://en.wikipedia.org/wiki/Pandas_\(software\)](https://en.wikipedia.org/wiki/Pandas_(software)).