1 Environment

Throughout these exercises, we will be working with Python. Python is a scripting language, which is widely used nowadays. In particular, Python is among the most popular programming languages for data science and machine learning. This tutorial gives you a brief introduction of the overview of the environments that we will use and brief instructions to setup your own environments.

1.1 Required libraries/packages

During the exercise sessions, the following packages will be required in your working environment:

- Python: for the first exercise, we will use Python 3.x.
- Numpy: a fundamental package for scientific computing with Python [1]. It contains
  - a powerful N-dimensional array object
  - sophisticated (broadcasting) functions
  - tools for integrating C/C++ and Fortran code
  - useful linear algebra, Fourier transform, and random number capabilities

  In general, Numpy allows us to perform efficient operations with matrices in an easy way, like we do in Matlab.
- scikit-learn [2]: a simple and efficient tool for data mining and data analysis. It contains efficient implementations of many common machine learning models.
- Matplotlib [3]: a Python 2D plotting library. We will need it to draw figures, graphs, etc. from a Python program
- Jupyter Notebook [5]: a handy web application that allows us to write and execute Python code in the web browsers. We will mostly work with it through the exercise sessions.

1.2 Setting up

Google Colab: It is recommended that you work online with Google Colab. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs [6]. You can access Google Colab via the following url `https://colab.research.google.com/`

Anaconda/Miniconda: However, you also have the choice to setup everything at your own personal computer. For that you should install either Anaconda or Miniconda (a minimal version of Anaconda with just the basic packages installed that saves you some disk space) [4]. Anaconda is one of the most popular Python data science platforms. It is packed with several common data science packages, including the ones listed in 1.1. Note that these packages except for Python are not included in Miniconda and you should manually install them using the pip command. We will go through that in the next steps. Anaconda also allows us to create, maintain virtual environments, and easy installations of additional data science packages in case it is needed.

Below is a step-by-step guide to set up your environment. In the steps below you can find how to install Miniconda and the packages that we are going to use on Linux. Note that although the instructions are for Linux, it is straightforward to install Miniconda on Windows or Mac OS, and then follow the steps from step 2 onwards.

1. Install Miniconda: Download and install Miniconda from the link `https://docs.conda.io/en/latest/miniconda.html`. Don't forget to choose the right version for your OS. You can download the file with Python 3.9 pre-installed. When you download your file open a linux terminal and run the following commands:

   ```
   chmod +x your_file.sh
   ./your_file.sh
   ```

2. Then create a new python environment with Python 3.9 (or any other version of python 3 that you prefer). To do so run:

```
conda create –n python3.9 python=3.9
```

3. Activate your newly installed environment:

   ```
   conda activate python3.9
   ```

4. Install Numpy by typing:

   ```
   pip install numpy
   ```

5. Install scikit-learn by typing:

   ```
   pip install sklearn
   ```

6. Install matplotlib by typing:

   ```
   pip install matplotlib
   ```

7. Install scikit-learn by typing:

   ```
   pip install notebook
   ```

After executing the previous steps without any errors, you should have a working Python 3.9 environment. You can validate that your installation is working properly using the following steps.

1. Open your terminal and activate your Python 3.9 environment (execute this step in the case that you are not already inside the Python 3.9 environment):

   ```
   conda activate python3.9
   ```

2. Invoke Python:

   ```
   python
   ```

3. Import Numpy:

   ```
   import numpy
   ```

4. Import matplotlib :

```
import matplotlib
```

5. Import scikit-learn:

   ```
   import sklearn
   ```

6. Close the previous terminal or exit() from the python terminal. Open you jupyter notebook and navigate with you browser into the .ipynb file that you want to execute:

   ```
   jupyter notebook
   ```

If you can import these packages, everything is working properly with your installation.

References

[1] http://www.numpy.org/.

[2] http://scikit-learn.org/.

[3] https://matplotlib.org/.

[4] https://www.anaconda.com.

[5] http://jupyter.org/

[6] https://research.google.com/colaboratory/faq.html