1 Environment

Throughout the excercises, we will be working with Python. Python is a scripting language, which is widely used nowadays. In particular, Python is among the most popular programing languages for data science and machine learning practices.

This tutorial gives an overview of the environments which we will use and a brief instructions to setup your own environments.

1.1 Required libraries/packages

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

- Python: version 2.7
- Numpy: a fundamental package for scientific computing with Python. It contains [1]
  - 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 tools 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 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 excercise sessions.

1.2 Setting up

Any distribution of Python 2.7.x, together with installations of the required packages, will be enough for the excercise sessions. Nevertheless, it is recommended to use Anaconda Python platform [4], espcially if you use Windows.

Anaconda is one of the most polular Python data science platform. It is packed with several common data science packages, including the ones listed in 1.1. 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. Note that this guide only focuses on Windows OS, which are used at the lab. Nevertheless, it is even more convenient to work with Python and setup your environment in Linux or Mac OS.

1. Install Anaconda: Download and install Anaconda from the link below (Don't forget to choose the right version for your OS):
   https://www.anaconda.com/download/.

2. Open anaconda command prompt

3. Install Numpy by typing:

   ```
   pip install numpy
   ```

4. Install scikit-learn by typing:

   ```
   pip install scikit-learn
   ```

5. Install matplotlib by typing:

   ```
   pip install matplotlib
   ```

6. Install scikit-learn by typing:

   ```
   pip install notebook
   ```

After finishing these steps without any errors, you should already have a working environment. You can check your install by the following steps:

1. Install Anaconda: Download and install Anaconda from the link below (Don't forget to choose the right version for your OS): https://www.anaconda.com/download/.
2. Open anaconda command prompt
3. Invoke Python:

```
python
```

4. Import Numpy:

```
import numpy
```

5. Import matplotlib :

```
import matplotlib
```

6. Import scikit-learn:

```
import sk-learn
```

If you can import those packages, you are ready to solve the exercises.

References

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

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

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

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

[5] http://jupyter.org/