

AdvNLP Practical Work: Introduction to the Environment

Summary

The goal of this notebook is to get you introduced to an environment that will enable you to run experiments in text analysis for the AdvNLP course. We will use the [Python 3](#) programming language, with [Jupyter](#) notebooks to write instructions and log your results.

To be able to use Python and Jupyter Notebooks, you have three options:

1. environment-specific installation with Conda
2. simple installation (platform-wide)
3. no local installation, use an online service: Google Colab.

First option: Conda environment manager

If you want to manage several versions of Python and its modules, a good idea is to install an environment manager such as Conda. Before installing Python and Jupyter locally, installing Conda will help you manage different versions of Python and of other modules on your computer, for instance for several courses. (You may wish to install Conda before Python, but you can also do it later, in which case it will install its own Python libraries.)

Miniconda is a minimal environment manager, while Anaconda is a larger software distribution that includes also Python and R software.

- To install Conda, follow the instructions at <https://conda.io/projects/conda/en/latest/user-guide/install/index.html> and consider installing Miniconda. Follow the instructions for your OS: Windows, Linux or MacOS.
- You will be prompted to get a suitable installer from <https://conda.io/en/latest/miniconda.html> (an executable for Windows, or a `bash` script for Linux).
- In Linux, run `wget https://repo.continuum.io/miniconda/Miniconda3-latest-Linux-x86_64.sh` and then `bash Miniconda3-latest-Linux-x86_64.sh`
- Conda commands are the same on all OSs -- the most important ones are for creating and loading an environment
- after you install Conda, commands are performed from the command line (command prompt, a Power Shell, Anaconda Prompt, etc.)
- try `conda info` for a start.
- to get familiar with Conda: <https://conda.io/projects/conda/en/latest/user-guide/getting-started.html>
- cheatsheet of Conda commands: <https://conda.io/projects/conda/en/latest/user-guide/cheatsheet.html>
- to create an environment called *AdvNLP* for this course and activate it:
 - `conda create --name AdvNLP python=3.11`
 - `conda activate AdvNLP`
 - `conda list` to see what is installed in the current environment
 - `conda deactivate` to exit from the environment

Please continue with the second option (except its first bullet point) to complete your installation of the Jupyter server, which will allow you to open, edit, and run notebooks.

Second option: system-wide installation of Python

With a local installation you will be able to work offline, and gain experience with Python and Jupyter, so we recommend this solution, with or without Conda.

- You need to install [Python 3](#) (the latest stable Python 3 version).
- Install then the [Jupyter package](#).
- A good idea is to use the [Package Installer for Python \(pip\)](#) to add new packages.
- All these operations are done in the command line (on Windows: `cmd` or Power Shell).
- The `MSE_AdvNLP_Introduction.ipynb` must be saved into your repository of Jupyter notebooks.
- Start Jupyter with `jupyter notebook --no-browser`
- Open the Jupyter interface at <http://localhost:8888> and launch this notebook.

To use Jupyter with Conda

Jupyter can be aware of Conda environments. A notebook can use a kernel associated to a specific Conda environment. Once Jupyter is installed (preferably in the same Conda environment that you will use), run these commands:

- `conda activate AdvNLP`
- `python -m ipykernel install --user --name AdvNLP --display-name "AdvNLP"`

Once you have done this, you can just choose this kernel (environment) when you open a new Jupyter notebook.

Third option: use Colab online (no local installation)

You can use a free online environment provided by Google, named [Colab](#). You will need a Google account (the same as Gmail if you are using it), and your files will be stored on Google Drive. Once you open a session in Colab, you will need to import the Jupyter notebooks to Colab. This environment is offered by Google for machine learning experiments, with CPU / GPU / TPU computation, but not for time-consuming jobs. All Python packages that we will use are installed.

Running simple tests

To test your settings, run some a simple operation first.

```
In [ ]: 1 + 2 * 3 / 4
```

Create a list with the first ten numbers (1 to 10) and then print the list of their squares.

```
In [ ]: numbers = list(range(11))
squares = [n ** 2 for n in numbers if n > 0]
print(squares)
```

Getting more familiar with Python

If you need to get familiar with Python, one option is to go through [this notebook](#) from the [University of Edinburgh NLP course](#). (They also offer a useful [introduction to the numpy package](#).) The notebook above also gives a very short introduction to [Matplotlib](#), which will allow you to plot graphics in Jupyter.

A more thorough introduction is provided by J.R. Johansson in his [Scientific Python Lectures](#) accompanied by Jupyter notebooks for practicing. And of course the definitive reference is the [Python documentation](#) itself.