Installing Python and Jupyter Notebook

All course material will be written in Python 3.

1 I already have Python

If you already have Python installed and if you are familiar with installing Python packages with `pip`, you can use this to install Jupyter. Please make sure you that you have upgraded to the latest version of `pip` using the following command:

```
pip3 install --upgrade pip
```

Next install the Jupyter notebook with:

```
pip3 install jupyter
```

In this course we will use many external packages that need to be installed. To manage them smoothly, you are recommended to use the **Anaconda** Python distribution. If you do not have Anaconda installed, please see Section 2 for how to install it. If you already have Anaconda installed, please go to Section 3 to create a virtual environment with all the required packages for this module.

2 I am new to Python

2.1 Download Anaconda

Anaconda is a Python distribution for data science and machine learning. It can be freely downloaded here.

2.2 Install Anaconda

Follow these instructions to install Anaconda distribution on your computer. The instructions support Windows, Mac and Linux.

2.3 Test the installation

Jupyter notebook should now be successfully installed on your local machine. Please open a new terminal window with one of the commands below:

Windows: press `[WIN]+[r]` and type `cmd`. Then press `[Enter]`.
Mac: press `[command]+[space]` to open Spotlight. Then type terminal and press `[enter]`.

Linux: in Unity/Gnome press `[Ctrl]+[Alt]+[T]`.

Type the following command in terminal to run the notebook.

```
jupyter notebook
```

This should open a browser and a webpage that looks similar to the illustration in Figure 1. (It is possible that the webpage shown on your local machine make look slightly different).

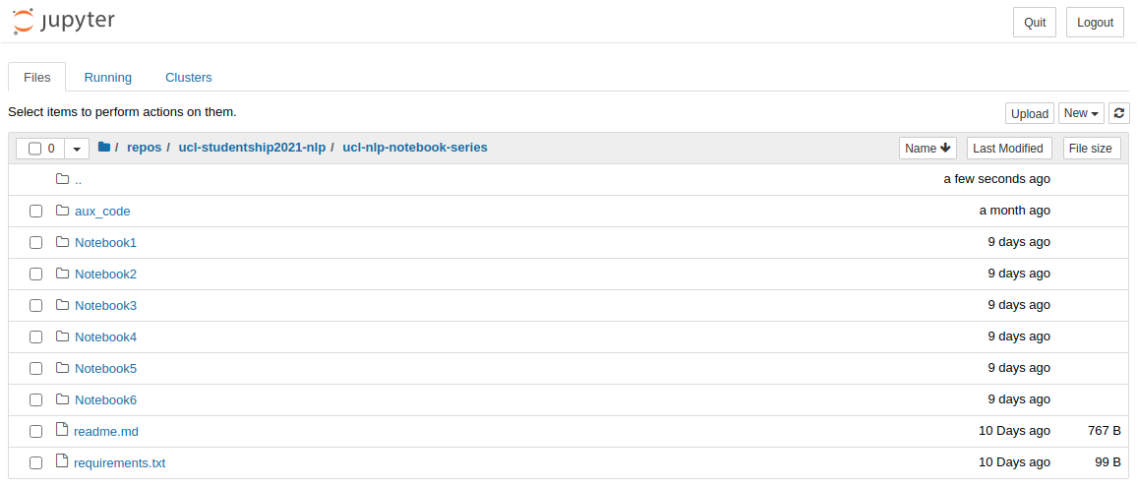

Figure 1: Jupyter notebook in browser

3 Create virtual environment

A virtual environment is a named, isolated, working copy of Python that that maintains its own files, directories, and paths so that you can work with specific versions of libraries or Python itself without affecting other Python projects.

To create a new virtual environment that contains all packages needed for this module and a stable Python version please run the following two commands (replace `*name*` with your preferred name):

```
conda create --name *name* python=3.8.10 numpy pandas scipy
scikit-learn matplotlib seaborn gensim=4.0.1 spacy nltk
```

```
conda install --name *name* -c conda-forge notebook ipykernel
langdetect swifter imbalanced-learn emoji
```

To activate your newly created virtual environment, run the following command:

```
conda activate *name*
```

We need to install one more thing, this time using `pip`. After activating your virtual environment run:

```
pip3 install PyMuPDF
```

4 Check package versions

You can check the versions of the installed packages by opening up a notebook, importing all the packages and checking the `__version__` attribute.

```
import numpy as np

np.__version__
```

Below you can see the versions for each package that guarantee the notebooks to run smoothly:

- matplotlib == 3.4.2
- numpy == 1.20.3
- pandas == 1.3.0
- gensim == 4.0.1
- PyMuPDF == 1.18.17
- spacy == 2.3.5
- scikit-learn == 0.24.2
- nltk == 3.6.2
- seaborn == 0.11.2
- swifter == 1.0.9
- langdetect == 1.0.9
- imbalanced-learn == 0.8.0
- emoji == 1.4.2
- ipykernel == 6.3.1
- notebook == 6.4.3