Syllabus for Reading Group "Python Programming" @ PhD ECOSTAT

Pietro Battiston

February 16, 2020

All classes are intended to be interactive, with students bringing their own laptops, (or using the lab's computers), experimenting live, and solving small exercises I will assign in real time.

Topics in [square brackets] might be skipped in the interest of time, and others could be added, depending on participants' interests.

Python 3 will be used, inside a **Jupyter** (or **JupyterLab**) environment. For students who have no experience with Python, the use of **anaconda**[1] or **miniconda**[2] is recommended, in particular if using Microsoft Windows or OS X (Linux user can alternatively rely on the system's package manger, complemented by **pip/pip3**).

Exercises proposed below are purely indicative of the level that students are expected to reach.

PART I: General introduction

- Lecture 1 (Feb. 26): introduction to the Python syntax and to the Python docs, overview of fundamental data structures, basics of object oriented programming
- Lecture 2 (Mar. 4): control flow, how to import/create a module, [how to create a `class`, subclassing], how to read/write to a file, [web

[1] `https://www.anaconda.com/`

[2] `https://conda.io/miniconda.html`

scraping], plotting with **matplotlib**

Exercises:

- load a text from disk and apply replacement patterns defined inside a specific module
- download a web page and look in it for given items

PART II: Python for data analysis

- Lecture 3 (Mar. 11): introduction to **numpy** and **pandas**: types, vectorization, indexing, I/O, [reshaping data, `MultiIndex`, handling dates and times], groupby operations
- Lecture 4 (Mar. 18): data analysis: **statsmodels**, **scikit-learn** for machine learning, [interaction with **R** through **rpy2**]

Exercises:

- download macroeconomic data with **pandaSMDX**, select a given country, plot the desired variable averaged by year together with a filtered version, save the result to a file
- download reviews from a given website (e.g. trustpilot.com), analyse most frequent terms, use their count to predict whether a review is positive or negative

PART III: projects exposition

- Lecture 5 (Apr. 22): presentation of group projects written in Python. The projects will be chosen by participants, or assigned by the me, during Lecture 4.