

Python #0

Python Setup

Location: https://people.sc.fsu.edu/~jburkardt/classes/python_2022/python00/python00.pdf

Freely adapted from the Python lessons at <https://software-carpentry.org/>



Python

Python is one of the most popular modern programming languages.

- *Predominant language for machine learning;*
- *Free;*
- *Windows, MacOS, and Linux;*
- *Use interactively, or as a program, or mixed with other programs*

1 Install anaconda

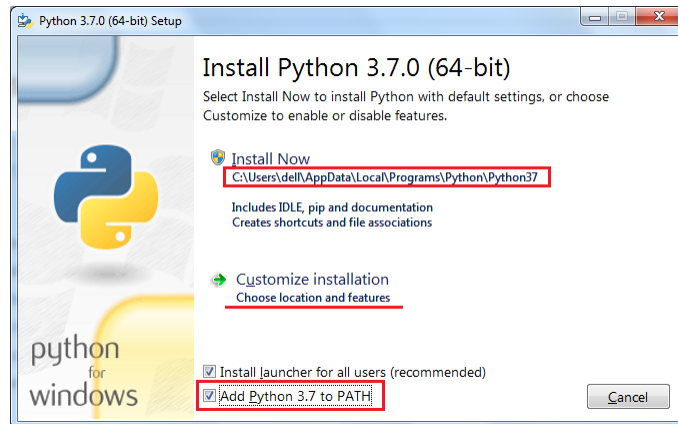
This class will involve many programming examples, exercises, and homework. The programming language will be Python, version 3.0. Along with Python, we will need associated libraries `numpy`, `scipy`, `matplotlib`, `PIL`, `pandas`, `seaborn` as well as the specialized machine learning library `scikit-learn`.

If you are a computer whiz, you may be able to download and install these applications on your own. If you have a Linux or Macintosh system, you might already have some of these installed by default. However, in any case, it is best to set up your programming environment by downloading Anaconda, and its package manager `conda`.

Go to the site

<https://www.anaconda.com/products.individual>

and download a copy of Anaconda for your operating system.



Anaconda Download for Windows

2 Verify your installation

Whether you installed Anaconda, or downloaded things on your own, or are lucky enough to be using a computer that seems to have the proper Python environment, it's important that you check that everything is there and accessible.

You work with Python in a terminal window, so we need to get that first:

- Windows: an Anaconda powershell prompt should appear in your Start menu. Click on it;
- MacOS: In the Finder, open the /Applications/Utilities folder, then double-click Terminal;
- Linux: right click on the background screen, to get a menu that includes Open Terminal.

Within the terminal window, you can verify that the appropriate packages are now available. To do this, you will run a Python session. If things are going well, then each of the following commands should execute silently. If there is a problem, you will get an error message like *Command not found*, or *No module named 'numpy'*. Assuming there are no such warnings, then the `quit()` command will terminate the Python session, and we have confirmed that we have all the packages we need.

```
python3
>> import numpy
>> import scipy
>> import matplotlib
>> import pandas
>> import PIL
>> import seaborn
>> import sklearn
>> quit()
```

Each of these commands is asking whether a particular library is accessible. If you get an error message that states something like

```
ModuleNotFoundError: No module named 'numpy'
```

then something is wrong with your setup.

It's possible that the Pitt IT Help Desk will be able to help you. Their web address is

```
https://www.technology.pitt.edu/247-it-help-desk
```

3 Look ahead at Python

We will try to teach you Python starting with the basics. Nonetheless, it can be helpful for you to try to get an advance look at the language. If you have no programming experience at all, this may take some time to get used to. If you have experience in programming in C, C++, Fortran, Java, or other procedural languages, then a lot of the features of Python will be familiar. If, better yet, you have worked with an interactive language like MATLAB or R, then Python will seem very similar, and it will just be a question of “translating” how you do things.

One excellent way to start is to refer to the online tutorial available at:

<https://docs.python.org/3/tutorial/>

and work through some of the initial examples.

You don’t need to become an expert, but you should start by focusing on:

- how you define and use data;
- how vectors and arrays organize multiple data items;
- how `for` and `while` carry out repetitive operations on data;
- how you define and use functions to modify data;
- how the `import` command makes a library accessible;

4 The Python lessons






This is a class on machine learning. After learning about various problems and algorithms, you will need to do some programming in a lab exercise. This programming will be done using the Python language. Previously, we’ve tried to cram an introduction to Python into the first week of class. Students don’t seem to absorb all the information instantly, and in later weeks, they find themselves having trouble with Python itself, when what we really want is for the students to be having trouble with the machine learning concepts.

So for this class, we have decided to include a regular, weekly lesson in Python. These lessons form a connected project, which doesn’t focus on machine learning, but which introduces many concepts needed for your machine learning work. We hope that this approach will reduce the frustrations that many students have expressed in the past, when they feel that their inadequate understanding of Python has kept them from successfully mastering the machine learning exercises.

You are required to do these Python exercises and submit the results for each lesson.

5 The Python project

Inflammation data

| Patients | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Day 6 | Day 7 |
|---|-------|-------|-------|-------|-------|-------|-------|
|  | 0 | 0 | 1 | 3 | 1 | 2 | 4 |
|  | 0 | 1 | 2 | 1 | 2 | 1 | 3 |
|  | 0 | 1 | 1 | 3 | 3 | 2 | 6 |
|  | 0 | 0 | 2 | 0 | 4 | 2 | 2 |
|  | 0 | 1 | 1 | 3 | 3 | 1 | 3 |

Our imaginary colleague “Dr. Maverick” has invented a new miracle drug that promises to cure arthritis inflammation flare-ups after only 3 weeks since initially taking the medication! Naturally, we wish to see the clinical trial data, and after months of asking for it we have been provided with a CSV spreadsheet. The CSV file contains the number of inflammation flare-ups per day for the 60 patients in the initial clinical trial, with the trial lasting 40 days. Each row corresponds to a patient, and each column corresponds to a day in the trial. Once a patient has their first inflammation flare-up they take the medication and wait a few weeks for it to take effect and reduce flare-ups.

As part of this project, we will see how to read the data file into the computer, how to compute some simple statistical measurements, how to make plots and visualizations of the data, and make some judgments about whether the data indicates that the new drug is effective.