Introduction to Python

Erik Dam
Raghavendra Selvan
Sandeep Sengar

Data Science Lab
Faculty of SCIENCE

UNIVERSITY OF COPENHAGEN

Data Science Lab

Goal: Enrich research at SCIENCE

In both Statistics and CS data science, we have triple:

- Introduction to R / Python
- PhD: Methods in Statistics / Machine Learning and Imaging
- PhD: Projects course ("write paper using Methods")

In addition, the Data Science Lab offers small consultations and larger collaborations.

Our hope here is that you get started with Python, start working with your own data, and then come back for advice, discussions, and potentially collaborations.

Python?

For Data Science software, why choose Python?

Criteria:

- Ease of prototyping/development
- Performance (speed/memory)
- Environment (IDE, debug, profiler)
- Open source communities
- Data Science relevancy
- Cost ($)

Performance + IDE ➔ C, C++, C#, …

… + Ease + Data Science ➔ Matlab

… + Communities + Cost ➔ **Python**

Aims of a Programming Course

Level of Programming skills:

1. Read a program
2. Modify and extend a program
3. Develop program from scratch
4. Develop large program framework

Depending on your background, the aim here is that you get the foundation to reach level 1, 2 or 3.

Advertisement

Programming is just a tool …

The Data Science Lab offers a PhD course in May:

Machine Learning and Imaging Methods
(MLI, 4.5 ECTS) is a Methods course where the PhD students are introduced to data science methods for extraction and cleaning of data, analysis of images/text/…, and machine learning methods.
We will use Python for the examples.

https://datalab.science.ku.dk/english/course

Data Science at SCIENCE/ERDA

In this course, we show examples of Python interactive notebooks using Jupyter / JupyterLab on ERDA.

Because:

- You are required to have reproducible versions of your research on ERDA (electronic research data archive).
- ERDA has storage for data
- ERDA has storage for sensitive data (SIF)
- ERDA has virtual computers
- ERDA will get GPUs also

Program

The course consists on 6 lecture. A "lecture" is a mix of lecture, demonstration, exercise, discussion, and a break.

Keep your mind, fingers, and eyes busy!!

	Tuesday	Wednesday
09.00	0. Intro 1. Basics, variables, loops, … (Erik)	4. I/O, Visualization (Sandeep)
11.00	2. Data structures (Raghav)	5. Machine Learning basics (Erik)
	Lunch Break	
13.30	3. Function, Modules, Objects (Sandeep)	6. Putting it all together (Raghav)

Background material

For Python Basics:

- www.python.org
  The official site offers documentation, tutorials, …
- http://cs231n.github.io/python-numpy-tutorial/
  A nice tutorial on Python basics including some numpy/scipy

Python Data Science Handbook:
http://github.com/jakevdp/PythonDataScienceHandbook
An excellent book, freely available online with Jupyter notebooks for the examples. We picked some of the exercises from here.