NOVA IMS Information Management School

Programming for Data Science

Fall Semester

Introduction to Python

Leon Debatin
Sofia Pereira

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Agenda

- Tools:
  - Python
  - Anaconda
  - Jupyter Notebook
- Setting up your environment

Ground Rules of practical classes

- Each group has 2 practical classes per week (Practical + Lab).
- Each week, the practicals will focus on implementing the techniques covered in the theoretical classes
- **Attendance in the practical classes is not strictly mandatory to complete the course (but heavily recommended)**
- **Attend the schedules assigned to your group**

Grading

The practical component of the course will have an individual component (Labs) and a final group project.

Evaluation Labs (50%)

- At the start of the lab, you may be informed that you need to deliver it by the end of class (it can happen any week)
- 6 labs
- The best 5 (out of the 6) will be accounted for in this component

Group Project (40%)

- You will be provided with a repository containing the possible projects with varying levels of difficulty.
- You can select any project out of the options provided.
- Grading will take into account our perceived difficulty of the project you choose.
- More details later in the semester.

Disclaimers

- Participation in the Project is required even if you opt for non-continuous assessment.
- Continuous assessment requires participation in at least 4 evaluation labs.
- The minimum grade in the project and practical exercises is 8.00 (out of 20).

Tools - Python

Python is a programming language that lets you work quickly and integrate systems more effectively.

Python is a widely used ***high-level programming language*** *for* ***general-purpose programming***[...]

Tools - Python

- Python language is one of the most popular tools for data science and analytics
- Easy to Read, Learn and Use
- Hundreds of Python Libraries and Frameworks
- Supportive Python Community
- **Support for Big data**, **Machine Learning** and **Data Mining**

Tools - Anaconda

- A very popular platform for Data Science
- Easy to install and use
- Provides easy access to many libraries often used for data manipulation and Machine Learning tasks

Tools – Jupyter Notebook

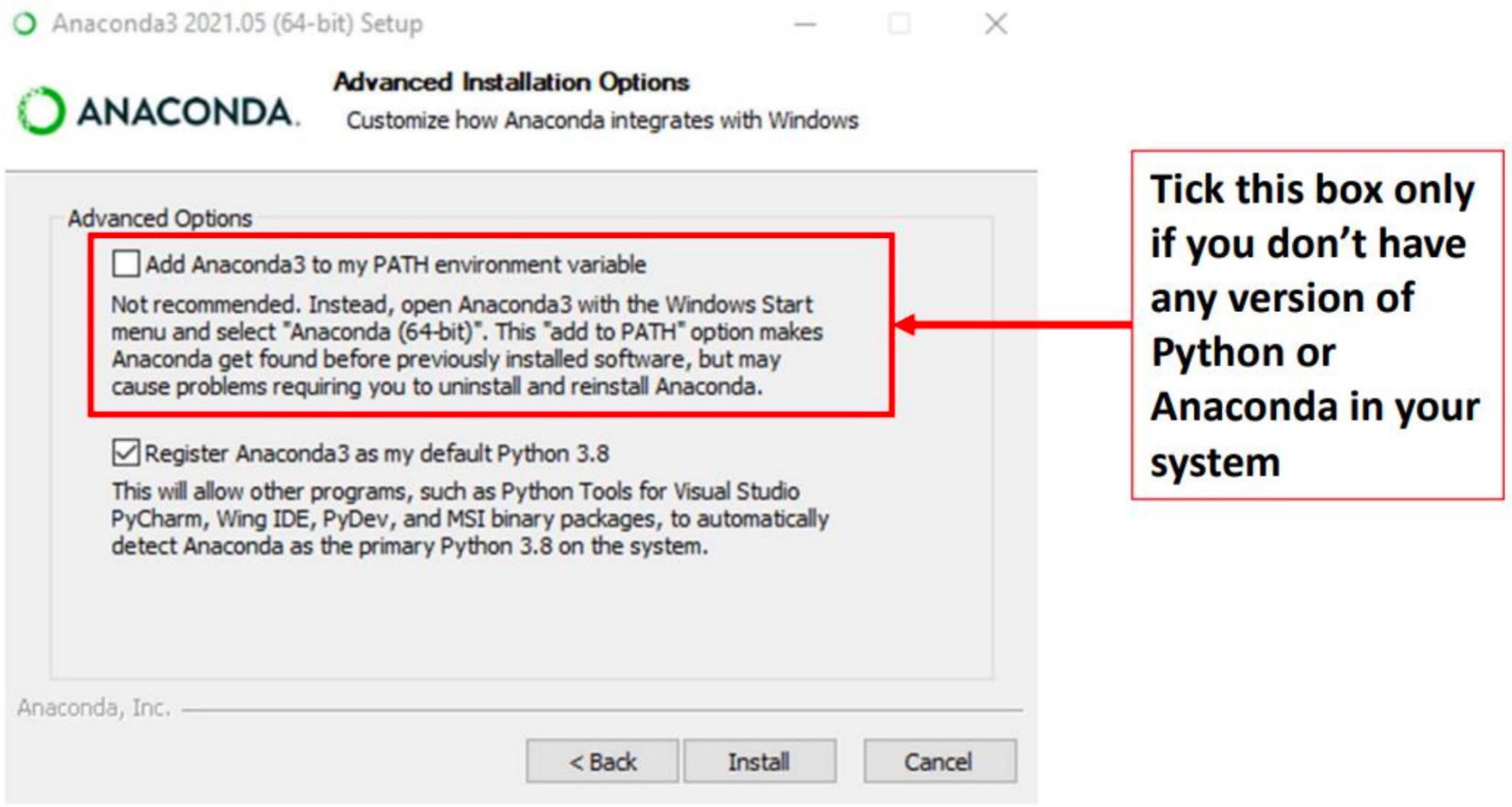
- A web-based interactive computing platform
- Cell-based structure, very convenient to test code
- Often adopted by cloud computing providers

Set up your environment

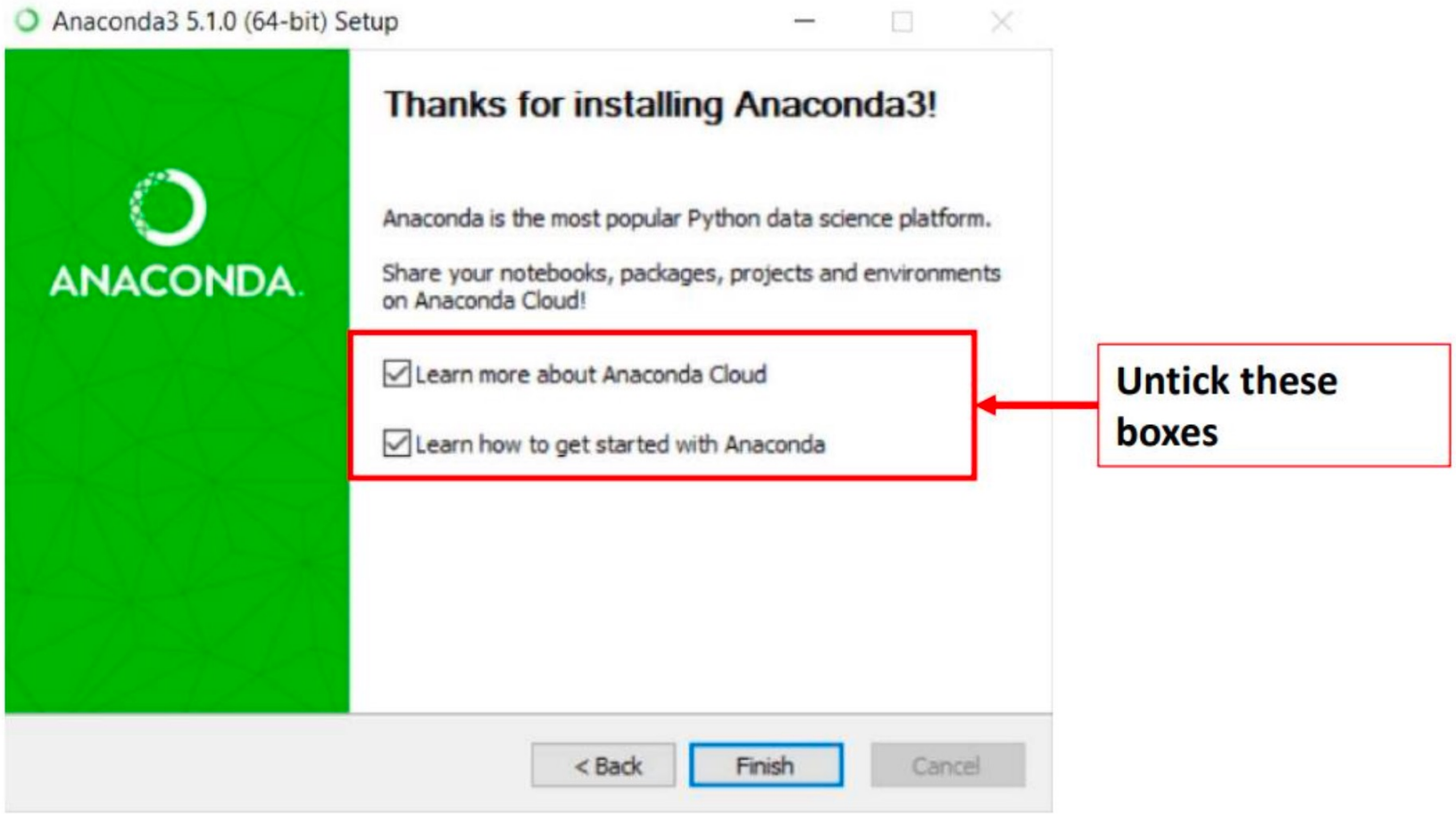
1. Installing Anaconda
2. Opening Jupyter Notebook

1. Installing Anaconda

1. Download and install Anaconda (https://www.anaconda.com)

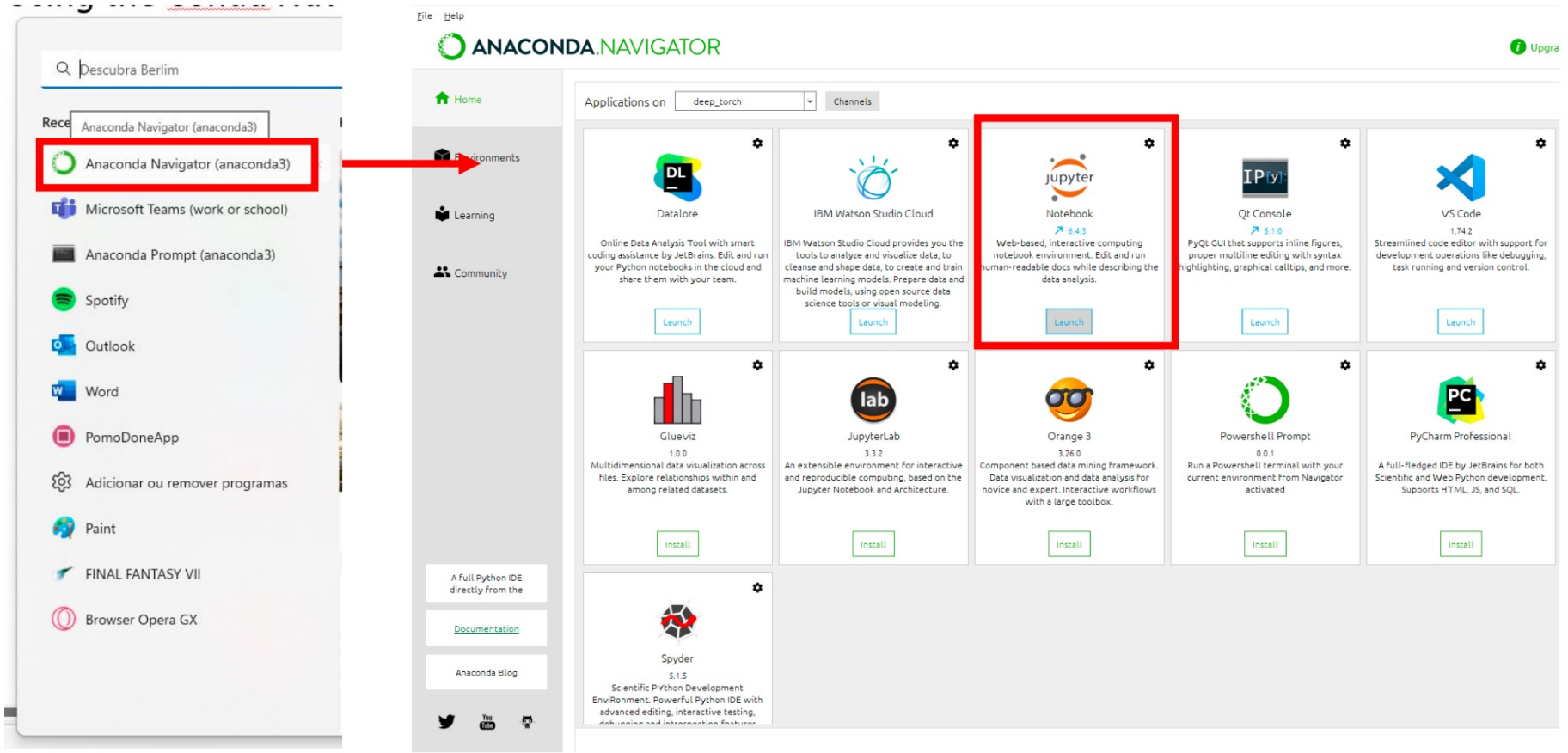


1. Installing Anaconda



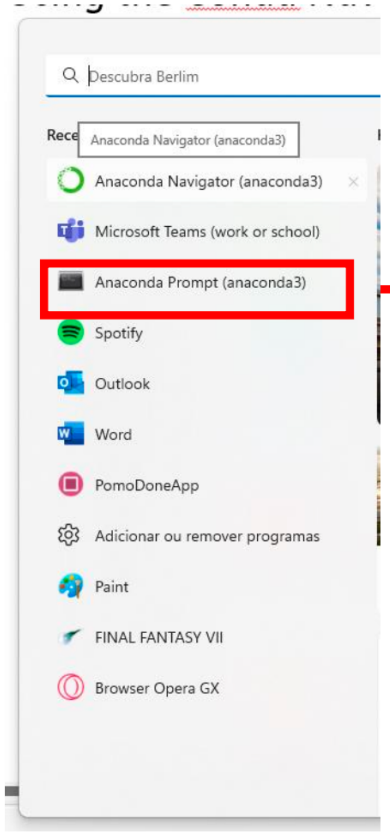
2. Opening Jupyter Notebook

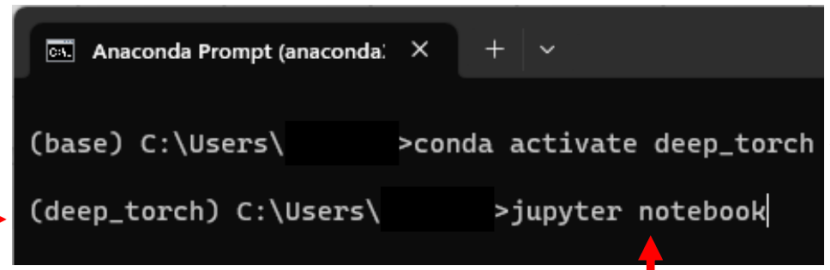
a. Using the Anaconda Navigator



2. Opening Jupyter Notebook

b. Using the Anaconda Prompt



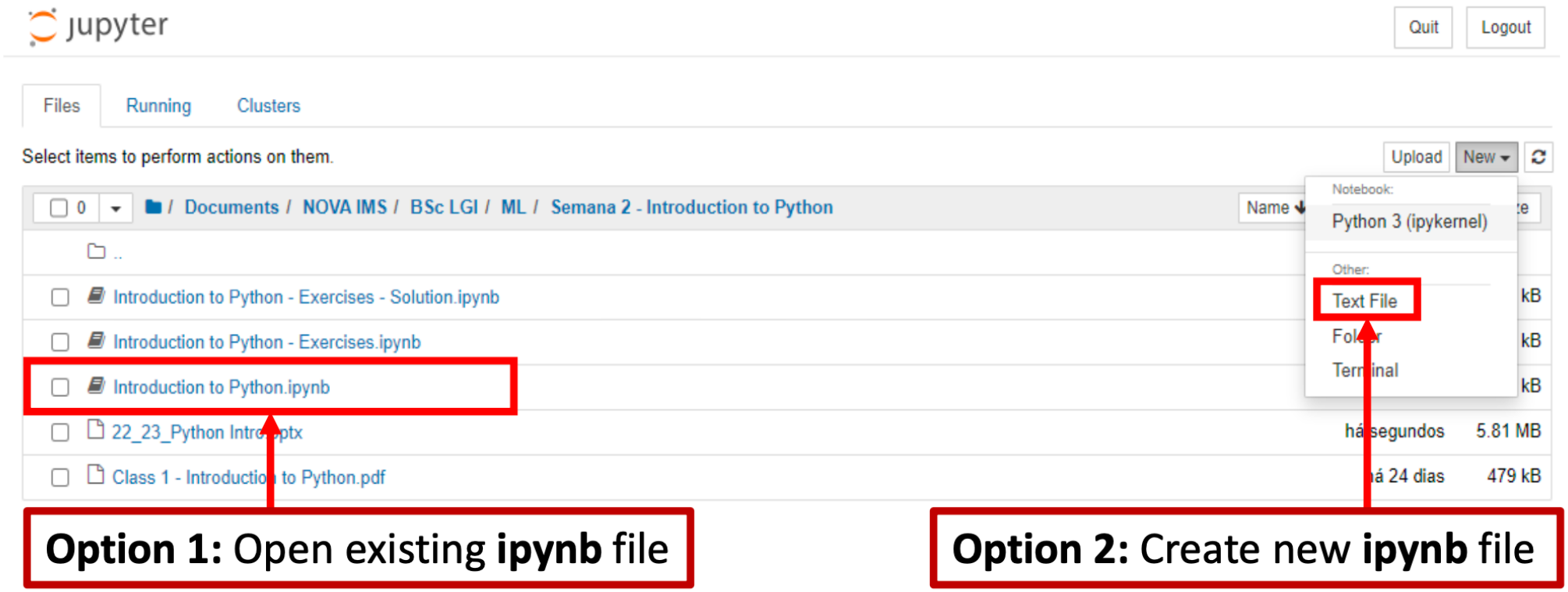


This command activates a conda environment

This command opens jupyter

2. Opening Jupyter Notebook

2. Opening Jupyter Notebook



3. Working with Jupyter Notebooks

- Markdown cell
- Code cell
- Adds one cell below
- Switches between modes (Markdown/Code)

4. Creating an Environment

Creating a new environment in Anaconda ensures **dependency isolation** and **avoids conflicts**, providing a clean and controlled workspace.

More information on setting a conda environment can be found here:
https://conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html

Tools – Jupyter Notebook

We'll use the following Python packages often:

- ***pandas*** – handles data analysis and manipulation (Excel on steroids)
- ***numpy*** – for intricate mathematical operations
- ***Matplotlib & seaborn*** – used for data visualization
- ***Scikit-learn*** – Multi-purpose package that with useful implementations of functions & Machine Learning algorithms
- **Other packages will be installed on-need**

Thank you!

Morada: Campus de Campolide, 1070-312 Lisboa, Portugal
Tel: +351 213 828 610 | Fax: +351 213 828 611

Acreditações e Certificações da NOVA IMS

Cofinanciado por