NOVA IMS Information Management School

Machine Learning
Fall Semester

Practical Classes and Anaconda

Rafael Pereira

Ricardo Santos

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

Acreditações e Certificações

Introduction

1 2 9 0

UNIVERSIDADE D COIMBRA

- **BSc in Biochemistry (2012)**
- **Postgraduation in Biochemistry (2013)**
- **BSc in Economics (2020)**

- **Msc in Data Science and Advanced Analytics (2023)**
- **PhD in Information Management (2023-2026)**

Office hours: Thursdays 3pm-4pm:

Office 136

Upon scheduling:

rcsantos@novaims.unl.pt

NOVA IMS Information Management School

Agenda

- Ground rules for practical classes
- Tools
  - Python
  - Anaconda
  - Jupyter Notebook
- Setting up your environment

Ground rules of practical classes

- Each group has 1 practical classes per week
- In each week, 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 you are assigned to**

Grading

The practical component of the course will have one final group project.

Group Project (30%)

- Practical implementation of ML to solve a classification/regression problem
- You will receive a project specification, some labelled data (for training) and unlabelled data (for testing)
- You will need to follow the steps of a ML project to create a predictive model with your training data and use that model to make predictions with your test data
- We will use machine learning to form the groups (each group will be made out of 4 to 5 elements
- More details in a couple of weeks

Disclaimers

- Participation in the project is required to obtain approval in the course
- You will need to attend a project defense at the end of the semester
- Minimum grade in project is 8.00 (out of 20)

NOVA IMS Information Management School

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***[...]

NOVA IMS Information Management School

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**

NOVA IMS Information Management School

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

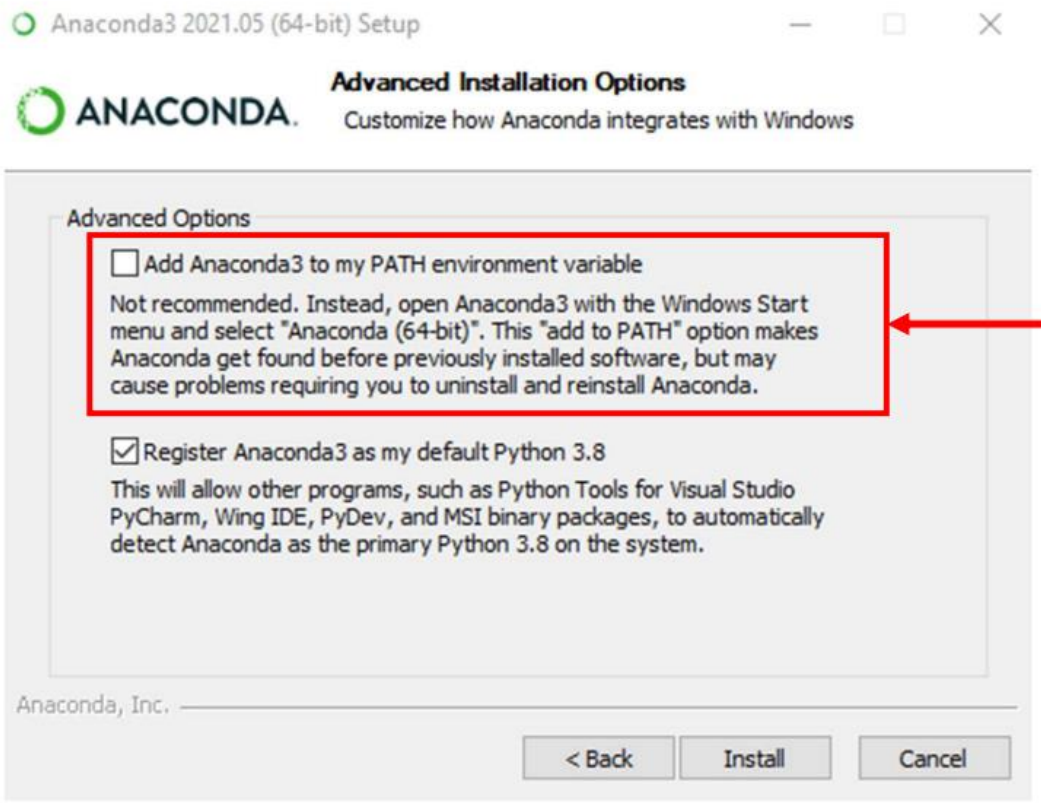
NOVA IMS Information Management School

Set up your environment

1. Installing Anaconda
2. Opening Jupyter Notebook

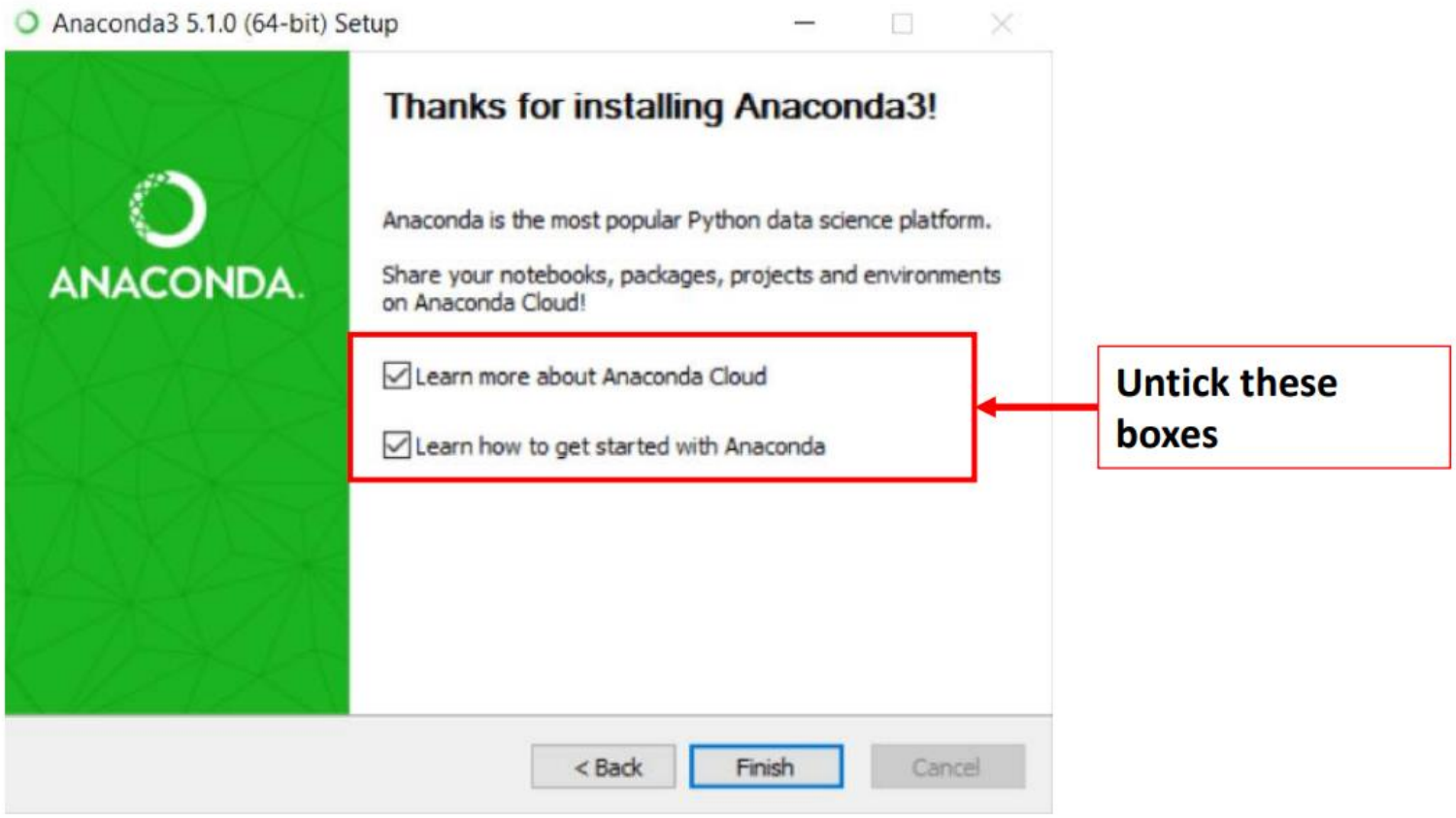
1. Installing Anaconda

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



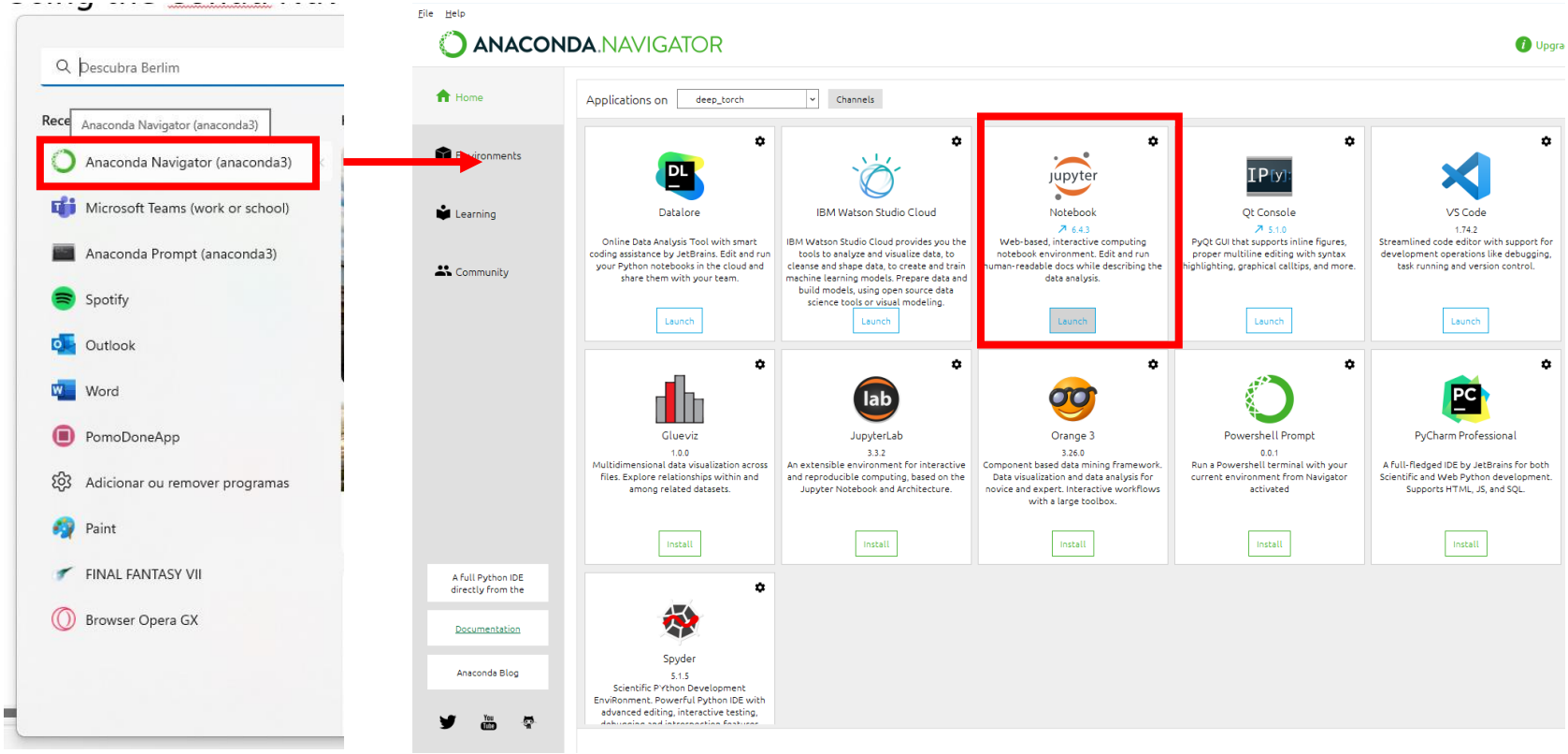
Tick this box only if you don't have any version of Python or Anaconda in your system

1. Installing Anaconda



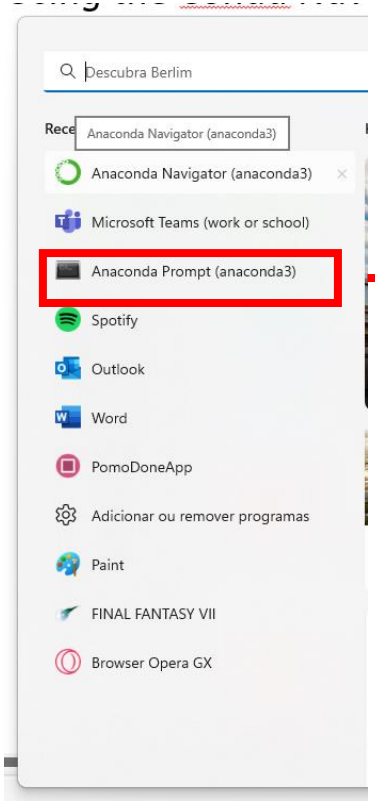
2. Opening Jupyter Notebook

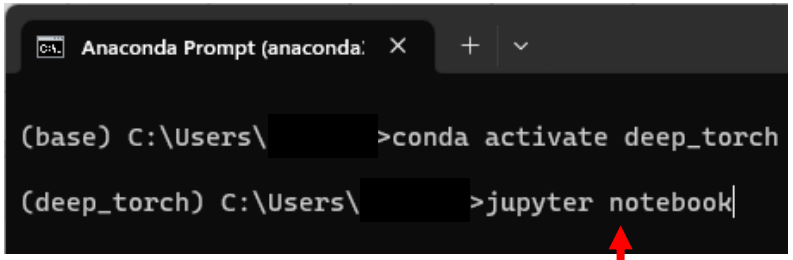
a. Using the Anaconda Navigator



2. Opening Jupyter Notebook

b. Using the Anaconda Prompt



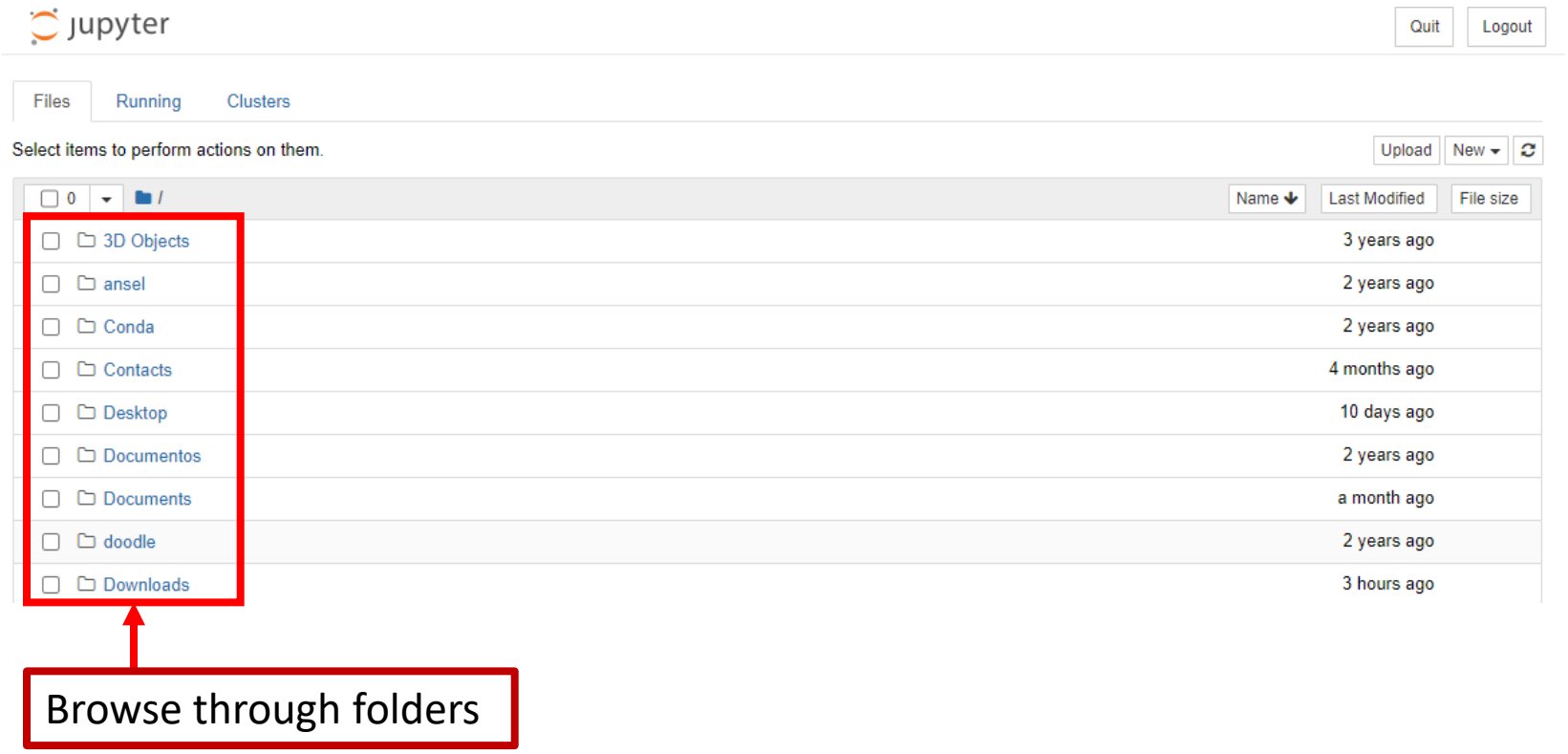


This command activates a conda environment

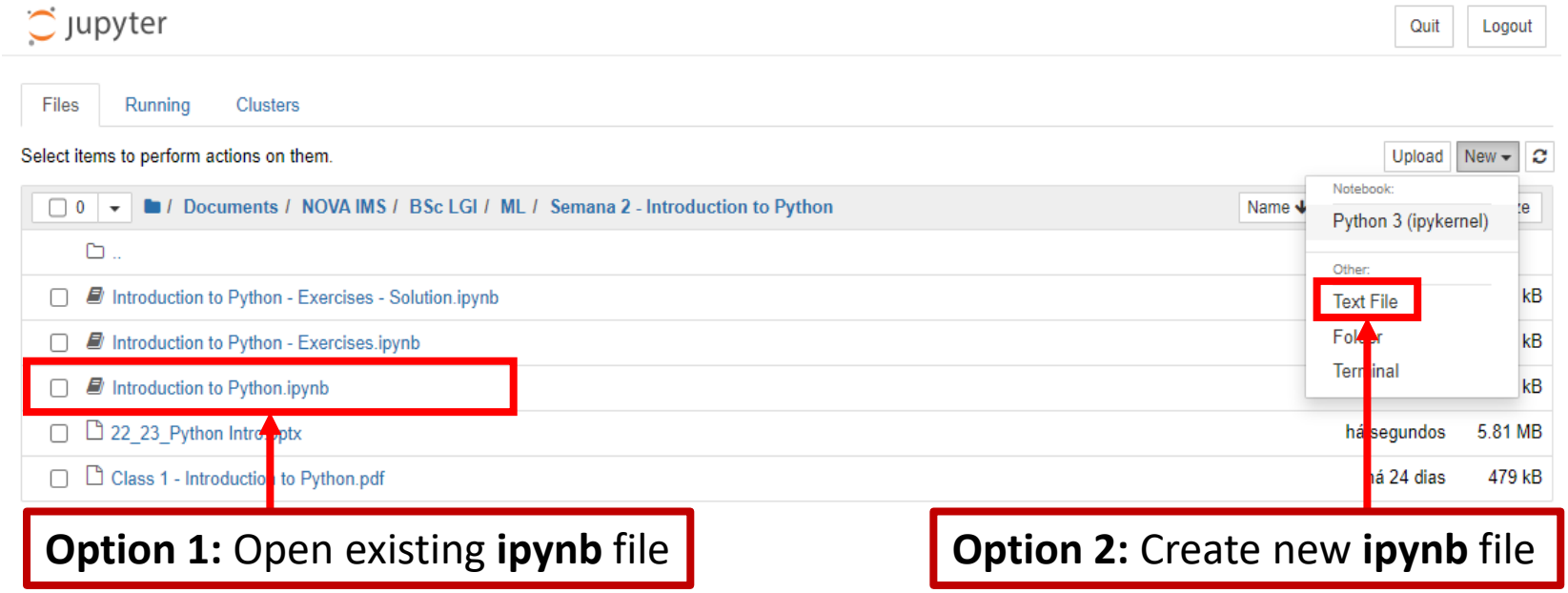
This command opens jupyter

More information on setting a conda environment can be found here: https://docs.conda.io/projects/conda/en/latest/user-guide/getting-started.html#managing-environments

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)

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

Address: Campus de Campolide, 1070-312 Lisboa, Portugal
Phone: +351 213 828 610 Fax: +351 213 828 611

Acreditações e Certificações

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