

NOVA

IMS

Information
Management
School

Machine Learning Fall Semester

Practical Classes and Anaconda

Rafael Pereira

Ricardo Santos

Introduction



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



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)

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



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



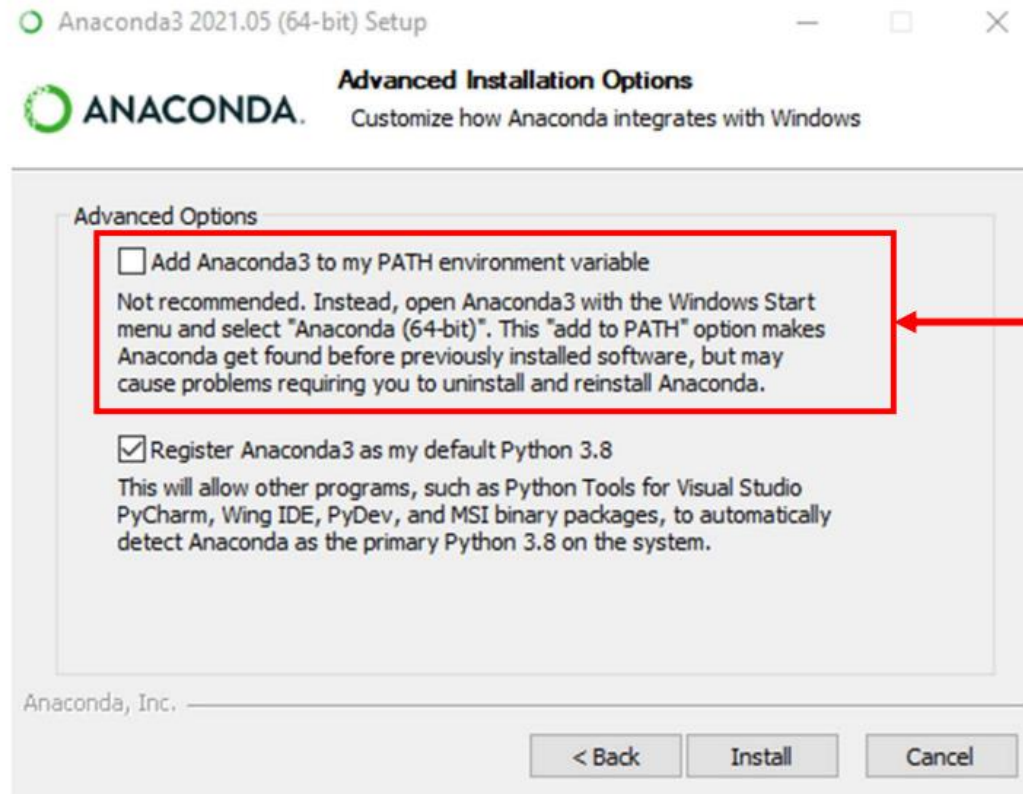
- A web-based interactive computing platform
- Cell-based structure, very convenient to test code
- Often adopted by cloud computing providers

Set up your environment

1. Installing Anaconda
2. Opening Jupyter Notebook

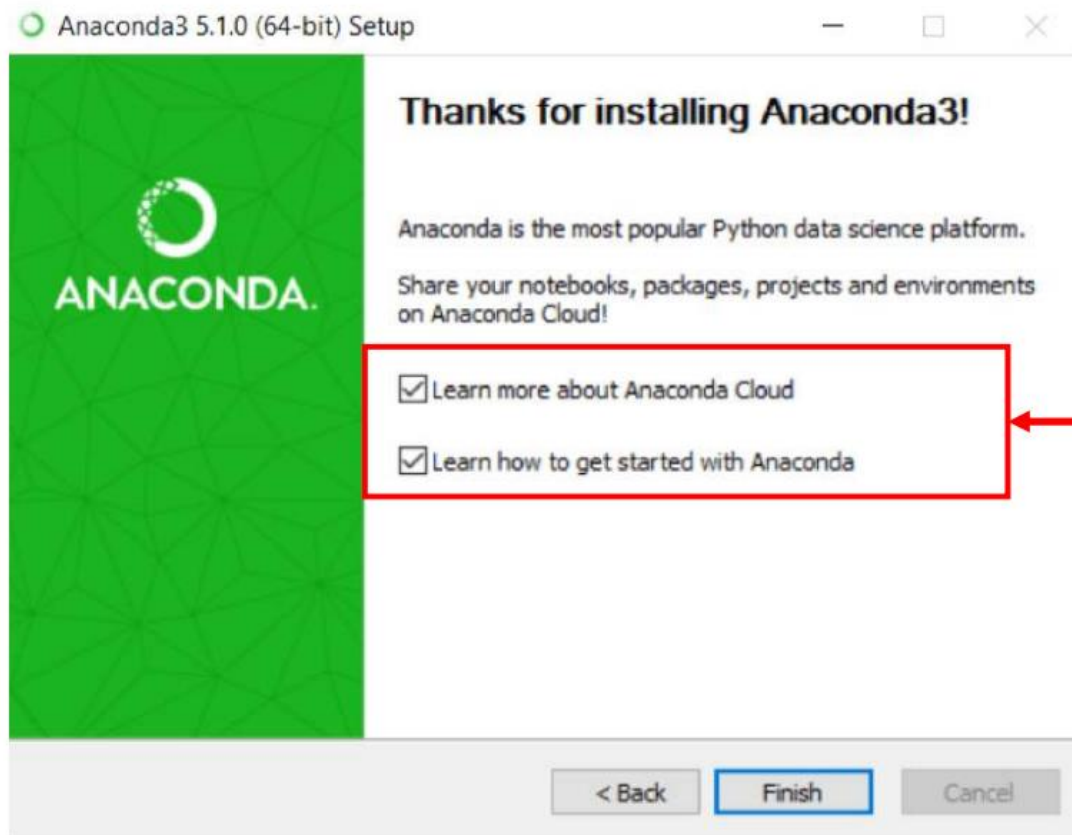
1. Installing Anaconda

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



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

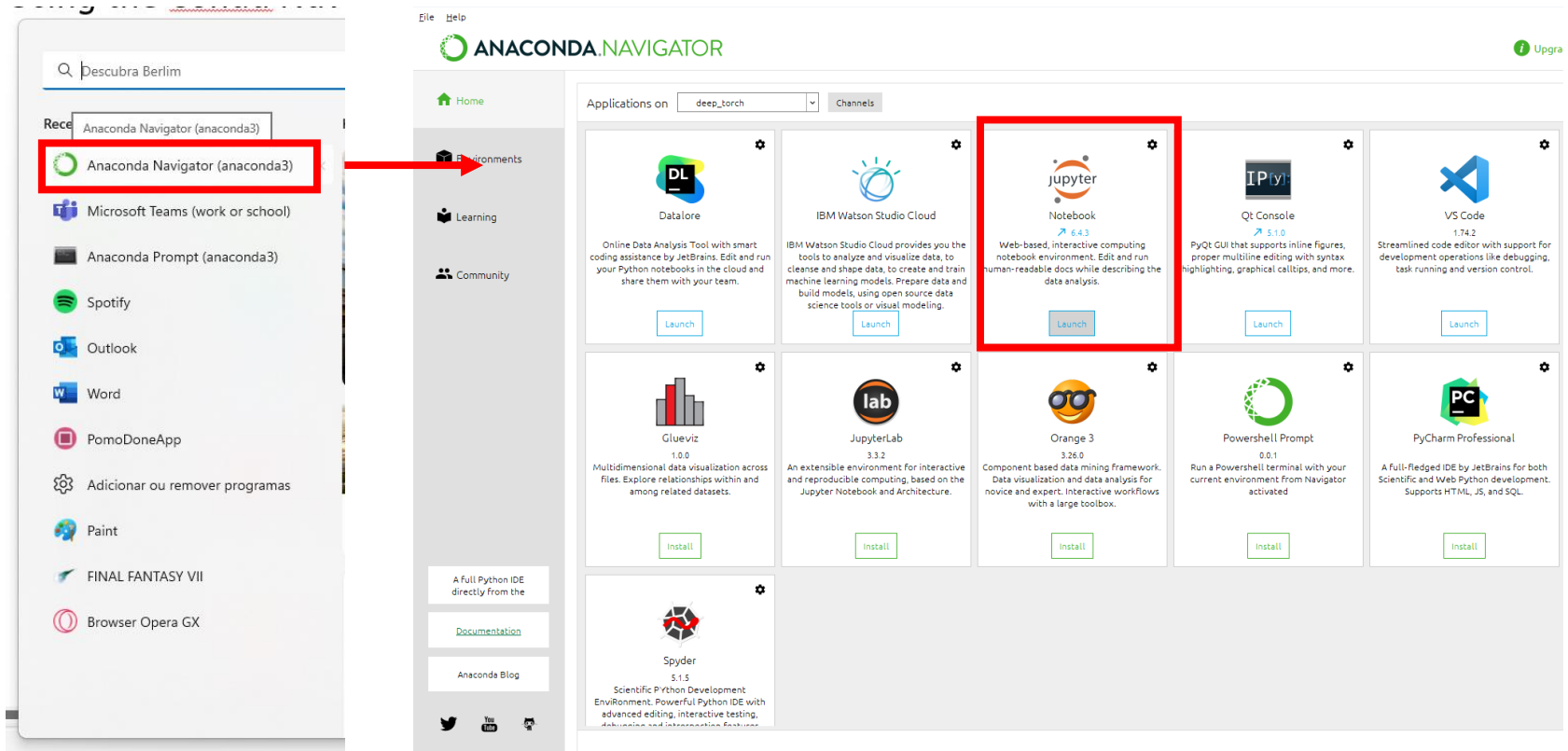
1. Installing Anaconda



Untick these
boxes

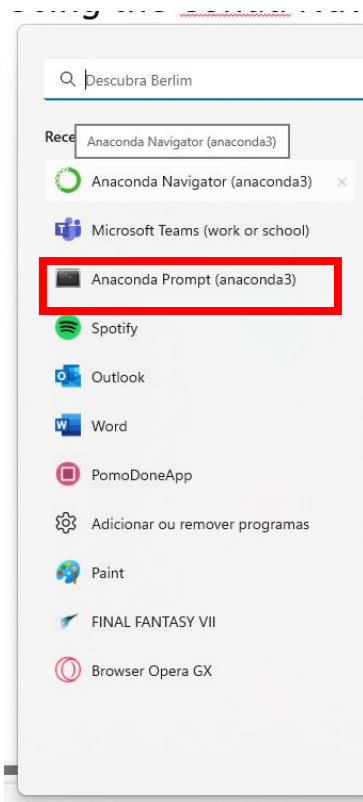
2. Opening Jupyter Notebook

a. Using the Anaconda Navigator



2. Opening Jupyter Notebook

b. Using the Anaconda Prompt



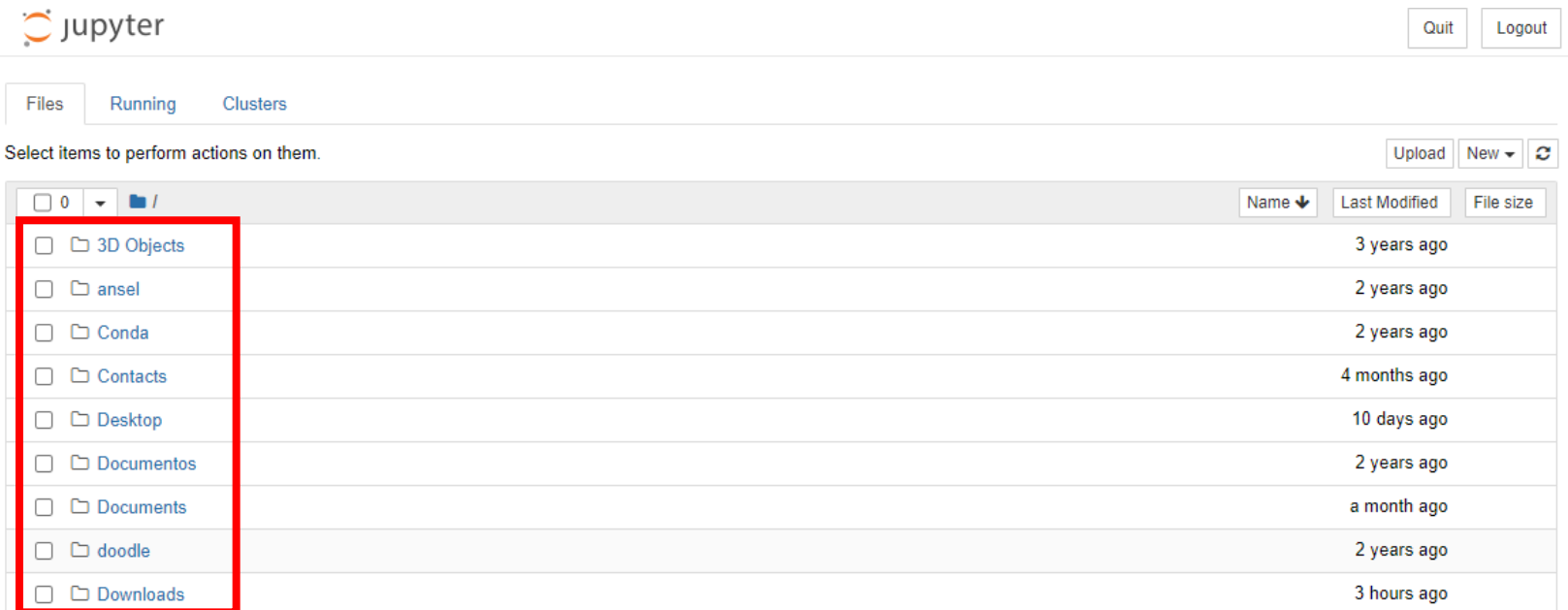
```
Anaconda Prompt (anaconda: x + v  
(base) C:\Users\[redacted] >conda activate deep_torch  
(deep_torch) C:\Users\[redacted] >jupyter notebook|
```

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



The screenshot shows the Jupyter web interface. At the top, there's a 'jupyter' logo and 'Quit' and 'Logout' buttons. Below that are tabs for 'Files', 'Running', and 'Clusters'. A message says 'Select items to perform actions on them.' with 'Upload', 'New', and a refresh icon. The main area is a file browser showing a list of folders. A red box highlights the following folders: 3D Objects, ansel, Conda, Contacts, Desktop, Documentos, Documents, doodle, and Downloads. A red arrow points from a text box below to this list.

| | Name | Last Modified | File size |
|--------------------------|------------|---------------|-----------|
| <input type="checkbox"/> | 0 | | |
| <input type="checkbox"/> | / | | |
| <input type="checkbox"/> | 3D Objects | 3 years ago | |
| <input type="checkbox"/> | ansel | 2 years ago | |
| <input type="checkbox"/> | Conda | 2 years ago | |
| <input type="checkbox"/> | Contacts | 4 months ago | |
| <input type="checkbox"/> | Desktop | 10 days ago | |
| <input type="checkbox"/> | Documentos | 2 years ago | |
| <input type="checkbox"/> | Documents | a month ago | |
| <input type="checkbox"/> | doodle | 2 years ago | |
| <input type="checkbox"/> | Downloads | 3 hours ago | |

Browse through folders

2. Opening Jupyter Notebook

The screenshot shows the JupyterLab interface. At the top, there's a 'jupyter' logo and 'Quit' and 'Logout' buttons. Below that, there are tabs for 'Files', 'Running', and 'Clusters'. The 'Files' tab is active, showing a file browser. The breadcrumb path is '/ Documents / NOVA IMS / BSc LGI / ML / Semana 2 - Introduction to Python'. The file list includes:

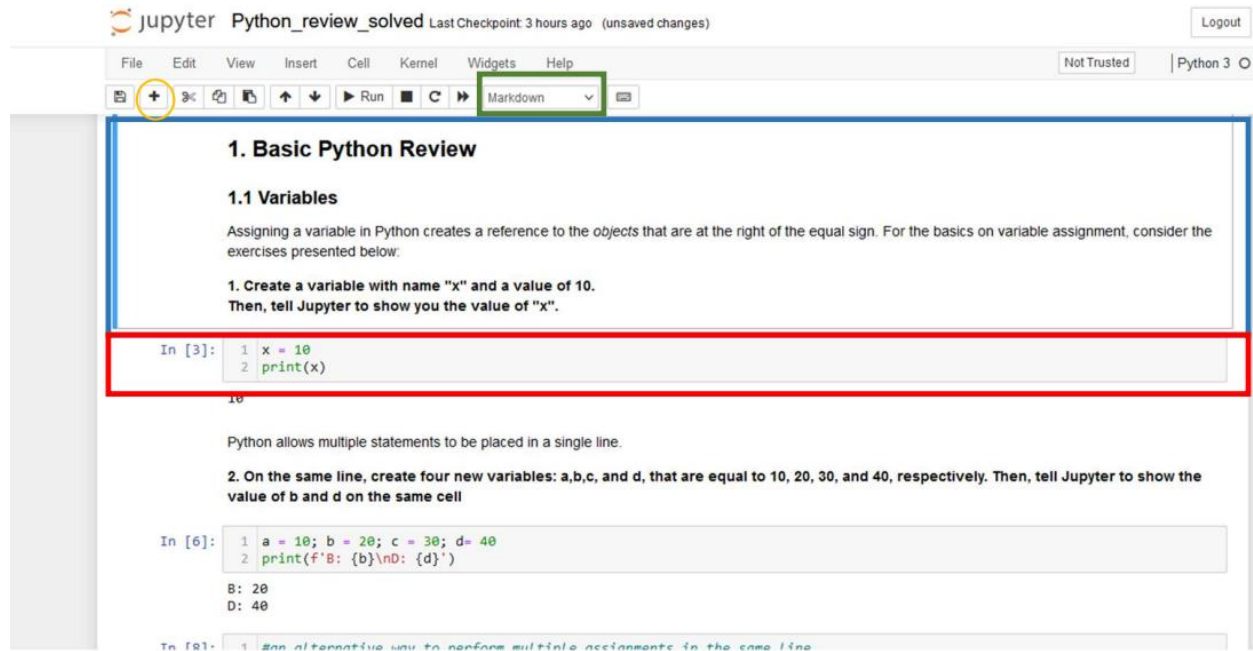
- 0
- ..
- Introduction to Python - Exercises - Solution.ipynb
- Introduction to Python - Exercises.ipynb
- Introduction to Python.ipynb (highlighted with a red box)
- 22_23_Python Intro.pptx
- Class 1 - Introduction to Python.pdf

On the right, there are 'Upload' and 'New' buttons. The 'New' button has a dropdown menu open, showing 'Notebook: Python 3 (ipykernel)' and 'Other: Text File' (highlighted with a red box), 'Folder', and 'Terminal'. A red arrow points from the 'Text File' option to a text box labeled 'Option 2: Create new ipynb file'. Another red arrow points from the 'Introduction to Python.ipynb' file to a text box labeled 'Option 1: Open existing ipynb file'.

Option 1: Open existing ipynb file

Option 2: Create new ipynb file

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



UNIGIS



A3ES



iSchools



Computing
Accreditation
Commission



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