Machine Learning

Laboratory on scikit-learn

Samuele Bortolotti Steve Azzolin

University of Trento

12 November, 2025

Set-up on Google Colab

If you prefer to run the notebook in the cloud, use the `Google Colab` link provided in the repository or scan the QR code below.

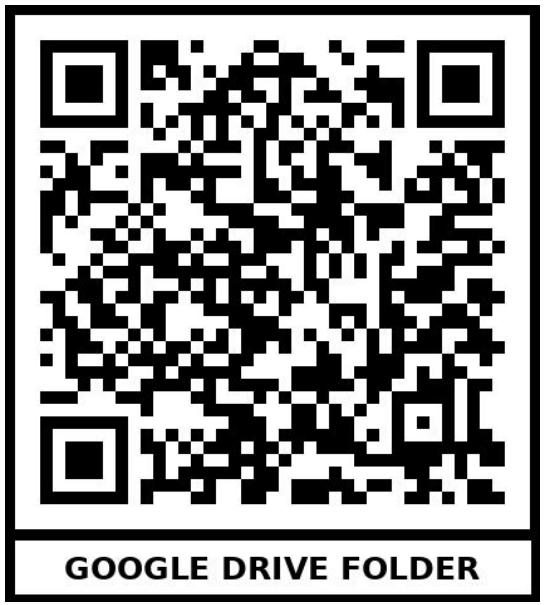

Figure 1: QR code to open Colab

No local installation is required; everything runs in your browser.

UNIVERSITÀ DI TRENTO

Preliminaries (on your PC)

1. Download the Material

Download or clone the course repository from:

https://github.com/samuelebortolotti/machine-learning-lab

UNIVERSITÀ DI TRENTO

Preliminaries (on your PC)

2. Install Jupyter Notebook

Open your terminal and install Jupyter using `pip`:

```
pip install jupyter
```

Note: If you are using Anaconda, Jupyter is already included.

Preliminaries (on your PC)

3. Navigate to the Project Folder

Use the terminal to move into the `scikit` directory:

```
cd path/to/scikit
```

Then launch Jupyter Notebook:

```
jupyter-notebook
```

UNIVERSITÀ DI TRENTO

4. Open Jupyter in Your Browser

Once Jupyter starts, your browser will open automatically showing the current working directory.

Locate and open the notebook file:

`scikit-learn.ipynb`

Figure 2: Jupyter file browser

UNIVERSITÀ DI TRENTO

Running Cells in Jupyter

To execute a cell, you can either:

- Click the **Run** button in the toolbar, or
- Press **Shift** + **Enter** on your keyboard.

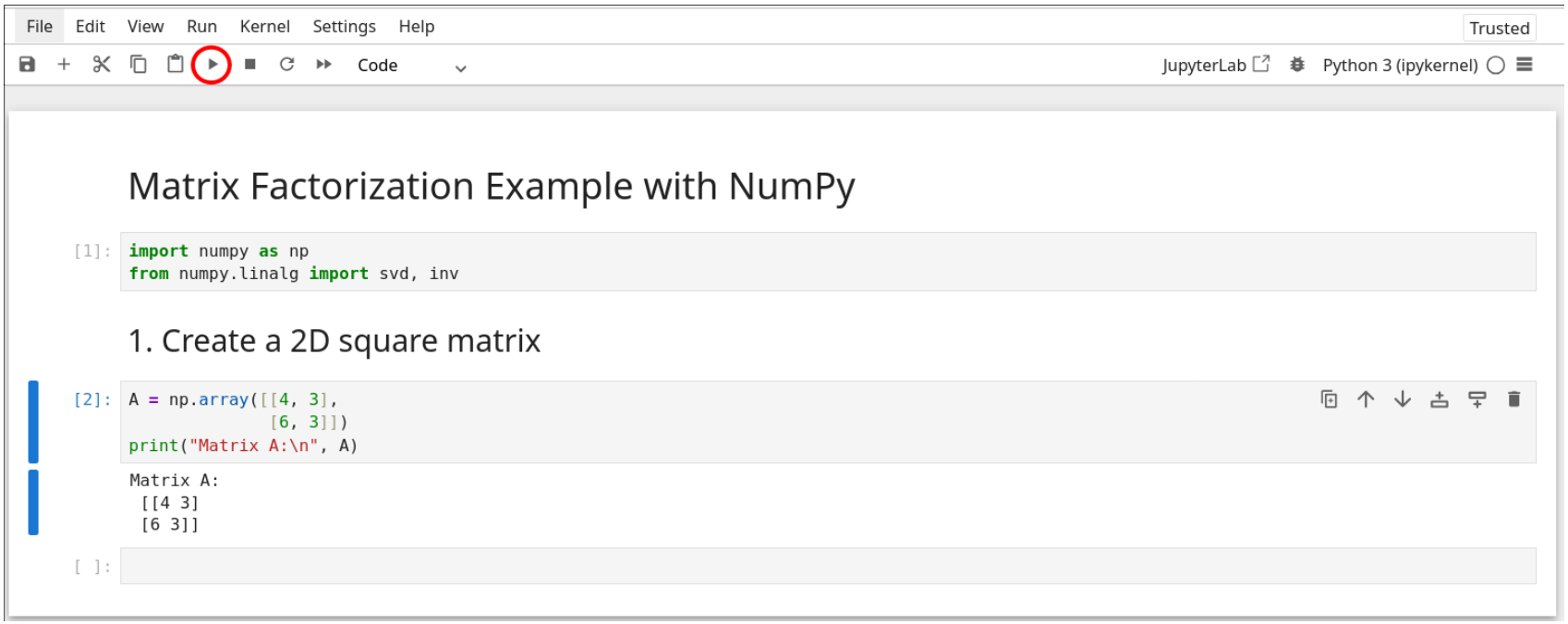

Figure 3: Running a cell in Jupyter

Take-home Exercise

As a self-assessment exercise, complete a classification task using **Scikit-Learn**.

Steps:

1. Choose a dataset.
2. Train and tune a classifier (e.g., perform hyperparameter optimization).
3. Evaluate the model on the test set.
4. Compare performance across different classifiers.

UNIVERSITÀ DI TRENTO

Suggested Datasets

You can find datasets in the UCI Machine Learning Repository:

- Spambase
- Optical Recognition of Handwritten Digits
- Abalone
- Additional datasets available at https://archive.ics.uci.edu/

Model Comparison

Compare the performance of multiple classifiers such as:

- Decision Tree
- Random Forest
- SVM

Perform hyperparameter tuning and report the evaluation metrics (e.g., accuracy, precision, recall, F1-score).

UNIVERSITÀ DI TRENTO