Learning to Use the 'Scikit-Learn' Library

Saptarshi Pyne, Assistant Professor
Department of Computational and Data Sciences (CDS)
Indian Institute of Science Education and Research (IISER) Kolkata
Mohanpur, West Bengal, India 741246

CS5103 Applied Machine Learning Lecture 2
August 5, 2025

What is Scikit-Learn

"Scikit-learn is an open source machine learning library that supports supervised and unsupervised learning. It also provides various tools for model fitting, data preprocessing, model selection, model evaluation, and many other utilities."

Source: https://scikit-learn.org/stable/getting_started.html

French data scientist David Cournapeau (pronounced "Courna-pooh") started it as a Google Summer of Code (GSoC) project to develop a **"scientific toolkit for machine learning"**. Slowly more developers joined in to make the Python library a huge success.

How to Install Scikit-Learn?

Official instructions:
https://scikit-learn.org/stable/install.html#installation-instructions

The aforementioned instructions create a 'virtual environment' and install scikit-learn inside this environment.

- A virtual environment is an isolated directory with:
  - Its own directory tree,
  - Its own Python version, and
  - Its own packages with their specific versions installed.

How to Install Scikit-Learn? (contd.)

The official instructions uses Python's built-in environment manager to create the virtual environment.

However, we can use specialized environment managers such as Conda for this purpose: https://docs.conda.io/projects/conda/en/stable/

Conda

Conda is a software that manages packages, environments, and dependencies for any programming language.

When we install Anaconda Navigator, conda also gets installed.

How to Use Conda

Let us open an Anaconda prompt from the Anaconda Navigator and issue the following commands.

List the existing virtual environments.
The default virtual environment is called 'base'.

> conda env list

Create a new virtual environment named 'sklearn-env'

> conda create --name sklearn-env

How to Use Conda (contd.)

Switch to 'sklearn-env' (from 'base')
> conda activate sklearn-env

Install scikit-learn inside 'sklearn-env'
> pip install -U scikit-learn

[To learn about useful Conda commands, please check out the Conda cheat sheet at
https://docs.conda.io/projects/conda/en/4.6.0/_downloads/52a95608c49671267e40c689e0bc00ca/conda-cheatsheet.pdf]

How to Use Scikit-Learn (sklearn)

Now, let us launch the 'useSklearn.ipynb' Jupyter notebook and do some machine learning with it.

Keep Calm
and
Learn Machine Learning

Thank You