Unit 1
Introduction to Machine Learning

Essential libraries and tools

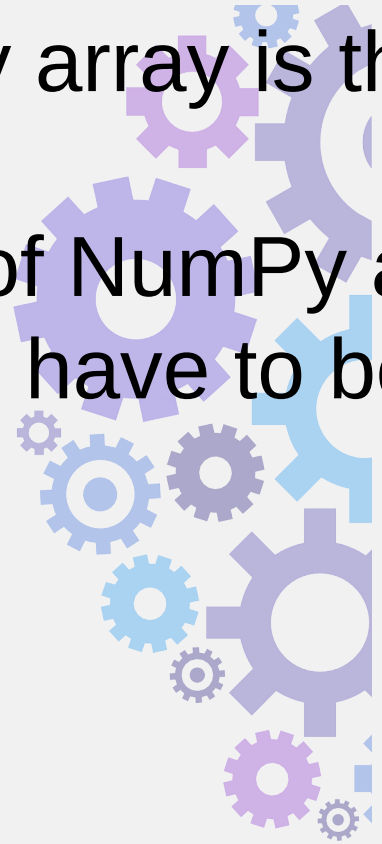
- Jupyter Notebook:
- The Jupyter Notebook is an interactive environment for running code in the browser.
- It is a great tool for exploratory data analysis and is widely used by data scientists.
- While the Jupyter Notebook supports many programming languages, we only need the Python support.

Essential libraries and tools

- NumPy:
- NumPy is one of the fundamental packages for scientific computing in Python. It contains functionality for multidimensional arrays, high-level mathematical func- tions such as linear algebra operations and the Fourier transform, and pseudorandom number generators.

Essential libraries and tools

In scikit-learn, the NumPy array is the fundamental data structure. scikit-learn
takes in data in the form of NumPy arrays.
Any data you're using will have to be converted to a NumPy array.



Essential libraries and tools

- SciPy
- SciPy is a collection of functions for scientific computing in Python.
- It provides, among other functionality, advanced linear algebra routines, mathematical function optimization, signal processing, special mathematical functions, and statistical distributions. scikit-learn draws from SciPy’s collection of functions for implementing its algorithms.

Essential libraries and tools

- Matplotlib:
- matplotlib is the primary scientific plotting library in Python. It provides functions for making publication-quality visualizations such as line charts, histograms, scatter
- plots, and so on.
- Matplotlib is used for all our visualizations.
- When working inside the Jupyter Notebook, you can show figures directly in the
- browser by using the %matplotlib notebook and %matplotlib inline commands.

Essential libraries and tools

- Pandas:
- Pandas is a Python library for data wrangling and analysis. It is built around a data structure called the DataFrame that is modeled after the R DataFrame.
- Simply put, a pandas DataFrame is a table, similar to an Excel spreadsheet. pandas provides a great range of methods to modify and operate on this table; in particular, it allows SQL-like queries and joins of tables.

Essential libraries and tools

- In contrast to NumPy, which requires that all entries in an
- array be of the same type, pandas allows each column to have a separate type (for
- example, integers, dates, floating-point numbers, and strings). Another valuable tool
- provided by pandas is its ability to ingest from a great variety of file formats and data-
- bases, like SQL, Excel files, and comma-separated values (CSV) files