

Machine Learning

Python Libraries

Curtis Larsen

Utah Tech University—Computing

Spring 2025

Objectives

Objectives:

- ▶ Control Python with `venv`
- ▶ Create programs with `numpy`
- ▶ Create programs with `pandas`
- ▶ Create programs with `matplotlib`
- ▶ Create programs with `scikit-learn`

Python Virtual Environments

Why?

Why Virtual Environments?

- ▶ Controls Python packages
- ▶ Avoids package conflicts between projects
- ▶ Makes project prerequisites repeatable

Documentation

How?

How Virtual Environments?

- ▶ `echo ".venv" >> .gitignore` - *Do not commit the derived files*
- ▶ `python3 -m venv .venv` - create a location to store the configuration
- ▶ `source .venv/bin/activate` - activate the environment
- ▶ Store list of packages in `requirements.txt`
- ▶ `pip install -r requirements.txt` - installs packages in the virtual environment
- ▶ `./program.py` - uses the virtual environment to find python related content
- ▶ `deactivate` - deactivate the environment

NumPy

NumPy

Why NumPy?

- ▶ Efficient storage
- ▶ Efficient manipulation
- ▶ Compatibility with other libraries

Documentation

Pandas

Pandas

Why Pandas?

- ▶ Spreadsheet-like data structure
- ▶ `.csv` file format
- ▶ Named columns
- ▶ Queries

Documentation

matplotlib

matplotlib

Why matplotlib?

- ▶ Visualize data
- ▶ Interactive
- ▶ Image/Figure generation

Documentation

scikit-learn

scikit-learn

Why scikit-learn?

- ▶ Prebuilt algorithms/classes
- ▶ Uniform API
- ▶ Classification
- ▶ Regression
- ▶ Clustering
- ▶ Dimensionality reduction
- ▶ *Not* neural networks

Documentation