Python Packages

CIS*4500 (Fall 2019)

Why Python for Machine Learning?

- Python as an interpreted language is slow in execution, but packages like NumPy and SciPy have been built to support fast and vectorized operations for multi-dimensional arrays
- Packages like Tensorflow allow parallel execution for multidimensional arrays by mapping the operations to the multi-processors in a GPU
- Many other Python packages and implementations of machine learning techniques are publicly available, making it convenient to develop new applications
- Python and the related packages are easy to learn with many references and tutorials on the Web

Two Major Versions of Python

- Latest releases are Python 2.7.15 and Python 3.7.2, both of which are available for Windows, macOS, and Linux at https://www.python.org/

- A good summary of the differences between the two versions is given at https://wiki.python.org/moin/Python2orPython3

- We recommend the latest Python 3, and once installed, use the "pip" utility that comes with Python 3 to install and upgrade the related packages. On a mac machine, do the following:
  - Pip install <package-name>
  - Pip install <package-name> --upgrade
  - Pip uninstall <package-name>

Scikit-Learn Package for ML

- To use scikit-learn for ML implementations, install the following packages with the given versions or higher:
  - NumPy 1.12.1 (store and manipulate multidimensional arrays)
  - SciPy 0.19.0 (use NumPy for scientific functions)
  - Scikit-learn 0.18.1 (support many machine learning algorithms)
  - Matplotlib 2.0.2 (visualize data and results)
  - Pandas 0.20.1 (built on top of NumPy for high-level support)

VirtualEnv for Python & Packages

- Certain implementations require different Python and package versions.

- Create a virtual environment for such an implementation as follows:
  - pip install virtualenv
  - virtualenv –p python <venv-dir>
  - source <venv-dir>/bin/activate
  - … (do things in this virtual environment)
  - deactivate (exit the virtual environment)

References for Python & Packages

- Sebastian Raschka and Vahid Mirjalili. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. Second Edition. Packt Publishing, 2017
  - https://github.com/rasbt/python-machine-learning-book-2nd-edition

- Allen Downey, Jeffrey Elkner, and Chris Meyers. How to Think like a Computer Scientist: Learning with Python. Green Tee Press, 2002
  - http://greenteapress.com/thinkpython/thinkCSpy/thinkCSpy.pdf

- Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python: Analyzing Text with Machine Learning Toolkit. O'Reilly Media, 2009
  - http://www.nltk.org/book/

Sample Python Programs

- Many examples are available in Raschka and Mirjalili (2017)

- ch03a.py: a sample program taken and adapted from Raschka and Mirjalili (2019) that classifies iris.txt with Logistic Regression, SVM's, Decision Trees, and K-Nearest Neighbors in Scikit-learn