

Additional Resources

Python Resources:

To be successful in the course, you will need an understanding of [Python](#) and the packages commonly used in data science:

[NumPy](#) - the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

[Pandas](#) - an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

[matplotlib](#) - a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.

[scikit-learn](#) - Scikit-learn remains very popular in practice for trying out canonical machine learning techniques, particularly for applications in experimental science and in data science. A lot of what we provide is now very mature. But it can be costly to maintain, and we cannot therefore include arbitrary new implementations. Yet Scikit-learn is also essential in defining an API framework for the development of interoperable machine learning components external to the core library.

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Downloadable Reference Materials:

- [Computer Age Statistical Inference \(CASI\)](#)
- [Elements of Statistical Learning \(ESL\)](#)
- [An Introduction to Statistical Learning \(ISLR\)](#)
- [Machine learning materials with Stanford professors Trevor Hastie and Robert Tibshirani](#)