A Brief Intro to Scikit-learn

Outline

- Introduction
- Supervised Problem - Random Forest
- Unsupervised - Color Space Compression

What is scikit-learn

- Collection of machine learning algorithms and utilities
  - SVM, RF, K-Means, MoG, manifold learning, all sorts of regression
  - See http://scikit-learn.org/stable/modules/classes.html
- Python based - numpy, scipy, matplotlib
  - Some library wrappers + Cython
- Built in parallelization via joblib!
- BSD + MIT licenses

Supervised Problem - Digits

- Built in data sets
- Random forests/classification
- Cross validation
- Grid search
- Parallelization

Unsupervised Problem - Color Space Compression

- K-means clustering

Questions?

- Additional Info
  - Jake van der Plas tutorial
    http://astronml.github.com/sklearn_tutorial
  - Additional Sklearn tutorials
    http://scikit-learn.org/stable/tutorial/basic/tutorial.html
    http://scikit-learn.org/stable/tutorial/statistical_inference/index.html