

Session 04

A first look at scikit-learn

What is `scikit-learn`?

- *The* python machine learning library
- Focus: ML, *not* stats (i.e. won't get error estimates on model parameters)
- One package to rule them all
 - Feature engineering
 - Models
 - Supervise, Unsupervised
 - Linear, SVM, Trees, Ensembles, Perceptron, ...
 - Model evaluation
 - Train test split, k-fold, ...

- Metrics
- Pipelines ...
- Alternatives
 - Statsmodels (for more stats based approach)
 - Tensorflow/Keras (for serious neural nets)

What we will learn today

- KNN model
- Feature engineering
- Fitting the model
- Evaluating the model
- Exercises