

# Machine Learning Practice

## DSA 5970 / CS 5970

Dr Andrew H. Fagg  
Fall, 2019

Machine learning is the data-driven process of constructing mathematical models that can be predictive of data observed in the future. In this course, we will study the use of a range of supervised, semi-supervised and unsupervised methods to solve both classification and regression problems. In particular, we will focus on methods that can robustly address data that are non-linear, noisy, heterogeneous and/or high-dimensional. We will also study methods for evaluation of the resulting models. In our homework and project work, we will make use of several python-based tool kits, including Scikit-Learn and Keras/TensorFlow.

## Topics

Topics will include:

- Decision trees: ensemble methods, random forests, and boosting
- Regression and combating overfitting: ridge regression, Tikhonov regression, lasso, elastic nets, support vector regression
- Nonlinear dimensionality reduction: kernel PCA, local linear embedding, ISOMap, multidimensional scaling
- Semi-supervised learning: label spreading, label propagation
- Unsupervised learning: spectral clustering, mixture models, DBSCAN
- Evaluation in ML: metrics, cross-validation, statistics, addressing the multiple comparisons problem

## Resources

- Text Book: Aurélien Géron (2019) **Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow**, 2nd Edition, O'Reilly Media, ISBN-13: 978-1492032649

## Prerequisites

- Artificial Intelligence (CS 4013/5013) OR Data Mining (CS 5593) OR Intelligent Data Analytics (ISE/DSA 5103).