Applied Machine Learning Syllabus

Economics 213R

Professor: Tyler Folkman

Contact Information

I have created a slack group that we will use for most communication. I will provide more information on how to join and use slack when the course starts.

You can also always reach me at tylerfolkman at gmail

Office Hours

To be determined. I will have an hour twice a week for office hours, but I want to discuss with the class the best times. My office hours will most likely be remote using Zoom video conference.

Grading

Final Project: 30%

Homeworks: 50%

Participation: 10%

Quizzes: 10%

Grading will be done on a curve and not upon absolute thresholds.

Class Material

I will be leveraging material from many sources as well as my own content. Below is a listed of sources I used. Many are freely (and legally) available online. Not required to purchase any, though, I find [1] to be quite good as well as [4]. [4] is available for free online.

1. Hands-On Machine Learning with Scikit-Learn and Tensorflow
2. Data Science From Scratch
3. Python for Data Analysis
4. Introduction to Statistical Learning
5. Coursera: Machine Learning
6. Fastai course 1
7. https://distill.pub/2016/misread-tsne/

Course Description

The amount of data being created in the world is staggering. More data has been created in the past two years than in the entire previous history of the human race and companies are taking notice. Some of the largest companies in the world - Google, Microsoft, Facebook, and Amazon to name a few - are leveraging these data to create valuable products. Amazon's Alexa, Google's search, Facebook's timeline are all powered by machine learning. These same companies are also scrambling to get as much machine learning talent as they can.

This course will teach you the fundamental building blocks of machine learning. We will go from how to gather and clean data to writing Python code to create and evaluate our own models. We will cover enough theory to ensure an understanding of machine learning models, but the focus is on the application of machine learning. To that end, we will also cover machine learning in industry and how to avoid some of the pitfalls that cause machine learning initiatives to fail.

The course will use Python heavily and while not a pre-requisite any previous experience with coding and/or Python will be very helpful.

Prerequisites

Econ 110 and Econ 378

Tentative Outline

Introduction to Machine Learning

1. What, why, how, and challenges
2. Differences between stats, ML, and econometrics

References: [1] Chapter 1

Introduction to Python

1. Functions, Lists, Dictionaries, Counters, Sets, List Comprehensions, Generators
2. Data science stack: numpy, pandas, scikit-learn, scipy, jupyter notebooks
3. Git and github

References: [2] Chapter 2, [3] All Chapters

How to get and clean data

1. Loading from text, SQL, and web pages
2. Cleaning data: missing, outliers, scaling data, reformatting

References: [3] Chapters 5-7, [2] Chapters 9-10 and 23, [1] Chapter 2

How to visualize and describe data

1. Data description techniques: mean, median, percentiles, correlations
2. Data visualization techniques: various plots, Bokeh

References: [1] Chapter 2, [2] Chapters 3 and 5, [3] Chapters 5, 8 and 9

Regression

1. Linear Regression
2. Logistic Regression
3. Regularized Regression

References: [2] Chapters 14-16, [1] Chapter 4, [4] Chapter 3, [5] Weeks 1-3

Gradient Descent

1. Cost functions, learning rates, and gradients

References: [2] Chapter 8, [1] Chapter 4

How to evaluate and tune models

1. Training/Testing and Cross validation
2. MSE
3. Confusion matrix, precision and recall, ROC
4. Learning Curves
5. Bias and variance / over-fitting under-fitting
6. Tuning hyperparameters
7. Feature importances

References: [2] Chapter 11, [1] Chapter 2 and 3, [4] Chapter 2, 5, [5] Week 6

Classification Models

1. Naive Bayes, K-nearest neighbors

References: [2] Chapters 12-13, [4] Chapter 4

SVMs

1. Max-margin, kernel trick, more than 2 classes

References: [1] Chapter 5, [4] Chapter 9, [5] Week 7

Trees and Ensembles

1. Decision Trees, Random Forests, and Gradient Boosted Trees
2. XGBoost

References: [2] Chapter 17, [1] Chapters 6-7, [4] Chapter 8

Dimensionality Reduction

1. Curse of dimensionality
2. PCA and SVD
3. T-SNE

References: [1] Chapter 8, [7]. [4] Chapter 6, 10, [5] Week 8

Clustering

1. K-means, Hierarchical, DBScan

References: [2] Chapter 19, [4] Chapter 10, [5] Week 8

Recommender Systems

1. Collaborative filtering and deep collaborative filtering

References: [2] Chapter 22, [5] Week 9

Data Science at Scale

1. Parallelism - dask and blaze
2. Spark and Mapreduce
3. AWS
4. Online learning and stochastic gradient descent

References: [5] Week 10, [2] Chapter 24

Deep Learning

1. Perceptron, simple networks, and backprop
2. GPUs
3. Tensorflow, Keras, Pytorch
4. CNNs
5. RNNs and word2vec

References: [2] Chapter 18, [1] Chapters 9-16, [5] Weeks 4-5, [6] All Lessons

Deploying Machine Learning

1. Docker
2. Flask and REST Endpoints

Machine Learning in Industry

1. Guest lecturers / companies
2. Things Kaggle challenges won't teach you
3. Why most data science initiatives fail
4. Importance of communication and collaboration

University

Honor Code

In keeping with the principles of the BYU Honor Code, students are expected to be honest in all of their academic work. Academic honesty means, most fundamentally, that any work you present as your own must in fact be your own work and not that of another. Violations of this principle may result in a failing grade in the course and additional disciplinary action by the university. Students are also expected to adhere to the Dress and Grooming Standards. Adherence demonstrates respect for yourself and others and ensures an effective learning and working environment. It is the university's expectation, and my own expectation in class, that each student will abide by all Honor Code standards. Please call the Honor Code Office at 422-2847 if you have questions about those standards.

Preventing & Responding to Sexual Misconduct

In accordance with Title IX of the Education Amendments of 1972, Brigham Young University prohibits unlawful sex discrimination against any participant in its education programs or activities. The university also prohibits sexual

harassment—including sexual violence—committed by or against students, university employees, and visitors to campus. As outlined in university policy, sexual harassment, dating violence, domestic violence, sexual assault, and stalking are considered forms of "Sexual Misconduct" prohibited by the university. University policy requires all university employees in a teaching, managerial, or supervisory role to report all incidents of Sexual Misconduct that come to their attention in any way, including but not limited to face-to-face conversations, a written class assignment or paper, class discussion, email, text, or social media post. Incidents of Sexual Misconduct should be reported to the Title IX Coordinator at t9coordinator@byu.edu or (801) 422-8692. Reports may also be submitted through EthicsPoint at https://titleix.byu.edu/report or 1-888-238-1062 (24-hours a day). BYU offers confidential resources for those affected by Sexual Misconduct, including the university's Victim Advocate, as well as a number of non-confidential resources and services that may be helpful. Additional information about Title IX, the university's Sexual Misconduct Policy, reporting requirements, and resources can be found at http://titleix.byu.edu or by contacting the university's Title IX Coordinator.

Student Disability

Brigham Young University is committed to providing a working and learning atmosphere that reasonably accommodates qualified persons with disabilities. If you have any disability which may impair your ability to complete this course successfully, please contact the Services for Students with Disabilities Office (422-2767). Reasonable academic accommodations are reviewed for all students who have qualified, documented disabilities. Services are coordinated with the student and instructor by the SSD Office. If you need assistance or if you feel you have been unlawfully discriminated against on the basis of disability, you may seek resolution through established grievance policy and procedures by contacting the Equal Employment Office at 422-5895,D-285 ASB.