IEOR E4525: Machine Learning for OR and FE
Fall 2108 Syllabus and Course Logistics

Instructor: Christian Kroer

September 11, 2020

Location: Online
Time: Fridays, 10:10am to 12:40pm.
Office hours: TBD

1 Introduction

This is a graduate-level course on Machine Learning. We will cover some of the theoretical foundations for machine learning, and then cover a number of the most important ML algorithms that are used in practice.

We will have about 13 lectures total. Each lecture will consist of two blocks of slightly more than one hour each, with a 15m break in the middle.

2 Prerequisites

This is intended to be an advanced MS level course for MS students in Operations Research and Financial Engineering. Students should therefore have a good background in optimization and applied probability. Some familiarity with statistics and in particular, regression and maximum likelihood (ML) techniques, will also be useful. It is also important that students are comfortable with vector and matrix notation and are comfortable with concepts from linear algebra such as the rank of a matrix and the eigen decomposition of a square matrix.

Students should also be familiar with programming in python (it's likely possible to pick this up alongside the course, if you have other programming experience). The course will make extensive use of Jupyter notebooks as well as the packages `sklearn`, `numpy`, `xgboost`, and either `tensor flow` or `pytorch`.

3 Software Requirements

Homework assignments will require you to build ML models for various problems. Please complete these assignments in python 3.8. You will need `sklearn`,

`numpy`, `xgboost`, and a neural network library (we will most likely use Tensor Flow). The best way to get all this set up is to install the anaconda package from `https://www.anaconda.com/products/individual`.

4 Textbooks

There are no required books for this course. You are free to study the material using only the lecture slides. However, I highly recommend supplementing the slide materials with readings from the following books, which are freely available online:

- An Introduction to Statistical Learning with Applications in R [ISLR] by James, Witten, Hastie and Tibshirani. Free here: `http://www-bcf.usc.edu/~gareth/ISL/ISLR%20First%20Printing.pdf`
- Dive into Deep Learning [D2L] by Zhang, Lipton, Li, Smola. Free here: `https://d2l.ai/`

The following books are considered auxiliary reading. I will probably assign a few readings from HFT.

- Pattern Recognition and Machine Learning [PRML] by Christopher M. Bishop. We will use this book for the EM algorithm.
- The Elements of Statistical Learning [HTF] by Trevor Hastie, Robert Tibshirani, Jerome Friedman. Ebook available on CLIO.
- All of Statistics: A Concise Course in Statistical Inference [Wasserman] by Larry Wasserman

5 Tentative Grading Scheme

The following is a rough tentative grading scheme:

1. Assignments 30%
2. Midterm 25%
3. Final 45%

However, *please keep in mind* that this is the first iteration of this course being taught fully online. The grading scheme is more likely than usual to change because of this.

6 Tentative Syllabus

The below should be viewed as a suggestive outline. I am likely to cut down on the number of algorithms covered per topic, and potentially add a few different ones.

1. Intro to Machine Learning, Probability, Stats, and Optimization background
2. Regression: Linear, logistic, ridge, lasso
3. Model Selection: Cross Validation methods for model selection, Regularization.
4. Classification I: k-nearest neighbours, model assessment, Naive Bayes, LDA, QDA, logistic regression, trees, random forests
5. Unsupervised Learning:
   (a) Dimensionality Reduction and Clustering algorithms: PCA, k-means.
   (b) The EM Algorithm: applications include clustering via normal and Bernoulli mixture models
6. Neural Networks and Deep Learning:
   (a) Neural Networks and Back-propagation.
   (b) Convolutional Neural Networks with applications to image processing
   (c) Recurrent Neural Networks with applications to Natural Language Process- ing