Econ 490: Applied Machine Learning in Economics

Abdollah Farhoodi

Spring, 2019

E-mail: farhood2@illinois.edu
Office Hours: F (9:00-10:00am) Th (5:00-6:00pm)
TA Office Hours: Office hours room: DKH room 15
Cran Chen(M 10:00-11:30am) Gargy Deb(T 9:15-10:45am) Isaac Miller(W 3:30-5pm)

Course Description

This introductory course gives an overview of different concepts, techniques, and algorithms in machine learning and their applications in economics. We begin with topics such as classification, linear and non-linear regressions and end with more recent topics such as boosting, support vector machines, and Neural networks as time allows. This course will give students the basic knowledge behind these machine learning methods and the ability to utilize them in an economic setting. Students will be led and mentored to develop and solve an economic problem with machine learning methods introduced during the course.

Required Materials

The exercises in the course will require R programming. Previous work in this language is helpful but not necessary. Students that have no experience with R can use this course as a starting point.

- Text Book* (Required): An Introduction to Statistical Learning, (James, Witten,Hastie, Tibshirani, 2013)
  - Available online at: https://www-bcf.usc.edu/~gareth/ISL/ISLR
- More advanced book (optional): The Elements of Statistical Learning (James, Witten,Hastie, Tibshirani, 2013)
  - Available online at: https://web.stanford.edu/~hastie/Papers/ESLII.pdf
- An Introduction to R, (Venables, Smith and the R Core Team)
  - Available online at: https://cran.r-project.org/doc/manuals/R-intro.pdf

Prerequisites

- Econ 203 (Economic Statistics)
- Recommended: Econ 471 (Intro to Applied Econometrics)
- Econ 302 (Intermediate Microeconomics)

Course Objectives

After this course, students should be able to ...

1. understand and explain different machine learning algorithms and regression methods
2. apply machine learning tools to real data
3. create meaningful results and visual illustrations
4. compare and contrast different algorithms
5. analyze and discuss the meaning of their codes

Email Policy

Before you start writing an e-mail to a member of the course staff

- Please make sure your question is not already answered in the syllabus or course discussion board.
- Try to Google the error that you get (e.g. copy and paste it into Google). Since R is an open source program, a majority of your questions have been already answered on the web.

Course Structure

Class Structure

Lectures

Classes will be regular lectures on theories and methodologies of statistical methods.

Labs

During lab sessions, students apply the algorithms and practice using R. Students will finish a given example by the end of each lab session.

Exams and Class Projects

There will be one midterm and one final exam on theories and definitions discussed during the lectures. Students are also expected to submit biweekly project reports and a final project in which they apply the practiced methods to a pre-defined problem.

The reports should be submitted on compass. Your TA and one of the instructors will review and comment your work. You can work with any TA and attend any TA's office hours but your report will be reviewed by your assigned TA.

Late Submission Policy

Project reports and final projects should be submitted by the assigned deadlines. Any late submissions will not be accepted.

Grading Policy

The grading policy is as follow:

- **25%** Project reports
- **25%** Final Project
- **20%** Midterm
- **30%** Final

The letter grades are the following:

A+	97-100	C+	77-79
A	93-96	C	73-76
A-	90-92	C-	70-72
B+	87-89	D+	67-69
B	83-86	D	63-66
B-	80-82	D-	60-62

Schedule and weekly learning goals

The schedule is tentative and subject to change. This schedule should be viewed as a road map to the key concepts that students should learn and study before each exam.

Week 01, 01/14 - 01/18: Introduction

- **Lab 0:** Introduction to R, Jupyter notebook, and assignment of individual projects
- Required readings: chapter 1 and chapter 2 part 2.1

Week 02, 01/21 - 01/25: Introduction to Statistical Learning

- Supervised Versus Unsupervised Learning
- Regression Versus Classification Problems
- **Lab 1:** Introduction to R, and assignment of individual projects-continued
- Required readings: chapter 1 and chapter 2 part 2.1, 2.2.1, 2.2.2

Week 03, 01/28 - 02/01: Linear Regression

- Estimating the Coefficients
- Assessing the Accuracy of the estimated Coefficients
- F and t tests
- **Lab 2:** Linear Regression
- Required readings: chapter 3 pp 52-92

Week 04, 02/04 - 02/08: Classification

- Logistic and Probit Regressions
- **Lab 3:** Classification
- Required readings: chapter 4 parts 4.1, 4.2, 4.3, some parts of 4.4.3 (look at the slides), chapter 2

Week 05, 02/11 - 02/15: Resampling Methods

- Cross-Validation
- Bootstrap
- **Lab 4:** Resampling Methods
- Required readings: chapter 5

Week 06, 02/18 - 02/22: Linear Model Selection and Regularization

- LASSO and Ridge
- **Lab 5:** Shrinkage methods
- Required readings: chapter 6 parts 6.1, 6.2

Week 07, 02/25 - 03/01: Tree-based Methods

- Quiz 1
- Regression Trees
- Classification Trees
- **Lab 6:** Decision Tree
- Required readings: chapter 8 part 8.1

Week 08, 03/04 - 03/08: Midterm (Tuesday March 5th)

- **Lab:** Will be announced

Week 9, 03/11 - 03/15: Tree-based Methods

- Bagging, Random Forests
- **Lab 7:** Bagging, Random Forests
- Required readings: chapter 8

Week 10, 03/18 - 03/22: Spring Break

Week 11, 03/25 - 03/29: Tree-Based Methods

- Boosting
- Gradient Boosting
- **Lab 8:** Xgboost

Week 12, 04/01 - 04/05: Applications of ML in Economics

- Paper Discussion
- **Lab 9:** Hyper-parameters Tuning I

Week 13, 04/08 - 04/12: Support Vector Machines (SVM)

- Support Vector Machines (SVM)
- **Lab 9:** Hyper-parameters Tuning II
- Required readings: chapter 9

Week 14, 04/15 - 04/19: Support Vector Machines (SVM)

- **Lab 10:** Support Vector Machines (SVM)
- Required readings: chapter 9

Week 15, 04/22 - 04/26: Neural Networks

- **Lab 11:** Neural Networks

Week 16, 04/29 - 05/03: Limitations of Machine Leaning

- Inference and Causality

Week 17, 05/06 - 05/10: Will be announced