Econ 590: Machine Learning in Economics

Anuar Assamidanov

Spring, 2022

E-mail: aassamidanov@fullerton.edu
Class Room: SGMH 2504
Class Hours: M 7-9:45pm
Office Hours: M/Th 10-11 am

Course Description

This introductory course gives an overview of different concepts, techniques, and algorithms in machine learning (ML) and their applications in an economic setting. We begin with classification, linear and non-linear regressions, bagging, boosting, and end with more recent neural networks and deep learning models. We will also touch on the recent methods at the intersection of ML and econometrics, designed for causal inference, optimal policy estimation, estimation of counterfactual effects. This course will give students the conceptual knowledge behind these ML methods, emphasizing their practical application. Students will learn how to program machine learning algorithms in Python using cutting-edge libraries such as TensorFlow and Scikit-learn.

Required Materials

- Textbook (Required): Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, (Aurélien Géron, 2019)
- Textbook (optional): An Introduction to Statistical Learning, (James, Witten,Hastie, Tibshirani, 2013), https://www-bcf.usc.edu/~gareth/ISL/ISLR
- Textbook (optional): The Elements of Statistical Learning (James, Witten,Hastie, Tibshirani, 2013), https://web.stanford.edu/~hastie/Papers/ESLII.pdf

Prerequisites/Corequisites

Prerequisites: Statistics, Applied Econometrics

Course Objectives

1. Demonstrate the ability to use the existing machine-learning techniques for analyzing the given economic problem.
2. Compare and contrast different algorithms
3. Analyze and discuss the meaning of their codes

Course Structure

Class Structure

Lecture

Classes will be regular lectures on theories and methodologies of ML algorithms.

Labs

During lab sessions, students apply the algorithms and practice using Python. Students will finish a given example by the end of each lab session. To support theses in-class exercises, students should bring laptops to class. Software installation instructions will be provided on the first day of class. Laptops should only be used during class for these exercises and, optionally, for taking notes. If you don't bring a laptop to class, please pair with another student for these in-class exercises.

We will use Python. Most of the modern data science applications are written in Python, supplemented with data-science platforms such as Google TensorFlow and Scikit. An additional advantage of Python is that it is an open-source software. All the problem sets will use examples and data related to economics.

We will be using the Jupyter Notebook application to create and edit Python code. Here is this link to download Jupyter Notebook https://www.anaconda.com/products/individual. This link directs to the Anaconda distribution to install Jupyter and Python. Let the instructor know if you face any issues with installing it.

Exams and Assignments

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

We will have seven assignments during the term, which will consist of programming tasks designed to give you experience working with big and otherwise challenging data in the context of econometric analysis. In addition, you will complete a final project applying the methods you learn in the class to a dataset and question of your choosing. All assignments are due by **Friday, 11:59 pm** on the week listed below unless otherwise indicated.

Assignments and the project will be evaluated based on both functionality and the readability/organization of the code that you write. Part of your grade for the project will also be based

on a writeup of your application of the methods learned in this class.

Grading Policy

TThe grading policy is as follow:

- Assignments: **30%** (best 6 out of 7)).
- Midterm: **30%**
- Final project: **40%** (Paper 20%, Presentation 20%)

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, 01/24 - 01/28: Introduction to Python

- Lab 1: Introduction to Python, Jupyter notebook, Numpy and Pandas libraries
- Required readings: Chapter 1

Week 02, 01/31 - 02/04: Introduction to Machine Learning

- ML models, loss functions, optimization
- Lab 2: Pandas and Data Visualization libraries
- Required readings: Chapter 2
- **Assignment 1: Cleaning and Visualization of the House Prices Data**

Week 03, 02/07 - 02/11: Classification & Training Models

- Performance Measures, linear regression, gradient descent, logistic regression
- Lab 3: Training end-to-end ML model
- Required readings: Chapter 3 & 4

Week 04, 02/14 - 02/18: Classification & Training Models

- Ridge, Lasso, Elastic Net, early stopping
- Lab 4: Bias-variance trade-off, overfitting and underfitting, hyperparameter tuning
- Required readings: Chapter 3 & 4
- **Assignment 2: House Prices - Advanced Regression Techniques**

Week 05, 02/21 - 02/25: Decision Trees

- Classification Tree, Gini Impurity, Entropy, Regression Tree
- Lab 6: Decision Tree
- Required readings: Chapter 6

Week 06, 02/28 - 03/04: Decision Trees

- Voting Classifiers, Random Forests
- Lab 6: Random Forest
- Required readings: Chapter 6

- **Assignment 3: Predict Future Sales. Predict total sales for every product and store in the next month**

Week 07, 03/07 - 03/11: Tree Based Methods

- AdaBoost, Gradient Boosting, XGboost, Feature Importance techniques
- Lab 7: Boosting models, Hyper-parameters Tuning
- Required readings: Chapter 7

Week 08, 03/14 - 03/18: Midterm

- Kaggle Competition

Week 09, 03/21 - 03/25: Dimensionality Reduction

- PCA
- Lab 8: Main Approaches for Dimensionality Reduction
- Required readings: Chapter 8

Week 10, 03/28 - 04/01: No Class

- **Assignment 4: G-Research Crypto Forecasting. Use your ML expertise to predict real crypto market data.**

Week 11, 04/04 - 04/08: Perceptron & Neural Network

- MLP, Stochastic Gradient Descent
- Lab 11: Implementing MLPs with Keras
- Required readings: Chapter 10 & Chapter 11

Week 12, 04/11 - 04/15: Perceptron & Neural Network

- Deep Learning
- Lab 12: Loading and Preprocessing Data with TensorFlow
- Required readings: Chapter 12 & Chapter 13
- **Assignment 5: Time Series Forecasting. Use machine learning to predict grocery sales**

Week 13, 04/18 - 04/22: Perceptron & Neural Network

- RNN
- Lab 13: Time-Series Problem
- Required readings: Chapter 15

Week 14, 04/25 - 04/29: Processing Sequences Using RNNs and CNNs

- LSTM, GRU
- Lab 14: Text as data
- Required readings: Chapter 16
- **Assignment 6: Natural Language Processing with Disaster Tweets. Predict which Tweets are about real disasters and which ones are not**

Week 15, 05/02 - 05/06: Machine Learning and Causal Inference

- Matching/Balance using ML, Shrinkage estimators & variable selection
- Required Reading: Belloni & Chernozhukov (2013); Belloni, Chernozhukov & Hansen (2014); Abadie & Kasy (2017)
- Lab 15: Replication

Week 16, 05/09 - 05/13: Machine Learning and Causal Inference

- Counterfactual prediction, IV using ML
- Required Reading: Carvalho et al (2015); Athey et al. (2018); Belloni, Chernozhukov & Hansen (2012); Hartford et al. (2017)
- Lab 16: Replication
- **Assignment 7: Replication of the Paper on IV**

Week 17, 05/16 - 05/20: Final Presentation