SYLLABUS

Simon Fraser University Dept. of Economics

Machine Learning for Microeconometrics

by

Prof. A. Colin Cameron, Ph.D.

(University of California at Davis)

May 10-13, 2022

The course is taught in 4 sessions of 2 hours: T, W, R, F 10.30 – 12.30.

The slides provided include more than can be covered in eight hours, so some parts will be skipped.

Course outline

Tuesday:	Machine learning: Overview, terminology. Selection of regressors using goodness-of-fit or cross-validation. Regression for prediction: shrinkage methods(ridge, lasso, elastic net),
Wednesday:	Applications for prediction. Causal inference for partial linear regression model using lasso.
Thursday:	Other machine learning learning methods: nonparametric regression, principal components, basis functions (splines), neural networks, neural networks, regression trees, bagging, random forests and boosting.
Friday:	Causal inference for ATE in heterogeneous effects model. Very brief discussion of classification (categorical y) and unsupervised learning (no y).

The material will cover applications using Stata Version 17.
R and Python will not be used. They are the standard software for machine learning for prediction.

Material posted at Course Website http://cameron.econ.ucdavis.edu/sfu2022/

All slides will be posted.

All programs and datasets generating the slides will be posted.

Most papers should be accessible e.g. through JSTOR.

Key references are I strongly suggest getting either a pdf or hardcopy of James et al. *An Introduction to Statistical Learning: with Applications in R* – see below.

The course uses Stata version 16. But most of the basic methods of machine learning are well explained in *An Introduction to Statistical Learning: with Applications in R*, and there is much more machine learning code in R than in Stata.

Slides (posted) Not all slides will be covered.

ML_2022_part0_Overview	Cover all
ML_2022_part1_CrossValidation	Cover all
ML_2022_part2_Shrinkage_Estimators	Cover all
ML_2022_part3_Causal_Lasso	Cover to end slide 36
ML_2022_part4_More_Methods	Focus on regression trees and random forests
ML_2022_part5_More_Causal	Focus on ATE with heterogeneity
ML_2022_part6_Classification_Unsupervised	Brief discussion

Programs and Output Files and Data (posted)

ML_2022_part1.do (uses Stata addon crossfold, loocv, vselect)
ML_2022_part2.do
ML_2022_part3.do
ML_2022_part4.do (uses Stata addon rforest)
ML_2022_part5.do
ML_2022_part6.do

mus203mepsmedexp.dta
mus228ajr.dta

ML_2022_part1.txt
ML_2022_part2.txt
ML_2022_part3.txt
ML_2022_part4.txt
ML_2022_part5.txt
ML_2022_part6.txt

Key readings (only the first posted)

Chapter 28 "Machine Learning for prediction and inference" in A. Colin Cameron and Pravin K. Trrivedi, Microeconometrics using Stata, Second edition, forthcoming.
Posted as Cameron_Trivedi_MUS2_chapter_28.pdf

ISL2: Gareth James, Daniela Witten, Trevor Hastie and Robert Tibsharani (2021), An Introduction to Statistical Learning: with Applications in R, Second Edition, Springer.
A free legal pdf is at https://www.statlearning.com/

Next most important readings (not posted)

Sendhil Mullainathan and J. Spiess: "Machine Learning: An Applied Econometric Approach", Journal of Economic Perspectives, Spring 2017, 87-106.

Alex Belloni, Victor Chernozhukov and Christian Hansen (2014), "High-dimensional methods and inference on structural and treatment effects," Journal of Economic Perspectives, Spring, 29-50.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey and James Robins (2018), "Double/debiased machine learning for treatment and structural parameters," The Econometrics Journal, 21, C1-C68.

Other suggested readings include (not posted)

ESL: Trevor Hastie, Robert Tibsharani and Jerome Friedman (2009), The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer.
A free legal pdf is at http://statweb.stanford.edu/~tibs/ElemStatLearn/index.html

Bradley Efron and Trevor Hastie (2016), Computer Age Statistical Inference: Algorithms, Evidence and Data Science, Cambridge University Press.

Achim Ahrens, Christian Hansen, Mark Schaffer (2019), "lassopack: Model selection and prediction with regularized regression in Stata," arXiv:1901.05397

Susan Athey (2018), "The Impact of Machine Learning on Economics".
http://www.nber.org/chapters/c14009.pdf

Susan Athey and Guido Imbens (2019), "Machine Learning Methods Economists Should Know About."

Alex Belloni, Victor Chernozhukov and Christian Hansen (2011), "Inference Methods for High-Dimensional Sparse Econometric Models," Advances in Economics and Econometrics, ES World Congress 2010, ArXiv 2011.

Alex Belloni, D. Chen, Victor Chernozhukov and Christian Hansen (2012), "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain", Econometrica, Vol. 80, 2369-2429.

Alex Belloni, Victor Chernozhukov, Ivan Fernandez-Val and Christian Hansen (2017), "Program Evaluation and Causal Inference with High-Dimensional Data," Econometrica, 233-299.

Max Farrell (2015), "Robust Estimation of Average Treatment Effect with Possibly more Covariates than Observations", Journal of Econometrics, 189, 1-23.

Max Farrell, Tengyuan Liang and Sanjog Misra (2021), "Deep Neural Networks for Estimation and Inference: Application to Causal Effects and Other Semiparametric Estimands," Econometrica, 89(1), 181-213.

Jon Kleinberg, H. Lakkaraju, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan (2018), "Human decisions and Machine Predictions", Quarterly Journal of Economics, 237-293.

Hal Varian (2014), "Big Data: New Tricks for Econometrics", Journal of Economic Perspectives, Spring, 3-28.

Stefan Wager and Susan Athey (2018), "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," JASA, 1228-1242.