Economics 482-University of Washington

Winter 2018

MW 5:30-7:20 pm SMI 102
Instructor: Gregory M. Duncan
gmduncan@uw.edu
Office: Savery 341
Office Hours: MW 4-5:30PM
TA: Stephanie Liu
Lab W 7:30-8:20 pm SMI 102

1) Week 1
    a) Overview
        i) What is Statistical Learning
            (1) ILSR Ch 2 Statistical Learning
            (2) Shmueli
            (3) Tukey
            (4) Donoho
    b) Elements of R-Programming (in Sections 1 and 2)
2) Weeks 2-8
    a) Review and extension of regression models.
        i) Approximation by piecewise constants
            (1) ILSR Ch 7.2 Step Functions
        ii) Nearest neighbor ideas
        iii) Local Regressions
            (1) CASI Ch 19.8 Kernel Smoothing and Local Regression
            (2) ILSR Ch 7.6 Local Regression
    b) Tuning Parameters
        i) K-fold Cross Validation, Bootstrap
            (1) ILSR Ch 5 Resampling Methods
    c) Variable Selection
            (1) the LASSO and Double LASSO
                (a) CASI Ch 16 Sparse Modeling and the Lasso
                (b) Urminsky et al.
                (c) The Lasso and Applications (Class Notes)
            (2) elastic net
            (3) Shrinkage Methods
                (a) ILSR Ch 6.2 Shrinkage Methods
    d) Nonlinear estimation
        i) Review of central limit theorem and the law of large numbers.
            (1) Slutsky's Calculus

- ii) Maximum likelihood and GMM
  - (1) Maximum Likelihood
    - (a) CASI Ch 4 Fisherian Inference and Maximum Likelihood Estimation
    - (b) Numerical Optimization
    - (c)
  - (2) Classification and logistic regression
    - (a) CASI Ch 8 Generalized Linear Models and Regression Trees
    - (b) Multinomial logistic regression
  - (3) STATA and R
- iii) Non-parametrics
  - (1) Trees
    - (a) ISLR Ch 8.1 The Basics of Decision Trees
  - (2) Splines
    - (a) ILSR Ch 7 Moving Beyond Linearity
  - (3) The curse of dimensionality
  - (4) Large data approach
    - (a) Boosting
      - (i) CASI Ch 17 Random Forests and Boosting
        - 1. Ch 17.2-17.5
      - (ii) Bühlmann and Hothorn https://projecteuclid.org/euclid.ss/1207580163
    - (b) Bagging
      - (i) ISLR Ch 8.2 Bagging
    - (c) Random Forests
      - (i) CASI 17 Random Forests and Boosting
        - 1. Ch 17.1
    - (d) Neural Nets
      - (i) CASI 18 Neural Networks and Deep Learning

3) Weeks 9-10
   - a) The Experimental Model: A/B testing
     - i) Two Sample T-Test https://www.itl.nist.gov/div898/handbook/eda/section3/eda353.htm
     - ii) Permutation Tests http://faculty.washington.edu/kenrice/sisg/SISG-08-06.pdf
     - iii) A/B Tests https://hbr.org/2017/06/a-refresher-on-ab-testing
     - iv) False Discovery Rate Control
       - (1) CASI Ch 15 Large-Scale Hypothesis Testing and FDRs
   - b) Quasi-experiments
   - c) Structural Models
   - d) Threats to validity
     - (1) Endogeneity
     - (2) Selectivity
4) (If we have time)
   - a) The Structural Modeling Approach
     - i) The probability method in Econometrics

        (1) The identification Problem
        (2) Recursive Models
        (3) General Models
    ii) The Structural Causal Modeling Approach of Computer Science
        (1) Pearl's do() Calculus
b) Instrumental Variables Methods
    i) Quasi-experiments revisited
    ii) Selectivity
c) Multi-armed bandits

Texts and Other Material:

Required Texts:

Efron, B., and Trevor Hastie (2016). Computer Age Statistical Inference: Algorithms, Evidence, and Data Science (Institute of Mathematical Statistics Monographs). Cambridge: Cambridge University Press. (https://web.stanford.edu/~hastie/CASI_files/PDF/casi.pdf)

James, Gareth, Daniela Witten , Trevor Hastie, Robert Tibshirani (2013) *An Introduction to Statistical Learning: with Applications in R* (Springer Texts in Statistics), Springer. (http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf)

Harry J. Paarsch and Konstantin Golyaev (2016) A Gentle Introduction to Effective Computing in Quantitative Research: What Every Research Assistant Should Know. MIT Press.

Other Material:

Angrist, Joshua D. and Jörn-Steffen Pischke, (2008). *Mostly Harmless Econometrics: An Empiricist's Companion,* Princeton University Press

*Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection amongst high-dimensional controls. *Review of Economic Studies*, 81(2), 608-650.

Berk, Richard A., (2009) *Statistical Learning from a Regression Perspective,* Springer

Breiman, Leo. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). Statist. Sci. 16, no. 3, 199--231. doi:10.1214/ss/1009213726. http://projecteuclid.org/euclid.ss/1009213726

Bühlmann, Peter and Torsten Hothorn(2007) Boosting Algorithms: Regularization, Prediction and Model Fitting. Statist. Sci. 22, no. 4, 477--505. https://projecteuclid.org/euclid.ss/1207580163

Peter Bickel and Kjell Doksum, (2015) *Mathematical Statistics: Basic Ideas and Selected Topics*, Volume I, Second Edition Chapman and Hall/CRC

Peter Bickel and Kjell Doksum, (2015) *Mathematical Statistics: Basic Ideas and Selected Topics*, Volume II Chapman and Hall/CRC

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, and C. Hansen (2016): “Double Machine Learning for Treatment and Causal Parameters”. Preprint, arXiv:1608.00060. [237,258]

*David Donoho, 2015, *50 years of Data Science*, Tukey Centennial Workshop, http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf

Duncan, G. M. (2017) Notes on Boosting (Class Notes)

Duncan, G. M. (2018) The Lasso and Applications (Class Notes)

Ian Goodfellow and Yoshua Bengio and Aaron Courville, 2016, *Deep Learning*, MIT Press (http://www.deeplearningbook.org)

Haavelmo, T. (1943). "The Statistical Implications of a System of Simultaneous Equations". *Econometrica*, Vol. 11, 1–12. http://www.jstor.org/stable/1905714

Haavelmo, T. (1944). "The Probability Approach in Econometrics" *Econometrica*, Vol. 12, Supplement, iii-115 http://www.jstor.org/stable/1906935

Heckman, James J. and Rodrigo Pinto (2012) *Causal Analysis After Haavelmo: Definitions and a Unifed Analysis of Identification of Recursive Causal Models*, Causal Inference in the Social Sciences, University of Michigan

Heckman, James J. 2010. "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy." *Journal of Economic Literature*, 48w(2): 356-98. DOI: 10.1257/jel.48.2.356

Holland, Paul W. “Statistics and Causal Inference.” Journal of the American Statistical Association, vol. 81, no. 396, 1986, pp. 945–960. www.jstor.org/stable/2289064.

Morgan, Stephen L. and Christopher Winship, (2007) *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, Cambridge University Press

Pearl, Judea (1995) “Causal diagrams for empirical research” *Biometrika*, Volume 82, Issue 4, 1 December Pages 669–688, https://doi.org/10.1093/biomet/82.4.669

Pearl, Judea (2009) "Causal inference in statistics: An overview" *Statistics Surveys* Vol. 3 96–146

Pearl, Judea (2014) "Trygve Haavelmo and the Emergence of Causal Calculus" Econometric Theory, Special Issue on Haavelmo Centennial. http://ftp.cs.ucla.edu/pub/stat_ser/r391.pdf

Peters, Jonas, Dominik Janzing and Bernhard Schölkopf (2017) *Elements of Causal Inference Foundations and Learning Algorithms*, MIT Press (free online http://www.math.ku.dk/~peters/jonas_files/bookDRAFT15-online-2017-10-06.pdf)

*Shmueli, Galit. To Explain or to Predict?. Statist. Sci. 25 (2010), no. 3, 289--310. doi:10.1214/10-STS330. https://projecteuclid.org/euclid.ss/1294167961

*Tukey, John W. The Future of Data Analysis. Ann. Math. Statist. 33 (1962), no. 1, 1--67. doi:10.1214/aoms/1177704711. https://projecteuclid.org/euclid.aoms/1177704711

Urminsky, Oleg and Hansen, Christian and Chernozhukov, Victor, Using Double-Lasso Regression for Principled Variable Selection (2016). Available at SSRN: https://ssrn.com/abstract=2733374 or http://dx.doi.org/10.2139/ssrn.2733374

The Bickel and Doksum books are very hard, but they are at the level of the kind of statistics you would need in grad school. The chapter in Volume II on Machine learning is particularly good though very, very dense. These books are not needed for this course. The Paarsch and Golyaev is required only in the sense that you will need it if you get a job doing this stuff or if you want to be picked up as a research assistant in grad school. (You'll want to be a research assistant, it is the best mentoring you will ever have.)

Recommended Background Text:

Jeffrey M. Wooldridge, (2016) *Introductory Econometrics: A Modern Approach*, Cengage Learning; 6th edition

Florian Heiss, 2016, *Using R for Introductory Econometrics*. (free online) http://www.urfie.net/read/mobile/index.html#p=1

Joseph Adler, (2012) *R in a Nutshell* (In a Nutshell (O'Reilly)) O'Reilly Media

Hadley Wickham, (2014) *Advanced R* Chapman and Hall/CRC

The R books are useful but there are free sites all over the web.

The Wooldridge book has econometrics at the level I expect for people taking this course, I will often refer to it and will assign some readings. The Heiss book is the R version of the course I teach based on Wooldridge.
This course is an advanced continuation of Economics 482 and 483. It assumes a good background in regression at the level of the Wooldridge text above. It will cover topics such as Simultaneous Equations Modeling (Structural Modeling, Instrumental Variables), Non-linear modeling (non-linear regression, logit, probit, maximum likelihood, with a brief, heuristic, introduction to Generalized Method of Moments), Variable Selection using the LASSO, and Modern Non-parametric Modeling from a Machine Learning Perspective (Regression and Classification Trees, Bagging, Boosting, and Random Forests). The course is decidedly hands on emphasizing interpretation, not formal proofs. That said, it uses math and stat skills and concepts without apology or review. The course is ideal for double majors in Economics with Math/Stat/Computer Science or graduate students in Economics, Business, Public Policy or the other social sciences.

Prerequisites: Econ 482, Math 126 and familiarity with matrices and basic matrix operations (structure, transpose, inverse, multiplication); familiarity with vector spaces (bases, orthogonalization, eigenvectors). Basic multivariable calculus and optimization (Lagrangean Multipliers). (So ideally UW Math 124-126, 308, 324). Knowledge of one major statistical program (SAS, STATA, SPSS) and/or familiarity with R. Those who took my Econ 482 should have the background.

Learning Goals: By the end of the course the students will be able to use R to analyze large datasets using a variety of new tools taught in the course. These new tools include instrumental variables, non-linear estimation, the LASSO for variable selection, Neural Nets, various local estimators and Random Forests. Particular emphasis will be put on instrumental variables estimation in the Roy Model (average treatment effects), binary and multinomial logistic regression. They will understand which tools are called for by the different structures of the data and the underlying reason for analysis. So for example, for a label response variable, a logistic type regression model might be best for interpretation, but random forests might be best for prediction.
The overall learning goal include providing sufficient background in machine-learning, as applied to economic problems, so as to make the students able to get jobs as research assistants and analysts at organizations using or interested in using so called "analytics" and "big data" methods. Such places would include major consulting companies (e.g. NERA, Deloitte, Brattle), major technology companies (e.g. Amazon, Google, Tesla), major retailers (e.g. Nike, The Gap, Nordstrom) or government agencies (e.g. FTC, DOJ, IMF). The sufficient background alluded to includes the ability to setup, run and interpret the output of the methods learned in R and interpret results from any standard library of procedures.

Grading: 30% homework, which will be primarily computer oriented. 70% final, which will test interpretation of computer output, set up of analysis and tools and model identification.