University of Wisconsin-Madison

AAE 722 Machine Learning in Applied Economic Analysis

Credits: 4

Meeting Time and Location: TBA

Course Designations and Attributes: General education

Instructional Mode: All fact-to-face

How Credit Hours are met by the Course: This class meets for a total of 4 class period hours (two 75 minute class periods and one 75 minute lab session) each week over the summer/fall semester and carries the expectation that students will work on course learning activities (reading, writing, problem sets, studying, etc) for about 2 hours out of classroom for every class period. The syllabus includes additional information about meeting times and expectations for student work.

Instructor: Associate Professor Sheldon (Xiaodong) Du

Instructor Availability (Office hours): TuTh 1:30-2:30PM, or by appointment; 331 Taylor Hall

Instructor Email: xdu23@wisc.edu

Course Description

The basic methods, implementation and applications of machine learning for understanding contemporary economic issues using large datasets. Building upon understanding of standard econometric models, the topics include data mining techniques; regression model selection and regularization; post selection inference and economic applications; tree-based methods; neural networks; random forests and casual inference; and unsupervised learning.

Requisites

AAE 636 or Econ 704

Learning Outcomes

After completing this course, students will identify and explain, (1) the large-data statistical methods for estimation and causal inference in applied economic and policy analysis; and (2) how these methods complement traditional econometric techniques in applied economic and policy analysis. Students will be able to,

(i) Describe and explain the mechanics of the basic machine learning methods,

(ii) Employ data exploration and visualization tools for analyzing large amounts of data,

(iii) Select model and conduct post-selection inference of high-dimensional data,

(iv) Apply machine learning methods on large data sets for economic and policy analysis, and

(v) Demonstrate the ability to use the R statistical package for the methods covered in the course.

Grading

The final grade for the class will be calculated using the following weights:

Problem sets	20%
Final project	10%
Midterm Exams	40%
Cumulative Final Exam	30%

The final grade will be determined by the following percentages: A: ≥ 90% , AB: 85%~89%, B: 80%-84%, BC: 75%-79%, C: 70%-74%, D: 60-69%, F: <60%.

Laboratory Session

The weekly computer lab session will be led by TA to go over the R coding examples associated with the data exploration, visualization, and machine learning methods covered in the lectures.

Textbooks and Other Course Materials

Readings will be assigned from:

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd Ed. Springer (HTRJ).

The electronic version is available here.

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

The electronic version is available here.

For reference and coding examples, the following book is useful:

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning with Application in R*. 1st Ed. Springer.

The electronic version is available here.

Exams, Quizzes and Final Project

There will be two in-class midterm exams. The final exam will be comprehensive. Final project will allow you to apply what you have learned in class to a selected topic, which needs to be approved by the instructor. You are recommended to work in a group with no more than 3 people. Each group will submit a written report with project description, model construction and result analysis. Each group will present their work to the class in the final week.

Make-up exams will be given only under extenuating circumstances, for which appropriate documentation will be required, and if advance arrangements are made with the instructor.

Homework

About six problem sets will be assigned. Each student must do each assignment independently.

Assignments will be penalized 10 percentage points for each day they are late, unless the student has obtained prior permission, or in the case of an unforeseen emergency. Written notification from your advisor or doctor will be required in both instances.

Academic Integrity

By enrolling in this course, each student assumes the responsibilities of an active participant in UW-Madison's community of scholars in which everyone's academic work and behavior are held to the highest academic integrity standards. Academic misconduct compromises the integrity of the university. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these acts are examples of academic misconduct, which can result in disciplinary action. This includes but is not limited to failure on the assignment/course, disciplinary probation, or suspension. Substantial or repeated cases of misconduct will be forwarded to the Office of Student Conduct & Community Standards for additional review. For more information, refer to studentconduct.wiscweb.wisc.edu/academic-integrity/.

Accommodations For Students with Disabilities

McBurney Disability Resource Center syllabus statement: "The University of Wisconsin-Madison supports the right of all enrolled students to a full and equal educational opportunity. The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy (Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform faculty [me] of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. Faculty [I], will work either directly with the student [you] or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA." http://mcburney.wisc.edu/facstaffother/faculty/syllabus.php

Diversity & Inclusion

Institutional statement on diversity: "Diversity is a source of strength, creativity, and innovation for UW-Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals.

Tentative Schedule

Week 1: Introduction and overview
Readings: HTJR Chs. 1-2; Einav and Levin (2014)
Lab session: Introduction to R

Week 2: Linear regression model and classification
Readings: HTJR Chs. 3-4; Angrist and Pischke Chs. 3-4
Lab session: Linear and logistic regression in R

Week 3: Algorithm and inference
Readings: ET Chs. 1-3
Lab session: Regression inference in R

Week 4: Data visualization and transformation
Readings: Wickham and Grolemund (2017) Chs. 1-4; Pathak (2014) Ch. 4
Lab session: Data visualization and transformation in R

Week 5: Exploratory data analysis
Readings: Wickham and Grolemund (2017) Chs. 5-6; Pathak (2014) Ch. 5
Lab session: Data exploration in R

Week 6: Model assessment and selection
Readings: HTRJ Ch. 7.
Lab session: Linear model selection and regularization in R

Week 7: Model inference
Readings: HTRJ Ch. 8.
Lab: Bootstrap and Bayesian methods in R

Week 8. Post-selection inference
Readings: ET Chs. 15-16, 20; Chernozhukov, Hansen and Spindler (2015)
Lab: Post-selection inference in R

Weeks 9: Economic applications and case studies
Readings: Bajari et al. (2015); Chalfin et al. (2016)

Week 10: Treatment effect and causal inference with big data
Readings: Athey et al. (2018); Belloni et al. (2014); Imbens and Rubin (2015) Chs. 7-8; Shiffrin (2016); Athey (2015, 2017)

Week 11: Tree based methods and boosting
Readings: HTRJ Chs. 9-10;
Lab: Trees and boosting methods in R

Week 12: Random forest and causal inference
Readings: HTRJ Ch. 15; ET Ch. 17; Wager and Athey (2018)
Lab: Random forests in R

Week 13: Neutral networks, deep learning and applications

Readings: HTRJ Ch. 11; ET Ch. 18
Lab: Neutral networks in R

Week 14: Support vector machine
Readings: HTRJ Ch. 12; ET Ch. 19
Lab: Support vector machine in R

Week 15: Unsupervised learning
Readings: HTRJ Ch. 14
Lab: Principle component analysis and clustering methods in R

References

Angrist, J.D. and J. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Athey, S. 2015. Machine Learning and Causal Inference for Policy Evaluation. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Pages 5-6.

Athney, S. 2017. Beyond Prediction: Using Big Data for Policy Problems. *Science* 355(6324): 483-485.

Athey, S., D. Blei, R. Donnelly, F. Ruiz, T. Schmidt. 2018. Estimating Heterogeneous Consumer Preferences for Restaurants and Travel Time Using Mobile Location Data. *American Economic Review: Papers & Proceedings, forthcoming.*

Bajari, P., D. Nekipelov, S. Ryan, and M. Yang. 2015. Machine Learning Methods for Demand Estimation. *American Economic Review: Papers & Proceedings* 105(5): 481-485.

Belloni, A., V. Chernozhukov, and C. Hansen. 2014. High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives* 28(2): 29-50.

Chalfin, A., O. Danieli, A. Hillis, Z. Jelveh, M. Luca, J. Judwig, and S. Mullainathan. 2016. Productivity and Selection of Human Capital with Machine Learning. *American Economic Review* 106(5): 124-127.

Chernozhukov, V., C. Hansen, and M. Spindler. 2015. Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments. *American Economic Review: Papers & Proceedings* 105(5): 485-490.

Einav, Liran and J. Levin. 2014. Economics in the Age of Big Data. *Science* 346(6210): 1243089.

Imbens, G.W and D.B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Pathak, M.A. 2014. *Beginning Data Science with R*. Springer.

Shiffrin, R.M. 2016. Drawing Causal Inference from Big Data. *Proceedings from National Academy of Sciences* 113(27): 7308-7309

Wagner, S. and S. Athey. 2018. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of American Statistical Association*, forthcoming.

Wickham, H. and G. Grolemund. 2017. *R for Data Science: Import, TIDY, Transform, Visualize and Model Data*. O'Reilly.