Economics 690: Machine Learning for Economists
University of Wisconsin - Madison
Fall 2019
Syllabus

Instructors:

Lorenzo Magnolfi
email: magnolfi@wisc.edu
office: 6426 Social Science
office hours: Tuesday, 5:30-6:45pm
(only when teaching)

Christopher Sullivan
email: cjsullivan@wisc.edu
office: 6458 Social Science
office hours: Tuesday, 5:30-6:45pm
(only when teaching)

TA:

Dan Mcleod
email: dpmcleod@wisc.edu

Course Website: https://canvas.wisc.edu/courses/170821

Course Description: This course will introduce students to the use of Machine Learning (ML) in economic analysis. While students will be taught basic techniques of machine learning, much attention will be devoted to evaluating the use of these tools in economics. In particular, students will learn how economists are integrating the tools of machine learning with econometric techniques in current empirical research. Students will also gain hands on experience in using these techniques to answer traditional questions of interest to economists. Topics include (i) an in-depth discussion of the differences and similarities in goals, empirical settings and tools between ML and econometrics, (ii) supervised learning methods for regression and classification, unsupervised learning methods, large data analysis and data mining, (iii) recent methods at the intersection of ML and econometrics, designed for causal inference, optimal policy estimation, estimation of counterfactual effects. We'll present the methods with an emphasis on their practical application.

The course will proceed in two parts. In the first part of the semester, students will attend lectures and work on hands-on assignments to implement the techniques taught in class. In the second part of the semester, students will gain hands-on experience in implementing the techniques and methods presented in lecture through the completion of a supervised group project. Successful projects leverage ML methods to generate insight from economic data

and should include in-depth data analysis, economic intuition, and careful interpretation of the results.

Instructional Mode: all face-to-face

Credits: 3. This 3 CR course has two 75 min lectures per week. Students are expected to work approximately 6 hours per week outside of class to complete assignments and learn the relevant material.

Prerequisites: The course is open to masters students in economics. Familiarity with undergrad-level probability and econometrics is assumed.

Lectures: Tuesday and Thursday, 4:15 pm — 5:30 pm, room 6203 Social Science. Lecture will meet for the first 10 weeks of the semester. We will post the lecture slides after each lecture. However, lectures will not be recorded.

Course Designations and Attributes: Breadth - Social Science; Level - Advanced; L&S Credit - Counts as Liberal Arts and Science credit in L&S

Readings: In addition to economics journal articles, readings will be drawn from the following texts:

> James, Witten, Hastie, and Tibshirani. *An Introduction to Statistical Learning.* Both the book and supplemental materials can be found at the following link: http://www-bcf.usc.edu/ gareth/ISL/ (ISL)
>
> Hastie, Tibshirani, and Friedman. *The Elements of Statistical Learning* which is available at https://web.stanford.edu/ hastie/Papers/ESLII.pdf (ESL)

The following article is a useful overview of many of the methods covered in the course:

> Athey, Susan, and Guido W. Imbens. "Machine learning methods that economists should know about." Annual Review of Economics 11 (2019).

Assignments: There will be problem sets assigned throughout the semester (about one a week). The purpose of the problem sets is to help you monitor your progress and assist you in mastering the methods so that you will be able to complete the group project. The problem sets will be graded on completeness and effort on a scale of ✓-, ✓, ✓+. Students

are encouraged to collaborate on the problem sets, however, each student must write and submit their own answers. In the second half of the course, students will work in groups to complete a final project. At the end of the semester, each group will submit a short paper detailing the empirical model, data, methodology, and results. All code used to complete the final project will be submitted with the paper.

Learning Outcomes: learn the basic concepts of ML from an economist's perspectve and how these tools complement the econometrics toolkit, understand differences and similarities between econometrics and ML, learn new methods at the intersection of ML and econometrics that can be used to answer economic questions, gain hands on experience in applying ML to datasets in economics, further develop and enhance analytic skills, learn to code in R.

Grading: Your final grade will be computed using the following weighting scheme:

25%: Problem Sets

70%: Final Project

5%: Participation

Details on the final project (and what constitutes a good one) will be distributed early in the semester. At the end of the semester, letter grades will be assigned as follows:
90+ A, 80-90 AB, 70-80 B, 60-70 BC, 50-60 C, 40-50 D, <40 F.

Lecture Topics

1. **Data, Regression and Visualization (week 1)**
    - Introduction to US Airline industry dataset used throughout the course.
    - Introduction to basic software tools: RStudio, ggplot2, R Markdown.
    - Application: visualizing firms' pricing patterns in complex markets.
        - Peng, Roger D., "R Programming for Data Science." (2017)
        - Tufte, Edward, and P. Graves-Morris. "The visual display of quantitative information." (2014).
2. **The "Two Cultures" of Empirical Modeling (weeks 2-3)**
    - Data science vs. Economics, Statistical models vs. Algorithmic Modeling

- Breiman, Leo. "Statistical modeling: The two cultures (with comments and a rejoinder by the author)." Statistical science 16.3 (2001): 199-231.
- Mullainathan, Sendhil, and Jann Spiess. "Machine learning: an applied econometric approach." Journal of Economic Perspectives 31.2 (2017): 87-106.
- Athey, Susan. "The impact of machine learning on economics." The economics of artificial intelligence: An agenda. University of Chicago Press, 2018.

3. **Supervised Learning: Linear Models (weeks 3-4)**

- *Methods: Regularized regression, Lasso, Shrinkage methods.*
- Bias/Variance Trade-off, Overfitting and Validation.
- Applications: sparsity in asset pricing and macro models, predicting patterns of entry in markets.
  - Selected chapters from ISL, ESL.
  - Gu, Shihao, Bryan Kelly, and Dacheng Xiu. "Empirical asset pricing via machine learning." No. w25398. National Bureau of Economic Research, 2018.
  - Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. "Shrinking the cross section." No. w24070. National Bureau of Economic Research, 2018.
  - Freyberger, Joachim, Andreas Neuhierl, and Michael Weber. "Dissecting characteristics nonparametrically." No. w23227. National Bureau of Economic Research, 2017.
  - Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri. (2018) "Economic predictions with big data: The illusion of sparsity."

4. **Supervised Learning: Nonlinear Models (weeks 5-6)**

- *Methods: Kernels, Trees, Support Vector Machines, Neural Networks.*
- Bagging, Boosting, and Ensemble methods.
- Application: predicting demand and promotional lift, predicting patients' hospital choice.
  - Selected chapters from ISL, ESL.
  - Bajari, P., D. Nekipelov, S. Ryan, and M. Yang, (2015), "Machine Learning Methods for Demand Estimation," American Economic Review, Papers and Proceedings, Vol. 105(5): 481-85
  - Chernozhukov, V., Goldman, M., Semenova, V., & Taddy, M. (2017). "Orthogonal machine learning for demand estimation: High dimensional causal inference in dynamic panels." arXiv preprint arXiv:1712.09988.

- Raval, D., Rosenbaum, T., & Wilson, N. E. (2019). "How Do Machine Learning Algorithms Perform in Changing Environments? Evidence from Disaster Induced Hospital Closures."

5. **Dimensionality Reduction: Clustering and Principal Component Analysis (weeks 6-7)**

   - *Methods: PCA, K-means and Hierarchical clustering.*
   - Application: categorizing markets/market segmentation.
     - Selected chapters from ISL, ESL.
     - Reguant, Mar. "The Efficiency and Distributional Implications of Large-Scale Renewable Policies." Journal of the Association of Environmental and Resource Economics, forthcoming.

6. **Prediction, Causality and A/B Testing (weeks 8-9)**

   - Predictions vs. Counterfactuals, fundamentals of causality and experiments.
   - Application: DeepIV methods, estimating heterogeneous treatment effects, running and interpreting experiments.
     - Athey, Susan. "Beyond prediction: Using big data for policy problems." Science 355.6324 (2017): 483-485.
     - Davis, J., & Heller, S. B. (2017). "Using causal forests to predict treatment heterogeneity: An application to summer jobs." American Economic Review, 107(5), 546-50.
     - Hartford, J., Lewis, G., Leyton-Brown, K., & Taddy, M. (2016). "Counterfactual prediction with deep instrumental variables networks." arXiv preprint arXiv:1612.09596.
     - Burlig, Fiona, Christopher Knittel, David Rapson, Mar Reguant, and Catherine Wolfram. "Machine Learning from Schools about Energy Efficiency." working paper.

7. **Big Data: Natural Language Processing and "Text as Data" (weeks 9-10)**

   - *Methods: Bag of Words and related techniques.*
   - Application: using product-description data for demand prediction.
     - Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy. Text as data. No. w23276. National Bureau of Economic Research, 2017.

- Einav, L., and Levin, J., (2014), "Economics in the age of big data." Science, 346(6210), 1243089.
- Varian, H., (2014), "Big Data: New Tricks for Econometrics," Journal of Economic Perspectives, 3-27.

8. **Supervised work on group project (weeks 11-15)**

Key Dates

Date	Topic
09/05	Beginning of Classes
09/19	No Lecture
09/30	Provide Final Project group and topic info
10/15	No Lecture
11/12	End of Lectures
12/12	Submit Final Project

Grievance Procedure

The Department of Economics has developed a grievance procedure through which you may register comments or complaints about a course, an instructor, or a teaching assistant. The Department continues to provide a course evaluation each semester in every class. If you wish to make anonymous complaints to an instructor or teaching assistant, the appropriate vehicle is the course evaluation. If you have a disagreement with an instructor or a teaching assistant, we strongly encourage you to try to resolve the dispute with him or her directly. The grievance procedure is designed for situations where neither of these channels is appropriate.

If you wish to file a grievance, you should go to room 7238 Social Science and request a Course Comment Sheet. When completing the comment sheet, you will need to provide a detailed statement that describes what aspects of the course you find unsatisfactory. You will need to sign the sheet and provide your student identification number, your address, and a phone where you can be reached. The Department plans to investigate comments fully and will respond in writing to complaints.

Your name, address, phone number, and student ID number will not be revealed to the instructor or teaching assistant involved and will be treated as confidential. The Department needs this information, because it may become necessary for a commenting student to have a meeting with the department chair or a nominee to gather additional information. A name and address are necessary for providing a written response.

Misconduct Statement

Academic Integrity is critical to maintaining fair and knowledge based learning at UW Madison. Academic dishonesty is a serious violation: it undermines the bonds of trust and honesty between members of our academic community, degrades the value of your degree and defrauds those who may eventually depend upon your knowledge and integrity.

Examples of academic misconduct include, but are not limited to: cheating on an examination (copying from another student's paper, referring to materials on the exam other than those explicitly permitted, continuing to work on an exam after the time has expired, turning in an exam for regrading after making changes to the exam), copying the homework of someone else, submitting for credit work done by someone else, stealing examinations or course materials, tampering with the grade records or with another student's work, or knowingly and intentionally assisting another student in any of the above. Students are reminded that online sources, including anonymous or unattributed ones like Wikipedia, still need to be cited like any other source; and copying from any source without attribution is considered plagiarism.

The Dept. of Economics will deal with these offenses harshly following UWS14 procedures:

1. The penalty for misconduct in most cases will be removal from the course and a failing grade,

2. The department will inform the Dean of Students as required and additional sanctions may be applied.

3. The department will keep an internal record of misconduct incidents. This information will be made available to teaching faculty writing recommendation letters and to admission offices of the School of Business and Engineering.

If you think you see incidents of misconduct, you should tell your instructor about them, in which case they will take appropriate action and protect your identity. You could also choose to contact our administrator Tammy Herbst -Koel therbst@wisc.edu) and your identity will be kept confidential.

For more information, refer to https://www.students.wisc.edu/doso/academic-integrity/