Econ 4330 - Econometric Machine Learning methods and models
Fall, 2022

Course intructor: Karun Adusumilli

Email: akarun@sas.upenn.edu

Office: Perelman building, 631

Office hours: TBA

Teaching Assistant: TBA, Office hours: TBA

Course description: This course will teach you how to apply modern machine/statistical learning methods - potentially including (but not limited to) penalized regression, Random Forests, Neural networks, k-means clustering, and LDA - for analyzing economic data. While emphasis will be placed on employing machine learning methods for estimating causal effects, we will also discuss other economically relevant applications such as forecasting and text analysis. After completing this course, you should be able to carry out machine learning analyses yourself for addressing economically relevant questions.

Prerequisites: The prerequisite for this course is Econ 2310 (previously, Econ 104) or by permission from the instructor. To do well in this course you will need to be comfortable with Econometric methods typically covered in Econ 2310 such as regression, instrumental variables and limited dependent variables (logit regression).

You are also expected to be proficient with the basics of R programming.

Learning: Econ 4330 will be based around active learning. Each lecture is associated with an R Notebook called R Labs covering computing and other details for the material discussed in the lecture. I expect you to read this before the lecture. The workload for Econ 4330 will be fairly high, approximately 6–7 hours of time spent outside of class per week. The lectures themselves will be a mix of traditional ones, where I discuss machine learning methods, and R Lab discussions, where I discuss programming aspects.

The RNotebooks also describe the reading material expected of you before each lecture. So you should go over these carefully. I will be discussing these Notebooks in some but not all the lectures. Many others will be discussed by your TA in coding sessions. For the rest, you should go over them yourself, and ask questions on Piazza or in office hours if you have any questions.

Required Text: There are three required texts for this course:

- "An Introduction to Statistical Learning" (ISL) by James, Witten, Hastie, & Tibshirani: https://www.statlearning.com/

- "R for Data Science" (RDS) by Wickham & Grolemund: https://r4ds.had.co.nz/
- "Pattern Recognition and Machine Learning" (PRML) by Christopher M Bishop: http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%...%20Springer%20%202006.pdf

All books are freely available from urls listed above. Printed versions are available on Amazon. ISL and RDS are the main textbook references. PRML will be used to cover only a subset of the material.

Required Software: We will use the statistical package R via a front-end called RStudio throughout the course. Both R and RStudio are free and open source. Installation instructions are provided on Canvas. You will be taught machine learning methods through a series of R Notebooks, some of which are discussed in lecture as well. Please make sure to read the R instruction file on Canvas. Additional R resources are listed below. It is important that you download and update to the latest version of R.

Optional Texts: For students who want a deeper theoretical grounding in the material covered in Econ 4330, I will assign optional readings from:

- "The Elements of Statistical Learning" (ESL) by Hastie, Tibshirani, and Friedman http://www.web.stanford.edu/~hastie/ElemStatLearn/

Note that this is purely optional and will not appear on problem sets or exams. Like its counterpart ISL, ESL is available as a free download from the authors' website.

Course Policies

Class participation: The lectures are in person, unless the university decides otherwise. I would urge you to try and attend as many lectures as possible. Since Econ 4330 is an active learning course, you should make sure you are prepared before each lecture. To encourage participation, I will be awarding 2% of the final grade in participation credits. Participation can take many forms: Showing up in the lectures or TA sessions, asking and answering questions on Piazza, coming to the office hours. This does not mean that you need to attend lectures: What I want to see is evidence that you are actively participating and working on the course and keeping up with the material. Participation points are discretionary and given in consultation with the RI.

Academic Integrity: All suspected violations of the code of academic integrity as set forth in the Pennbook will be reported to the Office of Student Conduct. Confirmed violations will result in a failing grade for the course.

Department Policies: All Economics Department course policies are in force in Econ 4330 even if they are not explicitly listed on this syllabus. See: https://economics.sas.upenn.edu/ undergraduate/course-information/course-policies for full details.

Piazza: We will be using an online discussion forum called Piazza, accessible via Canvas, for all written communication in this course. We will use Piazza to make course announcements, answer questions about course material and respond to private messages from individual students regarding personal issues. By asking your question and getting an answer on Piazza, you create a positive externality: other students benefit from your questions and you benefit from theirs. You can even post anonymously if you like. The TA and I will actively moderate Piazza both to answer questions and approve (or correct) answers written by your fellow-students. All written communication for Econ 4330 should be directed to Piazza, not to the instructors' personal email accounts.

Grading and Assignments

Grades for this course will be determined based on 5 homeworks, 2 take home mid-terms, participation, and most importantly, a final project. Specifically,

$$\text{Overall score} = (20 \times \text{Homeworks}) + (13 \times \text{midterm1}) + (15 \times \text{midterm2})$$
$$+ (50 \times \text{Final project}) + (2 \times \text{Participation}).$$

Course Curve: There will be no curve in Econ 4330. This course will demand a larger amount of work than other courses, but provided that you put in the time and effort, you will do well.

Homeworks: I will assign 5 problem sets over the course of the semester. The problem sets are assigned on Tuesdays. Each problem set is due at 17:59pm on the Sunday evening the week after it was assigned (so you will have 12 days to do the homework). The exception to this is HW3, for which you will have 2 weeks to do it. Problem set solutions should be submitted electronically to canvas and include both an .pdf file generated from RMarkdown and the .Rmd file used to create it. (In week 1 we will explain RMarkdown and how to use it to generate pdf documents.) Your grade for this component of the course will be calculated by averaging your 4 highest problem set scores. Make sure to use your 'free skip' wisely; additional exemptions will not be granted except for sudden emergencies.

You may discuss problem set questions and how to solve them with your fellow students, but your code and write-up must be your own. Specifically, I expect you to adhere to an

"empty hands policy." If you meet another student to discuss the problem set, you should leave the room with "empty hands," i.e. no written or digital notes of your discussion. In particular, this means that you may not share code files with one another. If you discuss the problem set with your classmates, please indicate which students you discussed with at the top of your write-up. Failing to adhere to this collaboration policy constitutes a violation of academic integrity.

Exams: There is one mid-term on Tuesday, October 11, and a second on Thursday, November 17. The mid-term is a take-home one and you can work on it for up to 24 hours. Unlike the homeworks, the mid-terms focus on the conceptual understanding of the material. You are free to use any resource, including the textbook, course notes, recordings, and even the internet to help you with your exam. However, you are not allowed to discuss the exam with any 'human': this implies talking about the exam with your fellow students is not allowed, nor is posting on online discussion forums. In particular, you are not to email, text, call, chat, or talk to anyone about the exam except with me and your TA. There will be no makeup midterm. Sudden emergencies, of course, will be discussed and determined by the undergraduate chair.

Submitting exams: You should upload you solutions to Canvas as **PDF** files. You should also ensure that the pdfs are of good quality, and all your answers are legible. Failure to follow all these steps could result in your submission being declined. Make sure to doublecheck that your solutions uploaded to Canvas, as submissions after the deadline are not allowed.

Regrade Requests: Exam regrade requests must be made in writing within a week of receiving your graded exam. As we re-grade the entire exam, your score could rise or fall. You may not discuss your answers with an RI or the instructor before submitting a regrade request.

Final Project: This is the most important part of Econ 4330. It is an opportunity for you to show what you have learned in the course by carrying out a substantive research project on a topic of your choice. Final projects will be carried out in groups of 3–4. You are welcome to form your own group; if you do not wish to form your own group, we will be happy to assign one for you. Two class meetings have been set aside for you to work on your group projects and get help from the instructor and TA: see the tentative semester plan on page 4 of this document. You will also be expected to work on the project outside of class. Projects are due at 11:59pm on Sunday, December 5th. In our final two class meetings of the

semester, you and your group will give a short presentation sharing what you learned from your project. Full details and requirements for the final projects will be provided before the Fall Break.

Installing R and RStudio

See the instructions provided in R-instructions.pdf (available on Canvas). Here are links to some additional free resources to help you learn R:

- http://cran.r-project.org/other-docs.html
- http://www.twotorials.com/
- http://www.r-bloggers.com/google-developers-r-programming-video-lectures/
- http://cran.r-project.org/doc/contrib/Farnsworth-EconometricsInR.pdf
- http://www.ats.ucla.edu/stat/R/

Schedule

The schedule is tentative and will updated in the course of the semester. R Labs refer to R Notebooks. The R Notebooks also contain a reading list. You are expected to go over these before the relevant lectures. Note that date for the HWs refers to when it will be released, not when it is due.

TABLE 1. Schedule (Tentative)

Day	Date	Lecture	HW	R Lab	Exam
Tue	Aug 30	1. Introduction to ML			
Thu	Sept 1	2. R Lab 1: Data analysis in R		R Lab 1	
Tue	Sept 6	3. R Lab 2: Data visualization in R	HW1	R Lab 2	
Thu	Sept 8	4. Statistical learning and regression		R Lab 3	
Tue	Sept 13	5. Resampling methods		R Lab 4	
Thu	Sept 15	6. High dimensional methods 1		R Lab 5	
Tue	Sept 20	7. High dimensional methods 2	HW2	R Lab 6	
Thu	Sept 22	8. Tree based methods I		R Lab 7	
Tue	Sept 27	9. Tree based methods II		R Lab 8	
Thu	Sept 29	10. Neural networks		R Lab 9	
Tue	Oct 4	11. Causal inference and ML	HW3	R Lab 10	
Thu	Oct 6	No class: Fall break			
Tue	Oct 11	Mid-term I			Mid-term I
Thu	Oct 13	12. Causal and generalized forests		R Lab 11	
Tue	Oct 18	13. Instrumental variables I		R Lab 12	
Thu	Oct 20	14. Instrumental variables II		R Lab 13	
Tue	Oct 25	15. Classification methods I	HW4	R Lab 14	
Thu	Oct 27	16. Classification methods II		R Lab 15	
Tue	Nov 1	17. Unsupervised learning I: Clustering		R Lab 16	
Thu	Nov 3	18. Bayesian methods I		R Lab 17	
Tue	Nov 8	19. Bayesian methods II	HW5	R Lab 18	
Thu	Nov 10	20. Text analysis: Regular expressions		R Lab 19	
Tue	Nov 15	22. Latent Dirichlet Allocation		R Lab 20	
Thu	Nov 17	Mid term II			Mid-term II
Tue	Nov 22	23. Miscellaneous topics			
Thu	Nov 24	Thanksgiving break			
Tue	Nov 29	Final Projects			
Thu	Dec 1	Final Projects			
Tue	Dec 6	Final Projects			
Thu	Dec 8	Final Projects			