ECON 573
Machine Learning and Econometrics

Instructor: Andrii Babii

Time and Location: T and Th, 8 a.m. - 9:15 p.m., Gardner - Rm 001

E-mail: andrii@email.unc.edu

Office Hours: T and Th, 9:15 p.m. - 10:15 p.m., Gardner 208A

Prerequisites

Econ 400, 410, 420, and 470 with a grade of C or higher.

Course description

Students will learn how to explore, visualize, and analyze high-dimensional datasets, build predictive models, and estimate causal effects. The course introduces key concepts and tools demanded in the business environment.

Examples of techniques include an advanced overview of linear and logistic regression, model selection and regularization, LASSO, cross-validation, experiments, and causal inference, estimation of treatment effects with high-dimensional controls, networks, classification and clustering, latent variable models, bagging and the bootstrap, decision trees and random forests, textual analysis.

Students will learn basic underlying concepts and will build practical programming skills in R. Heavy emphasis is placed on the analysis of actual datasets, and on applications of specific methodologies. Examples may include consumer choice data, housing prices, asset pricing, network data, internet and social media data, sports analytics.

This course meets the **"Research and Discovery"** objective of the IDEAs in action curriculum. Students immerse themselves in a research project and experience the reflection and revision involved in producing and disseminating original scholarship or creative works.

Questions for Students:

1. How do I establish my point of view, take intellectual risks, and begin producing original scholarship or creative works?
2. How do I narrow my topic, critique current scholarship, and gather evidence in systematic and responsible ways?
3. How do I evaluate my findings and communicate my conclusions?

Learning Outcomes:

1. Frame a topic, develop an original research question or creative goal, and establish a point of view, creative approach, or hypothesis.

2. Obtain a procedural understanding of how conclusions can be reached in a field and gather appropriate evidence.
3. Evaluate the quality of the arguments and/or evidence in support of the emerging product.
4. Communicate findings in a clear and compelling ways.
5. Critique and identify the limits of the conclusions of the project and generate ideas for future work.

Textbooks

1. (BDS) Matt Taddy, *Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions*
2. (ISL) Gareth, Witten, Hastie, and Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Second edition. The book can be downloaded for free from https://www.statlearning.com/.

I strongly encourage you to read the relevant material **before** the class.

Problem sets

There will be approximately 6 problem sets over the course of the semester. Only 5 best will count towards the final grade. Problem sets are independent work and not a group project. You should hand in your assignments at the beginning of class the day they are due. Late problem sets (but before graded problem sets are given back) will be marked down by 50% with no exceptions.

Programming

Problem sets will involve data analysis using R. R is a very flexible, powerful, and popular language and environment for statistical computing and graphics. You can download and install it from https://www.r-project.org/. You may also want to check the R Studio GUI from https://rstudio.com/products/rstudio/.

I do not assume that you have used R in a previous class. I will provide in-class demonstrations, some limited statistical instructions, and code to accompany lectures and assignments. However, this is not a class on R. Like any language, R is only learned by doing. You should install it as soon as possible and familiarize yourself with basic operations.

Some useful R resources:

1. R Tutorial: http://r-tutorial.nl/
2. R in Action: https://livebook.manning.com/book/r-in-action-third-edition/welcome/v-2/
3. Tutorials: https://data.princeton.edu/R

4. YouTube tutorials, e.g., from Google Developers: https://www.youtube.com/playlist?list=PLOU2XLYxmsIK9qQfztXeybpHvru-TrqAP
5. Me and your classmates.
6. Web search: "do X in R". Try variations of X until you find an answer. You will find many answers on https://stackoverflow.com/.

Exam dates

1. Midterm: 10/7, Th (in class)
2. Research project hard copy: 11/16, Th
3. Research project presentations: 11/18, Th and 11/23, T
4. Final: 12/4, Sat (8 a.m., 3 hours)

Research project

For the research project, you will analyze a prediction or a causal inference question using methods learned in the course. You will write a paper, approximately 15 pages long, where you will explain the research question, data, methodology, and results. One possibility is to focus on a prediction problem trying various techniques learned in this class. Another possibility is to estimate causal effects. For the former numerous datasets can be found at https://www.kaggle.com/datasets, which is an online community of data scientists and machine learners. For the latter, you can find more examples on Sakai. There will be oral presentations at the end of the course. The grade will be based both on the oral presentation and the hard-copy of the paper.

Grading

Your final grade will be based on:

- 15% problem sets (5 best)
- 15% research project
- 30% midterm
- 40% final

There will be no make-up exam for the midterm. If you miss a midterm exam because of a medical or family emergency, the final exam score will be 70% of your final grade.

Classroom Etiquette

To maintain a good learning environment for everyone, you must turn off all cell phones, laptops, and other electronic devices during class.

Tentative Schedule

#	Date	Topics	Readings, Ch.	Problem Sets Posted	Due
1	8/19	Big Data and Statistical Learning I	BDS 0, ISL 1-2	PS1	
2	8/24	Big Data and Statistical Learning II	BDS 0, ISL 1-2		
3	8/26	Regression I	BDS 2, ISL 3		
4	8/31	Regression II	BDS 2, ISL 3	PS2	PS1
5	9/2	Regression III	BDS 2, ISL 3		
6	9/7	Uncertainty	BDS 1, ISL 5		
7	9/9	Resampling methods	BDS 1, ISL 5		
8	9/14	Model selection	BDS 3, ISL 6		
9	9/16	Regularization I	BDS 3, ISL 6	PS3	PS2
10	9/21	Regularization II	BDS 3, ISL 6		
11	9/23	Classification I	ISL 4		
12	9/28	Classification II	ISL 4		
13	9/30	Regression splines	ISL 7		
14	10/5	Local regressions	ISL 7	PS4	PS3
15	10/7	**Midterm exam**			
16	10/12	Regression trees I	BDS 9, ISL 8		
17	10/14	Regression trees II	BDS 9, ISL 8		
18	10/19	Random forests	BDS 9, ISL 8		
19	10/21	Factors	BDS 7	PS5	PS4
20	10/26	Clustering I	BDS 7, ISL 10		
21	10/28	Clustering II	BDS 7, ISL 10		
22	11/2	Treatment effects I	BDS 5-6		
23	11/4	Treatment effects II	BDS 5-6		
24	11/9	Treatment effects III	BDS 5-6		
25	11/11	Network data	Slides	PS6	PS5
26	11/16	Deep Learning	ISLR 10		
27	11/18	Research Project Presentation II			
28	11/23	Research Project Presentation II			
29	11/30	Text as data	BDS 8		PS6

Honor Code Statement

The University of North Carolina at Chapel Hill has had a student-administered honor system and judicial system for over 100 years. The system is the responsibility of students and is regulated and governed by them, but faculty share the responsibility. If you have questions about your responsibility under the honor code, please bring them to your instructor or consult with the office of the Dean of Students or the Instrument of Student Judicial Governance. This document, adopted by the Chancellor, the Faculty Council, and the Student Congress, contains all policies and procedures pertaining to the student honor system. Your full participation and observance of the honor code is expected (honor.unc.edu).

Attendance Policy

No right or privilege exists that permits a student to be absent from any class meetings, except for these University Approved Absences:

1. Authorized University activities
2. Disability/religious observance/pregnancy, as required by law and approved by Accessibility Resources and Service and/or the Equal Opportunity and Compliance Office (EOC)
3. Significant health condition and/or personal/family emergency as approved by the Office of the Dean of Students, Gender Violence Service Coordinators, and/or the Equal Opportunity and Compliance Office (EOC).

Accessibility Resources and Services

The University of North Carolina at Chapel Hill facilitates the implementation of reasonable accommodations, including resources and services, for students with disabilities, chronic medical conditions, a temporary disability or pregnancy complications resulting in barriers to fully accessing University courses, programs and activities.

Accommodations are determined through the Office of Accessibility Resources and Service (ARS) for individuals with documented qualifying disabilities in accordance with applicable state and federal laws. See the ARS Website for contact information: https://ars.unc.edu or email ars@unc.edu.

Counseling and Psychological Services

CAPS is strongly committed to addressing the mental health needs of a diverse student body through timely access to consultation and connection to clinically appropriate services, whether for short or long-term needs. Go to their website: https://caps.unc.edu/ or visit their facilities on the third floor of the Campus Health Services building for a walk-in evaluation to learn more. (source: Student Safety and Wellness Proposal for EPC, Sep 2018)

Title IX Resources

Any student who is impacted by discrimination, harassment, interpersonal (relationship) violence, sexual violence, sexual exploitation, or stalking is encouraged to seek resources on campus or in the community. Please contact the Director of Title IX Compliance (Adrienne Allison – Adrienne.allison@unc.edu), Report and Response Coordinators in the Equal Opportunity and Compliance Office (reportandresponse@unc.edu), Counseling and Psychological Services (confidential), or the Gender Violence Services Coordinators (gvsc@unc.edu; confidential) to discuss your specific needs. Additional resources are available at safe.unc.edu.

References

[1] Susan Athey and Guido Imbens. *Machine learning methods economists should know about*, **Annual Review of Economics**, Vol. 11:685-725, https://arxiv.org/abs/1903.10075.

[2] Leo Breiman. *Statistical modeling: the two cultures (with comments and a rejoinder by the author)*. **Statist. Sci.**, 16(3):199-231, 2001.

[3] Matthew Getzkow, Bryan Kelly, and Matt Taddy. *Text as data*. **Journal of Economic Literature**, 57(3):535-574, 2019.

[4] Sendhil Mullainathan and Jann Spiess. *Machine learning: an applied econometric approach*. **Journal of Economic Perspectives**, 31(2):87-106, 2017.