

APEC 8222: Big Data Methods in Economics

Syllabus

Instructor: Steve Miller
Office: Ruttan Hall 316E
Phone: 612-625-3212
Email: s-miller@umn.edu
Office Hours: Tuesdays 10:30am-11:30am

Course Description

Challenges, techniques, and opportunities presented by data that has one or more of the following characteristics: large, unstructured, high frequency, variable quality. The course will consist of three parts: 1) computational tools for applying standard econometric techniques on large datasets, 2) extracting summary information from unstructured data (e.g. images, text) for use in econometric analysis, 3) application of statistical learning techniques (e.g. classifiers, regression trees, machine learning) and the role of such techniques in causal inference.

Prerequisites

APEC 5031 or equivalent; APEC 8221 (Programming for Econometrics) or equivalent programming experience.

Objectives

You should leave this course with an understanding of and ability to estimate standard econometric models on large datasets, produce datasets for analysis from unstructured data, and apply several statistical learning techniques to answer economic questions. In particular, you will be able to

- Use parallelization and distributed computing tools to manipulate economic data
- Generate cross-sectional, time-series, or panel data from unstructured sources containing images or text
- Estimate standard econometric models on datasets containing millions of records
- Apply statistical learning techniques to generate predictions and estimate relationships
- Assess the strengths and weaknesses of statistical learning techniques and evaluate their applicability to questions of causal inference.

Grading

Homework 70%

Project 30%

There will be no opportunities for extra credit.

Homeworks and the project will be evaluated based on both functionality and the readability/organization of the code that you write. Part of your grade for the project will

also be based on a writeup of your application of the methods learned in this class to your research question of interest.

If you think that any grading was done incorrectly or unfairly, please come to my office hours. I am happy to revisit the grading, but will re-grade the entire assignment. The purpose of that policy is to be as fair as possible: if one problem needs re-grading, they probably all should have a second look.

Please see the UMN policy below on make-up work and policies concerning legitimate absences. In short, if a legitimate reason prohibits you from finishing an assignment on time, please let me know and I will make accommodations for you. If you're simply late turning in an assignment for some other (illegitimate) reason, you'll receive a zero for that assignment.

Textbook and Course Materials

Textbook: *The Elements of Statistical Learning, 2nd Edition*, (Hastie, Tibshirani, and Friedman)

This book is available on Robert Tibshirani's web site:

<http://statweb.stanford.edu/~tibs/ElemStatLearn/>

Additional readings and materials will be on the course web site; a list is provided below.

Software

We will use the programming language R. R is open source and freely available. I highly recommend that you also download and use RStudio, which is an integrated development environment (IDE) for programming in R. If you do not have access to a computer, both R and RStudio are available on the computers in Ruttan 314.

While we will be using R in this course, much of what you learn will be transferable to other programming languages – best practices and paradigms for accomplishing tasks are frequently similar across languages, though syntax, function names, etc. may differ. In particular, MATLAB and python have many similar capabilities to R.

Homeworks

We will have 4 homeworks during the term, which will consist of programming tasks designed to give you experience working with big and otherwise challenging data in the context of econometric analysis. In addition, you will complete a small project applying the methods you learn in the class to a dataset and question of your choosing.

Please read the course policies below on collaboration and working together. All homeworks are due by the start of class on the date listed below unless otherwise indicated.

Core readings (in addition to textbook)

- Abadie, A., & Kasy, M. (2017). Choosing among regularized estimators in empirical economics.
- Athey, S. (2017). The impact of machine learning on economics. In *Economics of Artificial Intelligence*. University of Chicago Press.
- Athey, S., Bayati, M., Doudchenko, N., Imbens, G., & Khosravi, K. (2018). *Matrix completion methods for causal panel data models* (No. w25132). National Bureau of Economic Research.
- Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353-7360.
- Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6), 2369-2429.
- Belloni, A., & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2), 521-547.
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2), 608-650.
- Carlos V. Carvalho, Ricardo Mansini, and Marcelo C. Medeiros. (2015). ArCo: An Artificial Counterfactual Approach for Aggregate Data.
- Einav, L., & Levin, J. (2014). The data revolution and economic analysis. *Innovation Policy and the Economy*, 14(1), 1-24.
- Hartford, J., Lewis, G., Leyton-Brown, K., & Taddy, M. (2017). Deep iv: A flexible approach for counterfactual prediction. In *International Conference on Machine Learning*(pp. 1414-1423).
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790-794.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3-28.
- Wager, S., & Athey, S. (2017). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*. 113(523): 1228-1242.

Course Policies

1. Class attendance is expected. You will not be graded on attendance, but I expect you to come: any material covered in class is fair game on assignments and exams. Equally importantly, the questions you and your peers ask and the comments you make during class will make for a better experience for everyone, so please make every effort to attend. That said, unexpected conflicts come up for all of us, so any lecture notes and slides used during class will be available online.
2. You may discuss and work on assignments with other students, but you must write up and turn in your own assignment. You all have different strengths and can learn from one another by working together, so I encourage you to do so. The point of this policy is so that you can learn more, not less, so please don't abuse the privilege and let someone else do the work and you simply copy their answer. That won't help you learn the material.

3. Do not violate the Student Conduct Code. The assignments you turn in and your answers on exams should reflect your own work. Simply copying someone else's work, or otherwise violating the Student Conduct Code (http://regents.umn.edu/sites/default/files/policies/Student_Conduct_Code.pdf) may result in a failing grade and/or additional University action. That may sound scary, and it's supposed to be, but know that if you turn in work that is your own, you have nothing to worry about. If you have any questions or concerns about this policy, please don't hesitate to ask.
4. Treat your classmates with respect. Everyone in the class is here to learn and has an equal right to be treated with respect. This means many things, but in summary, I will not tolerate any form of discrimination or sexual harassment. The University has official policies that you can read if you aren't sure what constitutes either of those things, but your common sense will take you a long way.

Other notes:

- If you require disability accommodations, please let me know at your earliest convenience and we will work together to arrange accommodations. If you have or think you may have a disability and have not yet contacted Disability Services (DS), please do so at 612-626-1333 to arrange for a confidential consultation.

University of Minnesota Policies

Student Conduct Code:

The University seeks an environment that promotes academic achievement and integrity, that is protective of free inquiry, and that serves the educational mission of the University. Similarly, the University seeks a community that is free from violence, threats, and intimidation; that is respectful of the rights, opportunities, and welfare of students, faculty, staff, and guests of the University; and that does not threaten the physical or mental health or safety of members of the University community.

As a student at the University you are expected adhere to Board of Regents Policy: *Student Conduct Code*. To review the Student Conduct Code, please see: http://regents.umn.edu/sites/default/files/policies/Student_Conduct_Code.pdf.

Note that the conduct code specifically addresses disruptive classroom conduct, which means "engaging in behavior that substantially or repeatedly interrupts either the instructor's ability to teach or student learning. The classroom extends to any setting where a student is engaged in work toward academic credit or satisfaction of program-based requirements or related activities."

Use of Personal Electronic Devices in the Classroom:

Using personal electronic devices in the classroom setting can hinder instruction and learning, not only for the student using the device but also for other students in the class. To this end, the University establishes the right of each faculty member to determine if and how personal electronic devices are allowed to be used in the classroom. For complete information, please reference: <http://policy.umn.edu/Policies/Education/Education/STUDENTRESP.html>.

Scholastic Dishonesty:

You are expected to do your own academic work and cite sources as necessary. Failing to do so is scholastic dishonesty. Scholastic dishonesty means plagiarizing; cheating on assignments or examinations;

engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; altering, forging, or misusing a University academic record; or fabricating or falsifying data, research procedures, or data analysis. (Student Conduct Code: http://regents.umn.edu/sites/default/files/policies/Student_Conduct_Code.pdf) If it is determined that a student has cheated, he or she may be given an "F" or an "N" for the course, and may face additional sanctions from the University. For additional information, please see: <http://policy.umn.edu/Policies/Education/Education/INSTRUCTORRESP.html>. The Office for Student Conduct and Academic Integrity has compiled a useful list of Frequently Asked Questions pertaining to scholastic dishonesty: <http://www1.umn.edu/oscai/integrity/student/index.html>. If you have additional questions, please clarify with your instructor for the course. Your instructor can respond to your specific questions regarding what would constitute scholastic dishonesty in the context of a particular class-e.g., whether collaboration on assignments is permitted, requirements and methods for citing sources, if electronic aids are permitted or prohibited during an exam.

Makeup Work for Legitimate Absences:

Students will not be penalized for absence during the semester due to unavoidable or legitimate circumstances. Such circumstances include verified illness, participation in intercollegiate athletic events, subpoenas, jury duty, military service, bereavement, and religious observances. Such circumstances do not include voting in local, state, or national elections. For complete information, please see: <http://policy.umn.edu/Policies/Education/Education/MAKEUPWORK.html>.

Appropriate Student Use of Class Notes and Course Materials:

Taking notes is a means of recording information but more importantly of personally absorbing and integrating the educational experience. However, broadly disseminating class notes beyond the classroom community or accepting compensation for taking and distributing classroom notes undermines instructor interests in their intellectual work product while not substantially furthering instructor and student interests in effective learning. Such actions violate shared norms and standards of the academic community. For additional information, please see: <http://policy.umn.edu/Policies/Education/Education/STUDENTRESP.html>.

Grading and Transcripts:

The University utilizes plus and minus grading on a 4.000 cumulative grade point scale. Please see <http://policy.umn.edu/education/gradingtranscripts> for details..

Sexual Harassment

"Sexual harassment" means unwelcome sexual advances, requests for sexual favors, and/or other verbal or physical conduct of a sexual nature. Such conduct has the purpose or effect of unreasonably interfering with an individual's work or academic performance or creating an intimidating, hostile, or offensive working or academic environment in any University activity or program. Such behavior is not acceptable in the University setting. For additional information, please consult Board of Regents Policy: <http://regents.umn.edu/sites/default/files/policies/SexHarassment.pdf>

Equity, Diversity, Equal Opportunity, and Affirmative Action:

The University provides equal access to and opportunity in its programs and facilities, without regard to race, color, creed, religion, national origin, gender, age, marital status, disability, public assistance status, veteran status, sexual orientation, gender identity, or gender expression. For more information, please consult Board of Regents Policy: http://regents.umn.edu/sites/default/files/policies/Equity_Diversity_EO_AA.pdf.

Disability Accommodations:

The University of Minnesota is committed to providing equitable access to learning opportunities for all students. The Disability Resource Center is the campus office that collaborates with students who have disabilities to provide and/or arrange reasonable accommodations.

If you have, or think you may have, a disability (e.g., mental health, attentional, learning, chronic health, sensory, or physical), please contact DS at 612-626-1333 to arrange a confidential discussion regarding equitable access and reasonable accommodations.

If you are registered with DS and have a current letter requesting reasonable accommodations, please contact your instructor as early in the semester as possible to discuss how the accommodations will be applied in the course.

For more information, please see the DS website, <https://diversity.umn.edu/disability/>.

Mental Health and Stress Management:

As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance and may reduce your ability to participate in daily activities. University of Minnesota services are available to assist you. You can learn more about the broad range of confidential mental health services available on campus via the Student Mental Health Website: <http://www.mentalhealth.umn.edu>.

Academic Freedom and Responsibility

Academic freedom is a cornerstone of the University. Within the scope and content of the course as defined by the instructor, it includes the freedom to discuss relevant matters in the classroom. Along with this freedom comes responsibility. Students are encouraged to develop the capacity for critical judgment and to engage in a sustained and independent search for truth. Students are free to take reasoned exception to the views offered in any course of study and to reserve judgment about matters of opinion, but they are responsible for learning the content of any course of study for which they are enrolled.*

Reports of concerns about academic freedom are taken seriously, and there are individuals and offices available for help. Contact the instructor, the Department Chair, your adviser, the associate dean of the college, or the Vice Provost for Faculty and Academic Affairs in the Office of the Provost.

** Language adapted from the American Association of University Professors "Joint Statement on Rights and Freedoms of Students".*