SPIA 586a: Machine Learning for Policy Analysis

Spring 2024

Last Edited: January 31, 2024

Brandon M. Stewart (he/him)
bms4@princeton.edu
brandonstewart.org
OH: Friday 10:00am-10:50am, Wallace 149

This is a class about using the tools of machine learning to study social data with applications to policy. The power of machine learning tools is their applicability for a wide range of tasks. There are huge opportunities for applying these tools to learn and make decisions about real people, but there are also important challenges. This course aims to both familiarize you with the language of machine learning and provide some of the critical social science training that shows how to apply these tools in practice.

1 Overview

Machine learning and data science are rapidly being deployed for application across all spheres of life. While typical machine learning courses cover the math and mechanics of standard tools for predicting an output given a set of inputs, few have the time to cover the nuances that arise when those inputs and outputs are generated by real humans embedded in a social system. This course aims to fill that gap.

1.1 Course Goals

The goal of this course is to explore the implications of machine learning tools when applied to *social* data. In the process, we will cover some of the basics of machine learning and social science research design. In particular students will learn—

- the basic philosophy and culture of ML (e.g. train/test splits, collaborative task framework etc.)
- key representative techniques in ML and how they fit together (e.g. regression, neural networks, tree models)
- the difference between various kinds of prediction **tasks** and the implications for the data scientists (e.g. prediction within population, prediction out of population, counterfactual prediction)
- the different kinds of **data** sources for social data and the implications that they have for inference (e.g. experiments, designed data, found data)

- the importance of core guiding **values** that shape the application of machine learning (e.g. privacy, fairness, interpretability)

Prerequisites Most participants in the course will have some parts that are familiar to them and some parts that are new. Students with different course backgrounds will bring different knowledge to the table and that's okay! What everyone needs to have is:

- a foundation in how to do basic programming in `R`.
  (we are going to start the course with assignments in R and thus won't be teaching basics of the programming language. If you are a strong programmer in a related language such as Python, this should be a reasonably easy transition that you can do on the fly.)
- a familiarity/comfort with linear regression
  (our first methods lecture begins with a brief review of linear regression before moving to more complex topics.

Does This Course Compare to Other Machine Learning and Statistics Courses? Many machine learning courses are *technique-based*—explaining how individual methods such as neural networks or kernel methods function. This course is more *task-based* we will focus on the logic that guides the use of essentially all machine learning techniques. We will explore some techniques in depth, but will focus more on the intuition for the settings where they would work than the derivation, implementation, or theoretical properties of those algorithms.

2 Course Content

The course is organized into three thematic modules: (1) the target task, (2) the data source, and (3) guiding values. Each module will get the focus of three weeks of classes but we will talk about all three components throughout. In addition to developing a theme, each week of class will include: at least one applied example, a set of opportunities and challenges posed by the theme and one machine learning technique.

The **target task** module covers the goal of the analysis: measurement, prediction and causal inference. In this section we start with the sociological impacts of quantification through measurement and reconsider what the numbers in our dataset even mean. We will then consider the distinction between three different types of prediction: prediction where the target looks like the training data (*prediction within population*), prediction where the target looks different from the training data (*prediction in a new population*) and causal inference (*prediction to a counterfactual population*).

The **data source** module will consider how the origins of different kinds of data that one may encounter in academia or industry change the kinds of inference that can be drawn from then. This module picks up from the topic of causal inference and discusses *experimental data* including A/B tests as conducted within modern tech companies. We then discuss *designed data*—such as surveys where analysts get to control the way data is measured. We will conclude with the setting of the majority of machine learning work—*found data*—where the data is originally collected for some other purposes. This includes administrative data, digital trace data, electronic health records and many other types of sources.

The final module will focus on a topic that will be addressed broadly throughout the class—**guiding values**. Picking up from the topic of found data, we start with a focus on the way such datasets can represent a threat to *privacy*. We discuss potential harms as well as constructive solutions. We then discuss *fairness* and racial justice, investigating the way that minority populations are often disproportionately harmed by applications of machine learning and algorithmic decision making. We will consider different definitions of fairness and discuss constructive solutions for a more just application of machine learning. The final week will tackle *interpretability* and discuss some of the reasons that we might want to prioritize interpretability of the estimated model and when that is important for accountability.

3 Course Assignments

The main graded components for the class are:

1. *Participation:* (10%)
   10% of the grade will be based on asking good questions and generally being a good citizen of the classroom

2. *Minor Assignments:* ($10\times, 10\%$)
   There will be weekly minor assignments to help you explore the application of machine learning techniques in `R` due each week (excluding the first) on Monday 11:59pm. These exercises will be relatively short and graded for completion. These assignments will be completed in `R` markdown.

3. *Major Assignments:* ($3\times, 30\%$)
   You will have three major assignments due during the semester, one per module which will synthesize core concepts across the weeks. You will have at least one week to complete each assignment. These will be due on the dates listed below on the schedule.

4. *Proposal:* ($2\times, 10\%$) For two of weeks 3–8, you will submit a short (max 800 words) proposal of how the methodology described that week could be applied to a policy/research topic of interest to you. Be prepared to talk about this in class.

5. *Proposal Reviews:* ($6\times, 10\%$) Each week of weeks 3–8 you will respond/review a research proposal of one of your colleagues which you have selected. We will talk in class about how to write effective reviews.

6. *Refined Final Proposal:* ($1\times, 20\%$) At the end of the semester you will work in teams of three to submit a refined/updated version of one of the research proposals (or something new) for comments and review from your classmates and teaching staff.

7. *Refined Proposal Review:* ($2\times, 10\%$) At the end of the semester you will be assigned two final proposals to review and comment on.

3.1 Proposal

In at most 800 words, your goal is to lay out how you could apply the class of methods or concepts discussed in the current week's readings to a policy problem of interest to you. This is a great opportunity to get feedback on potential projects. During the course you will write two proposals on your own, a group proposal, and receive several reply memos. The higher the quality of your proposal, the more helpful your feedback will be.

Guidance on the Proposal There aren't many hard and fast rules for the proposals other than the sharp 800 word cutoff. However, we expect that strong proposals will include: a clear question or area of interest, a plausible dataset, a proposal for analysis, and a clear statement of how the analysis will help you learn about the question of interest. It is this last step- connecting the method to the substantive topic of interest that can be very challenging. We want you to gain an intuition for what these methods do well and thinking about them in the context of problems you care about. If you have preliminary analyses on a set of data you have already collected you are more than welcome to include them but we do not expect that most people will have had the opportunity to do so. You may also pose questions to your reviewers on areas of the project where you need some help.

On the Length Cutoff and Deadlines We are asking your classmates to provide a thoughtful review of your proposal within a fairly short time-frame. Thus it is extremely important that you complete your proposal by the deadline and within the word limit. We explicitly use a word count rather than a page limit so that you can include figures if this would be helpful. Please do not mistake the 800 word limit as an indication that this is a simple assignment—proposals from previous years have turned into full-blown papers! Writing clearly and concisely is challenging and it may take you several drafts to articulate your idea well. There is less reading than a typical mini so we'd like you to reinvest that energy in these proposals.

Choosing a Proposal Topic You are only required to write two proposals. Try to choose topics which fit well with your interests. Note that choosing a less popular week will allow you to get more feedback as the total number of reviews is constant and the number of proposals is variable. Finally, the proposals are intended to be applications of existing methods to new policy questions. However, you are also welcome to propose the development of new methods within the area discussed that week if you are so inclined.

Posting your Proposals By Sunday at 5PM, you need to submit your proposal for peer review. We will provide a handout detailing the procedure for submitting a proposal.

3.2 Proposal Reviews

Each week you will review a proposal written by one of your classmates. The goal of the review is to provide constructive feedback on the proposed research project. The comments could address core areas such as data selection, the applicability of the methods or the policy relevance. Here again we have very few hard and fast rules because we want you to give the best comments that you can give based on your unique skills and background. These reviews will be assessed on how helpful they are to the author.

We do ask that you start each review with a short summary of what you think the author is proposing to do and why the author believes it is important. This establishes a common baseline so that the author knows how their work was read and what you are actually responding to. You would be amazed how often this part of the review is one of the most useful pieces!

Posting Reviews With $\frac{1}{3}$ of the class on average posting a proposal each week and everyone writing a review, the process of turning around reviews in two days has the potential to get messy quick. Please carefully follow the instructions we provide on how to post reviews and do so on time.

3.3 Refined Final Proposal

At the end of the course you will submit a final proposal in groups of up to three. This is intended to be a refined/updated version of a previously submitted proposals but can also be a new piece of work. If a new piece of work, feel free to weave together methods from across the different weeks if you like.

We have three requirements for the proposal:

1. Take a serious concrete step
   While the refined proposal is still a proposal, we ask that you take at least *one serious, concrete step* towards completing the project. This might be collecting a sample of the data or testing out the method on some related but easier to access data. The more forward progress you can make, the better.

2. Respond to the reviews
   Respond, at least implicitly, to the reviews you received in earlier rounds. You don't have to do what they were asking, but justify why you aren't.

3. Keep it to 1600 words or less
   This is double the original proposal. Use it wisely!

These proposals will be due on Dean's date.

3.4 Refined Proposal Review

On the deadline for the refined proposals you will receive the refined proposal for two other classmates. You will submit written comments on the work as previously done for the regular proposals. You will have one week to respond.

3.5 Late Submissions

Minor assignments received late will only be eligible for half credit. For major assignments, you have a total of four late days to be used across the three major assignments. Late days will be counted in 24 hour increments; for example, 5:01pm on a 5:00pm deadline is considered a full day late (we realize this is frustrating but it is a way to keep the bookkeeping from becoming unreasonable). Once you have used your four late days, assignment grades will be reduced by 20 percentage points per additional day.

3.6 AI-Assistance Policy

The question about what to do about ChatGPT and other AI-assistance tools goes to the very heart of what we are discussing in this class. We will discuss in class what the policy should be and how we should think about it. For now, you are welcome to use ChatGPT or other automated assistance tools in this class if you add a written acknowledgment that you are using them and how. However, keep three things in mind:

1. ChatGPT ranges a lot in its accuracy topic to topic. For example, it often invents sources that don't exist. You are responsible for what you write using it though!
2. plagiarism here is complicated. If you copy directly from ChatGPT without attribution, we consider that plagiarism. ChatGPT can also inadvertently plagiarize from others (and again you are responsible for what it does on your behalf!)
3. while it is quite good at producing code, learning how to do it yourself is sort of the point. Don't rob yourself of learning a valuable skill through over-reliance.

4 Course Schedule

4.1 Books

The course will use the following books (referenced below using the term between square brackets) as well as a series of individual papers.

> Salganik, Matthew J. *Bit by bit: Social research in the digital age.* Princeton University Press, 2019. [Salganik] `https://www.bitbybitbook.com/`
>
> Grimmer, Justin, Margaret Roberts and Brandon M. Stewart. *Text as Data.* Princeton University Press, 2022. [GRS]
>
> Barocas, Hardt and Narayanan *Fairness and machine learning: Limitations and Opportunities.* Book Manuscript. [BHN] `https://fairmlbook.org/`
>
> James, Witten, Hastie and Tibshirani. *An Introduction to Statistical Learning with Applications in R.* Springer, 2021. [JWHT] `https://www.statlearning.com/`

These books are used different amounts throughout the class and we will provide means for you to access necessary material digitally so please don't feel obligated to make purchases unless you want a physical copy (Salganik's *Bit by Bit* is the only one of the four where we will read over half). If you have substantial statistical or mathematical background, I'm more than happy to recommend other material for the machine learning techniques which leverages those prerequisites.

4.2 Course Outline and Readings

> **NB:** The course schedule is subject to change as we receive feedback. Please always check Canvas modules for the most up to date reference on readings.

A general content warning: many of the assigned and discussed articles tackle difficult subjects which may be emotionally distressing to engage with (e.g. police-involved shootings). I'll try to foreshadow possible concerns in the week before.

Introductions

Week 1: Introduction (Jan 30) Week 1 will discuss the syllabus, establish the three core themes of the class and lay out some of the basic philosophy and structure of machine learning. **Technique of the week: review of linear regression**.

- No Reading This Week! Just Come to Class!

Theme 1: Target Tasks

Week 2: Discovery and Measurement (Feb 6) This week will discuss the core ideas of discovery and measurement. **Technique of the week: regularization and basis functions**.

- Concepts:
  - Salganik Chapter 1: Introduction
  - GRS Chapter 2
  - Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting racial bias in an algorithm used to manage the health of populations." Science 366, no. 6464: 447-453.
- Methods
  - JWHT Chapter 3, 4.0-4.3, 6.0-6.2, 6.4, 7.0-7.4, 7.7

Week 3: Prediction Within Population (Feb 13) This week will consider prediction in the most straightforward setting where the training data is a random sample of the target population. We will include a discussion of how to evaluate performance. **Technique of the week: logistic regression, generalized additive models, CART**.

- Concepts:
  - Salganik et. al. 2020. "Measuring the predictability of life outcomes with a scientific mass collaboration" *Proceedings of the National Academy of Sciences.*
  - Lundberg, Brown-Weinstock, Clampet-Lundquist, Gold, Nelson, Yang, Edin and Salganik "The origins of unpredictability in life trajectory prediction tasks"
  - Blumenstock et al. 2015. Predicting Poverty and Wealth from Mobile Phone Metadata. *Science.*
- Methods:
  - JWHT Chapter 8

Week 4: Prediction in a New Population (Feb 20) This week will cover a more difficult problem: prediction when the target population comes from a different distribution than the training data. **Technique of the week: Random Forest**

- Concepts:
  - Lazer et al. 2014. The Parable of Google Flu: Traps in Big Data Analysis. *Science.*
  - Teng Ye, Rebecca Johnson, Samantha Fu, Jerica Copeny, Bridgit Donnelly, Alex Freeman, Mirian Lima, Joe Walsh, and Rayid Ghani. 2019. Using machine learning to help vulnerable tenants in New York city. In Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '19).

- Methods:
  - JWHT Chapter 8.2.3

Week 5: Prediction to a Counterfactual Population (Feb 27) This week will cover the basics of causal inference which involves prediction to a counterfactual population. We will also discuss why causal inference is so important for creating policy. **Technique of the week: deep learning**.

- Concepts:
  - Athey, Susan, 2017. Beyond prediction: Using big data for policy problems. Science, 355(6324), pp.483-485.
  - Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. "What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory." *American Sociological Review* 86, no. 3: 532–65.
- Methods:
  - JWHT Chapter 10

Major Assignment 1 Due Mar 1

Theme 2: Data Source

Week 6: Experimental Data (Mar 5) In this week we will cover data arising from experimental designs and when it can and can't facilitate inferences about counterfactual population. **Technique of the week: multiple testing and pre-analysis plans**.

- Concepts:
  - Salganik Chapter 4: "Running Experiments"
  - Experiment Papers TBD
- Methods:
  - JWHT Chapter 13

Spring Break

Week 7: Designed Data (Mar 19) In this week we will cover methods for designed data including surveys. **Technique of the week: post-stratification**.

- Concepts:
  - Salganik Chapter 3: "Asking Questions"
  - Bradley, V.C., Kuriwaki, S., Isakov, M. et al. 2021. "Unrepresentative big surveys significantly overestimated US vaccine uptake." *Nature* 600, 695–700.

Week 8: Found Data (Mar 26) In this week we discuss the implications of data not designed by a researcher and thus 'found' whether that be digital trace data, administrative data or consumer data. **Technique of the week: record linkage**.

- Concepts:
  - Salganik Chapter 2: "Observing Behavior"
  - Knox, D., Lowe, W., and Mummolo, J. (2020). Administrative records mask racially biased policing. *American Political Science Review*, 114(3), 619-637.
- Methods
  - Enamorado, Ted, Benjamin Fifield, and Kosuke Imai. 2019. "Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records" *American Political Science Review.*

Theme 3: Guiding Values

Week 9: Privacy (April 2) In this week, we discuss the value of privacy and the complex relationship with data science. **Technique of the week: differential privacy**.

- Concepts:
  - Salganik Chapter 6: Ethics
  - BHN Chapter 1: Introduction
  - Lundberg, Narayanan, Levy and Salganik. 2019. "Privacy, Ethics, and Data Access: A Case Study of the Fragile Families Challenge." *Socius*
- Methods:
  - Evans, G., King, G., Schwenzfeier, M., & Thakurta, A. (2023). Statistically valid inferences from privacy-protected data. American Political Science Review, 117(4), 1275-1290.

Week 10: Fairness (Apr 9) In this week we will consider fairness in machine learning and the way these ideas intersect with racial justice and technology. **Technique of the week: TBD**.

- Concepts:
  - BHN Chapter 5 "Testing Discrimination in Practice"
  - Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code.* United Kingdom: Wiley. Introduction, Chapters 2, 5
  - Bender, Emily M. and Gebru, Timnit and McMillan-Major, Angelina and Shmitchell, Shmargaret (2021) "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?"

Major Assignment 2 Due April 12

Week 11: Interpretability (Apr 16) In this week we will consider what makes a machine learning method interpretable and why interpretability might be something that we value in a method. **Technique of the week: rule lists**.

- Concepts:
  - Rudin, C. (2019). "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead." *Nature Machine Intelligence*, 1(5), 206-215.
  - TBD
- Methods:
  - TBD

Week 12: Machine Learning in Practice and Policy (Apr 23) We will discuss some of the difficulties of using machine learning in practice for research and policy.

- Concepts:
  - Sayash Kapoor and Arvind Narayanan. 2021. "(Ir)reproducible machine learning: a case study"
  - Johnson and Zhang. 2022. "What is the Bureaucratic Counterfactual? Categorical versus Algorithmic Prioritization in U.S. Social Policy" *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency.*
  - Wang, Kapoor, Barocas, and Narayanan. 2023. "Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy"

Major Assignment 3 Due Apr 26