SOC 306 / SML 306: Machine Learning with Social Data: Opportunities and Challenges

Spring 2022

Last Edited: March 6, 2022

Brandon M. Stewart
bms4@princeton.edu
brandonstewart.org
OH: Friday 11am-12pm, Wallace 149

Julia Byeon, Preceptor
ybyeon@princeton.edu
OH: Wednesday 4pm-5pm, Wallace 300

Ha-Joon Chung, Preceptor
hajoon.chung@princeton.edu OH: Monday 11am-12pm, Wallace 242

Hunter York, Preceptor
hyork@princeton.edu
OH: Wednesday 12pm-1pm, Wallace 110

This is a class about using the tools of machine learning to study social data. 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 (1) show social scientists and digital humanities scholars the potential of machine learning to help them learn about humans, make policy and help people while also (2) showing computer scientists how a social science research design perspective can improve their work and give them new outlets for their skills.

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)

1.2 Who Should Take This Course?

This course has something for everyone who is interested in the opportunities for machine learning with social data. This course has two archetypal audiences:

- Social Scientists and Digital Humanities Scholars
  Social scientists who have taken a prior statistics course such as ORF 245 (fundamentals of statistics) or POL 345 (Quantitative Social Science) may already be familiar with some topics around social science research design and causal inference. They will learn how machine learning tools can be used to achieve these same goals and some of the particular opportunities and pitfalls that come along with this change in technique.

- Computer Scientists and Data Scientists
  Computer/data scientists who have already taken a data science (SML 201) or machine learning course (COS 324, ELE 364) may already be familiar with some of the core machine learning philosophy and techniques discussed. They will learn elements of social science research design and how the use of these techniques in real world settings with people differs from textbook abstractions.

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. If you have done well in a course such as POL 345, SML 201, ORF 245, COS 324, ELE 364, you likely have the relevant expertise.)

What Requirements Does It Satisfy for the SML Certificate? The course is listed as a machine learning elective. However, during this semester (Spring 2022) you can request it satisfy the machine learning core requirement by emailing Susan Johansen at susan_j@princeton.edu.

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 to four 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 informative questions or making substantive comments about the readings or lectures either during Q&A or precept.

2. *Precept Assignments:* ($10\times, 20\%$)
   There will be ten precept assignments that explore the application of machine learning techniques in `R` due each week on Wednesday 11:59pm. These exercises will be relatively short. These assignments will be completed in `R` markdown based on instructions that will be distributed in precept.

3. *Problem Sets/Major Assignments:* ($3\times, 40\%$)
   You will have three large 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: Thursday February 24, Thursday March 31, and Thursday April 21.

4. *Final Exam* ($1\times, 30\%$)
   You will have a take-home, open-book final exam that synthesizes the core concepts for the class.

3.1 Late Submissions

Precept 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.

4 Course Schedule

4.1 How Does the Week Go?

Each week there will be: two classes, a pre-recorded methods lecture, a short precept assignment, reading, and a precept. There will also be optional office hours. Here is how those components break down along with their timing.

- *Classes* (Tuesday/Thursday 9am)
  Regular attendance is required (with the caveat that if you are sick, you should not physically, but may virtually, attend class). Generally in-person classes will focus on the week's theme and conceptual material. Half of Thursday's class will always be a Q&A session addressing your questions about the material (both this conceptual material material and the machine learning technique material). These lectures will generally assume that you may or may not have read the material on Tuesday, but that you have read the material by Thursday.

- *Pre-recorded Methods Lecture* (available by 9am Monday, watch by Wednesday night)
  1–2 short videos explaining the machine learning technique of the week. Each week supplemental readings will be available if you want to know more. The lectures often cover a little bit of code with the precepts covering more details on implementation.

- *Precept Assignment* (complete by Wednesday at 11:59pm)
  A short assignment that will always ask you to engage with the readings and complete a short coding assignment related to technical material covered up to the current work.

- *Reading* (complete by Wednesday night)
  You are responsible for completing the conceptual reading before Thursday's class (it will always be a part of the precept assignment). We also list methods supplement readings that track the pre-recorded methods lectures as closely as is feasible. These are optional but for many folks may be a valuable addition.

- *Precept* (Thursday or Friday)
  In precept we will discuss the readings in a small group setting and teach some coding that will be useful for your major assignments.

- *Office Hours* (optional and TBD)
  We will announce office hours as soon as we get a handle on the precept schedule.

4.2 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. 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.

4.3 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, AI systems designed to detect sexual orientation, etc.). I'll try to foreshadow possible concerns in the week before.

Introductions

Week 1: Introduction (Jan 25, 27) Week 1 will 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: linear regression with fixed basis functions**.

- Concepts:
  - Salganik Chapter 1: Introduction
  - GRS Chapter 2: Social Science Research and Text Analysis
- Optional Methods Supplement:
  - JWHT Chapter 3, 7.0-7.4

Theme 1: Target Tasks

Week 2: Discovery and Measurement (Feb 1, 3) This week will discuss the core ideas of discovery and measurement. **Technique of the week: logistic regression, regularization and generalized additive models**.

- Concepts:
  - GRS Chapter 10 Principles of Discovery
  - GRS Chapter 15 Principles of Measurement
  - 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.
  - Gelman, Andrew, Greggor Mattson, and Daniel Simpson. 2018. "Gaydar and the Fallacy of Decontextualized Measurement." *Sociological Science* 5: 270-280.
- Optional Methods Supplement:
  - JWHT Chapter 4.0-4.3, 6.0-6.2, 6.4, 7.7

Week 3: Prediction Within Population (Feb 8, 10) 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: CART and random forests**.

- Concepts:

  - Salganik et. al. 2020. "Measuring the predictability of life outcomes with a scientific mass collaboration" *Proceedings of the National Academy of Sciences.*
  - Blumenstock et al. 2015. Predicting Poverty and Wealth from Mobile Phone Metadata. *Science.*
  - BHN Chapter 2 [selections]
- Optional Methods Supplement:
  - JWHT Chapter 8

Week 4: Prediction in a New Population (Feb 15, 17) 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: boosting**.

- Concepts:
  - Lazer et al. 2014. The Parable of Google Flu: Traps in Big Data Analysis. *Science.*
  - Beauchamp, N., 2017. Predicting and interpolating state-level polls using Twitter textual data. *American Journal of Political Science*, 61(2), pp.490-503.
- Optional Methods Supplement:
  - JWHT Chapter 8.2.3

Week 5: Prediction to a Counterfactual Population (Feb 22, 24) 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: directed acyclic graphs and potential outcomes**.

- 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.
- Optional Methods Supplement:
  - BHN Chapter 4

Major Assignment 1 Due Feb 24

Theme 2: Data Source

Week 6: Experimental Data (Mar 1, 3) 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: neural networks**.

- Concepts:
  - Salganik Chapter 4: "Running Experiments"

- Optional Methods Supplement:
  - JWHT Chapter 13

Week 7: Spring Break

Week 8: Designed Data (Mar 15, 17) 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.
- Optional Methods Supplement:
  - Park, Gelman and Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls" *Political Analysis*.

Week 9: Found Data (Mar 22, 24) 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.
- Optional Methods Supplement:
  - 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 10: Privacy (Mar 29, 31) In this week, we discuss the value of privacy and the complex relationship with data science. **Technique of the week: support vector machines**.

- 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*

[noitemsep]

- Optional Methods Supplement:
  - JWHT Chapter 9

Major Assignment 2 Due Mar 31

Week 11: Fairness (Apr 5, 7) 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: data and pretraining**.

- 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
- Optional Methods Supplement:
  - BHN Chapter 7

Week 12: Interpretability (Apr 12, 14) 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:
  - Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
  - 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.
- Optional Methods Supplement:
  - TBD

Week 13: Machine Learning in Practice (Apr 19, 21) We will discuss and review the material from the class and consider some of the difficulties of implementing ML systems in practice **Technique of the week: TBD**.

- Concepts:
  - Grimmer, Roberts and Stewart. 2021. Machine Learning for Social Science: An Agnostic Approach. *Annual Review of Political Science*
  - Sayash Kapoor and Arvind Narayanan. 2021. "(Ir)reproducible machine learning: a case study"
- Optional Methods Supplement:
  - JWHT Chapter 10

Major Assignment 3 Due Apr 21