SOC 306 / SML 306: Machine Learning with Social Data: Opportunities and Challenges (Spring 2021)

Brandon M. Stewart Emily Cantrell Alejandro Schugurensky

Last Edited: February 18, 2021

Brandon M. Stewart
bms4@princeton.edu
brandonstewart.org

Emily Cantrell, Preceptor
emilymc@princeton.edu

Alejandro Schugurensky, Preceptor
as84@princeton.edu

This is a class about using the tools of machine learning to study social data. The power of machine learning tools is their applicability around 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 broad archetypal audiences:

- Social Scientists and Digital Humanities Scholars
  Social scientists who have taken a prior statistics course such as 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
  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 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 bases to the table and that's okay! What everyone needs to have is:

- some foundation in how to do basic programming in `R`.[1]
- basic statistics or machine learning course including a familiarity with linear regression (e.g. POL 345, SML 201, COS 324, ELE 364)

1.3 How 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.

[1] If you are a strong `Python` programmer, this should be a reasonably easy transition that you can do on the fly.

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.

The final week of the course will provide a review and unification of the themes.

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×, 20%)
   There will be ten precept assignments that explore the application of machine learning techniques in `R` due each week at precept. 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: Wednesday March 3 (Week 5), Wednesday March 31 (Week 9), and April 27 (Week 13).

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. Problem Sets/Major Assignments will include a late policy with the assignment.

4 Reading

The course will use the following books and manuscripts (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.* Book manuscript. [GRS]
- Barocas, Hardt and Narayanan *Fairness and machine learning: Limitations and Opportunities.* Book Manuscript. [BHN] `https://fairmlbook.org/`

We will discuss in class helpful resources for learning the individual machine learning techniques.

5 Course Schedule

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

Each week there will be two classes and a precept. The lectures broadly will focus on the core concepts and the intuitions behind the techniques while precepts will be used for discussion and applications of the methods.

Introductions

Week 1: Introduction 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.
Reading:

- Salganik Chapter 1: Introduction
- GRS Chapter 2: Social Science Research and Text Analysis

Theme 1: Target Tasks

Week 2: Discovery and Measurement This week will discuss the core ideas of discovery and measurement. Technique of the week: logistic regression, regularization and generalized additive models.
Reading:

- GRS Chapter 10 Principles of Discovery
- GRS Chapter 15 Principles of Measurement
- Gelman, Andrew, Greggor Mattson, and Daniel Simpson. 2018. "Gaydar and the Fallacy of Decontextualized Measurement." *Sociological Science* 5: 270-280.

Week 3: Prediction Within Population 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.
Reading:

- 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

Week 4: Prediction in a New Population 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.
Reading:

- 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.

Week 5: Prediction to a Counterfactual Population 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.
Reading:

- GRS Chapter 22 “Principles of Inference”
- BHN Chapter 4 “Causality”
- Athey, Susan, 2017. Beyond prediction: Using big data for policy problems. Science, 355(6324), pp.483-485.

Week 6: Review In midterms week we will pause to review the material and catch up on anything we have missed along the way. Technique of the week: neural networks. Reading:

- Grimmer, Roberts and Stewart. 2021. Machine Learning for Social Science: An Agnostic Approach. *Annual Review of Political Science*

Theme 2: Data Source

Week 7: Experimental Data Week 7 is a half week due to Spring Break. In our single course day, we will cover data arising from experimental designs and when it can and can’t facilitate inferences about counterfactual population. Technique of the week: None (due to half week)
Reading:

- Salganik Chapter 4: “Running Experiments”

Week 8: Designed Data In this week we will cover methods for designed data including surveys. Technique of the week: post-stratification.
Reading:

- Salganik Chapter 3: “Asking Questions”

Week 9: Found Data 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.
Reading:

- 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.

Theme 3: Guiding Values

Week 10: Privacy In this week, we discuss the value of privacy and the complex relationship with data science. Technique of the week: support vector machines.
Reading:

- 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*

Week 11: Fairness 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: Bayesian Additive Regression Trees.
Reading:

- 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

Week 12: Interpretability 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.
Reading:

- 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.

Conclusion

Week 13: Wrap-up (Half Week) In this final half-week we will review key themes from the class.