SML 515 (SOC 515): Machine Learning with Social Data: Opportunities and Challenges

Brandon M. Stewart*

Fall 2020

A graduate course on social data science and the particular opportunities and challenges that arise when applying machine learning to study humans. The course will meet once a week and involve annotating student-directed readings, presenting once, providing feedback on others, and producing a final project.

1 Overview

1.1 Course Goals

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. The goal of this reading course to explore the implications of these machine learning tools when applied to *social* data.

This reading course will provide us an opportunity to delve deeper into our collective interests in this area and discuss a series of individual books and articles in the broader emerging context of social data science. This course is also directly connected to a large undergraduate course that I am developing on Machine Learning with Social Data. Participants in this graduate readings course will have the opportunity to help shape the content and priorities of that class.

A secondary "practical skills" aspect to this course will focus on developing effective presentations.

1.2 Who Should Take This Course?

This class is designed primarily for social science graduate students with an interest in machine learning. Those with an interest in pedagogy and communication around these topics will find the course particularly stimulating. Many of the papers and books we read will be selected by participants in the class itself, but broadly speaking you should have an interest in engaging with the kind of readings listed below. You should have a working knowledge of statistics at the level of

*Last Edited: August 30, 2020

Soc500 and at least some familiarity with basic machine learning concepts. If you don't have these prerequisites but are interested, come talk to me and we will see what we can do!

2 Course Structure

2.1 Responsibilities and Logistics

We will meet once a week over the course of the semester (at a date and time best for those who end up enrolling). Each week following the first, we will have a student-selected reading (in consultation with me) that everyone will read in advance. The student will present on the topic and lead discussion.

Thus the main components of the class are:

1. *Reading:* (12×) You need to carefully read the material each week. That's the whole point of a reading class!

2. *Presentation* (1×) Once during the semester you will pick a reading or set of readings that is of interest to you. Everyone will read the work and then you will develop a presentation on the material. We will make these assignments early so that you can effectively prepare in advances. For inspiration, see some of the exemplar presentation from my reading group here: `https://scholar.princeton.edu/bstewart/sociology-statistics-reading-group`.

3. *Presentation Feedback* (2×) Twice during the semester you will sign up to offer up to two pages of written feedback on the presentation. This provides an opportunity to help your fellow classmates improve their presentation skills!

4. *Scribing* (1×) Each week one participant in the discussion will sign up to write and disseminate notes on the ensuing conversation.

5. *Final Project* (1×) Each participant will produce a final project chosen in consultation with me. The project can take on a variety of forms: a piece of research, a piece of pedagogical material related to the topic or anything else related to the topics we discuss.

6. *Peer Comments* (2×) Give comments to two other students on their final projects.

2.2 Conceptual Structure

In a traditional machine learning course, the flow of the course tracks the development of different statistical models (e.g. `https://www.cs.princeton.edu/courses/archive/fall18/cos324/`). By contrast, we are going to organize our readings by three aspects of data analysis: (1) the task, (2) the data source, and (3) values. These aspects of data analysis are often the untaught curriculum of machine learning and data science. Each theme will take approximately a third of the class in order to ensure that we get broad coverage.

Task covers the goal of the analysis: discovery, measurement, and inference. This section of the course will cover some of the subtle differences between these settings. For example, predicting whether or not a conflict will break out in a given country is a substantially different question than predicting whether or not a conflict will break out if we intervene. If we train an algorithm to predict disease diagnosis from a list of symptoms in one country, it may perform poorly when

we shift context to a different country. Readings in this section of the course will roughly track questions of task.

The **Data Source** will expose students to different kinds of data that they may encounter in academia or industry. While historically much data in the social sciences has been 'designed' (coded by a researcher for doing research like e.g. a survey), many common data sources now are 'found' (e.g. administrative data, digital trace data). Readings in this section will discuss the differences between these types of data (as well as other distinctions like streaming vs. static datasets). To give a concrete example, recent work by Knox, Lowe and Mummolo, demonstrate serious concerns with prior work using administrative data to analyze policing. They reexamine an influential finding from Roland Fryer which showed that during police stops black people were no more likely to be shot than white people and concluded that this showed evidence of no police bias in use of lethal force. Unfortunately this finding is premised on a flaw that bedevils many administrative datasets: selection. The dataset includes only people stopped by police. If being black increases the risk of being stopped, then black individuals with a range of behaviors are stopped while only the most dangerous white individuals are stopped. Thus, shooting these individuals at equivalent rates is actually consistent with racial discrimination. This section will focus on readings that push us to think carefully about the source of data and how that affects the inferences we can and cannot draw.

The final section of the class will cover **Values**. This portion of the class will include readings on interpretability, transparency, fairness and broader ethical considerations. For example, we will read Ruha Benjamin's new book *Race After Technology: Abolitionist Tools for the New Jim Code* which provides a sociological lens on the application of machine learning and algorithmic decision making. Building on the previous topics covered in the class we will assess applications of machine learning on different definitions of fairness and come to grips with the way the thoughtless deployment of machine learning tools can undermine core societal values like justice and privacy.

All three themes will hopefully run through all weeks of the class, but by dividing up the course roughly into these pieces we can ensure coverage while providing some thematic consistency.

3 Course Schedule

Each week will have a particular topic and (with some exceptions for book length readings or particularly deep articles) we will try to have two readings: one that captures the opportunities and one that captures the challenges. Thus we can balance optimism and skepticism.

Week 1: Introductions, Goals and Themes
On the first class we will discuss the three themes of the course and do introductions. I want the course to reflect the goals of the participants and so coming into this class I'd like you to think about what you hope to get out of the class. In order to help get the conversation started, I'm assigning a few readings to help orient us.

- Grimmer, Roberts and Stewart *Text as Data* Chapter 2
- Salganik, Matthew. *Bit by Bit: Social Research in the Digital Age* Chapters 1–2
- Barocas, Hardt and Narayanan *Fairness and machine learning: Limitations and Opportunities* Chapter 1
- Stewart 'A Talk on Talks: A Talk'

Roughly speaking these three pieces set the stage for the themes of Task, Data Source, and Values respectively. They should form a useful jumping off point for our first discussion which will help to identify the core ideas we want to get at in the subsequent 12 weeks.

Weeks 2–4: Tasks These weeks will focus on pieces that are related to the question of the analyst's task. A reading might be chosen because it targets a well-defined task that we want to explore, or because it raises crucial ideas around the importance or challenges of particular tasks.

Some possible readings include:

- Bansak et al (2018) "Improving refugee integration through data-driven algorithmic assignment" *Science*.
- Hand (2010) *Measurement Theory and Practice: The World Through Quantification*
- Salganik et al (2020) "Measuring the predictability of life outcomes with a scientific mass collaboration" *PNAS*.
- Watts et al (2018) "Explanation, prediction, and causality: Three sides of the same coin?"
- Hofman et al (2017) "Prediction and explanation in social systems"
- Gelman et al (2018) "Gaydar and the Fallacy of Decontextualized Measurement"
- Yeomans et al (2019) "Making sense of recommendations" *Journal of Behavioral Decision Making*
- Kleinberg et al (2018) "Human decisions and machine predictions" *Quarterly Journal of Economics*
- Athey (2018) "The impact of machine learning on economics" in *The economics of artificial intelligence: An agenda*
- West (2014) "How to improve the use of metrics: learn from game theory." *Nature* 465:871-872
- Bode et al (2020) "Study Designs for Quantitative Social Science Research Using Social Media" *PsyArXiv*
- Mullainathan and Spiess (2017) "Machine learning: an applied econometric approach" *Journal of Economic Perspectives*
- Bueno de Mesquita and Tyson (2020) "The Commensurability Problem: Conceptual Difficulties in Estimating the Effect of Behavior on Behavior" *American Political Science Review*

Weeks 6–8: Data Source These weeks will be themed around questions of the data source. A reading might be included here because it has a particularly novel or interesting data source. We will also look at work which problematizes particular data sources.

Some possible readings include:

- Knox et. al. (2020) "Administrative Records Mask Racially Biased Policing" *American Political Science Review*
- Lakkaraju et al (2017) "The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables" *KDD*
- Choi and Varian (2012) "Predicting the present with Google Trends" *Economic Record*
- Lazer et al (2014) "The parable of Google Flu: traps in big data analysis" *Science*
- Pechenick et al (2015) "Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution" *PloS one*

- Porton et al (2020) "Inaccuracies in Eviction Records: Implications for Renters and Researchers" *Housing Policy Debate*
- Grimmer et al (2018) "Obstacles to estimating voter ID laws' effect on turnout" *Journal of Politics*
- Chetty et al (2014) "Is the United States still a land of opportunity? Recent trends in intergenerational mobility" *American Economic Review*
- King et al (2013) "How censorship in China allows government criticism but silences collective expression" *American Political Science Review*
- Guess et al (2019) "Less than you think: Prevalence and predictors of fake news dissemination on Facebook" *Science Advances*

Weeks 9–11: Values During these weeks we will discuss readings around normative values that we might hold in data science such as privacy, fairness, interpretability and racial justice.

Some possible readings include:

- Abebe et al (2020). "Roles for Computing in Social Change". *In Conference on Fairness, Accountability, and Transparency (FAT* '20)*, January 27–30, 2020, Barcelona, Spain. ACM, New York, NY, USA,
- Obermeyer et al. (2019) "Dissecting racial bias in an algorithm used to manage the health of populations" *Science*
- Kleinberg et al (2018) "Algorithmic fairness" *AEA Papers and Proceedings*
- Benjamin (2019) *Race After Technology: Abolitionist Tools for the New Jim Code*
- Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases." Science 356, no. 6334 (2017): 183–86.
- Buolamwini and Gebru (2018) "Gender shades: Intersectional accuracy disparities in commercial gender classification" *FAT* '18*
- Rudin (2019) "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead" *Nature Machine Intelligence*
- Brayne (2020) "Predict and Surveil: Data, Discretion, and the Future of Policing" *Oxford University Press*
- Dwork (2008) "Differential privacy: A survey of results" *International conference on theory and applications of models of computation*
- Salganik (2019) *Bit by bit: Social research in the digital age* Chapter 6: Ethics
- Narayanan (2008) "Robust de-anonymization of large sparse datasets" *Security and Privacy*

Week 12: Final Thoughts The final week we will meet to share directions for the final projects and discuss broader themes from the course.