POLSCI 189FS: Intro to Machine Learning and Computational Models in the Social Sciences

Scott de Marchi (Professor of Political Science and Director of Decision Science)
demarchi@duke.edu

Time: W / F 1:45 pm
Location: Physics 205 (usually), or online (rarely)

Overview

Our goal as social scientists is to build models of the world and provide advice to policy makers. Given that human actors are often strategic and the games they play are complex, building and testing these models is difficult and distinct from common examples of machine learning. A task that is often used to motivate introductions to machine learning is teaching a model to recognize hand-written characters using MNIST data (https://www.tensorflow.org/datasets/catalog/mnist). Our task is harder: we must build models that involve forecasting human behavior. For example, when members of the US House vote on bills, what sorts of considerations guide these choices? Given the complexity of these sorts of problems, we will cover both human decision-making and machine learning in this course. The hope is that our machine learning models will be better when they are informed by theoretical models of behavior.

Substantively, we will pick a topic to explore / use for data throughout the course (e.g., bargaining models in PR systems, interstate conflict, voting behavior in mass electorates, or votes in the US Congress) based on student interest. Methodologically, we will work in two main areas. For an introduction to (strategic) behavior, we will cover game theory and computational models of decision-making. For an introduction to building statistical models, we will focus on the machine learning tradition with an emphasis on supervised learning.

Grading and Assignments

Projects (30%): Most weeks will have an assignment due and a final project is on the calendar (week 13) – these assignments will involve scraping data, writing code, and building statistical models.

Final exam (50%): the exam will cover material from the course and focus on creative applications of this material. We will have a practicum in class to determine the parameters of the exam.

Participation (20%): Different students have different preferences when it comes to participation, but engagement with the material and with your colleagues in class is essential. Accordingly, both in-class, out-of-class, and email participation will be considered in determining this portion of the grade.

Reading: Some of the articles in this course are challenging and contain mathematical notation you may be unfamiliar with. This is by design! If you are going to pursue further studies in this area, reading papers that contain mathematical models is a crucial skill to learn. Take time with the reading and come to class with questions. Be comfortable with the fact that you may not understand every step in these models and that multiple attempts may be required. This course is designed to get your feet wet without drowning you, so *don't panic*.

Note: given the pandemic and the possibility we will have to move to online instruction, you will be asked to submit questions on the reading and assignments the day before class **by noon** (via email or slack). I'll put together an "asynchronous set of answers" in response to these questions.

First assignment (due Wednesday, 8/30):

1. Install python 3.11 or later and anaconda on your machine (https://www.anaconda.com/) – use the 64-bit version.
2. Launch anaconda / jupyter and write a "hello world" python program. For help, see https://www.amazon.com/Python-Programming-Introduction-Computer-

Science/dp/1590282752/ref=sr_1_3?dchild=1&keywords=python+computer+science&qid=1597710154&sr=8-3 or any of the online tutorials on python 3.

3. Get a dropbox account (https://www.dropbox.com/). The free one has enough storage for our needs.
4. Get a slack and join the class channel at https://join.slack.com/t/slack-sxr6397/shared_invite/zt-22d91nk77-xCwhG2ayF~1fN_0kxL6H5A
5. Read https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1026&context=economicsperg_ppe (by Timur Kuran) by the **Friday class on 9/1**. Please do not skip the equations – we'll talk about how to go about this in detail on Friday though.

Readings / assignments:

Many of the readings are from An Introduction to Statistical Learning (James et al.). There are both R and python flavors of the text at https://www.statlearning.com/ -- you can use either. Which version is best for you depends on your major at Duke, but the default choice should be the python version.

Week 2 (9/6): Read chapter 1 and 2.1 in https://web.stanford.edu/~hastie/ISLRv2_website.pdf.

Week 3 (9/13): Read 2.2, 3.1, 3.2 in https://web.stanford.edu/~hastie/ISLRv2_website.pdf and watch https://www.youtube.com/watch?v=SjQyLhQIXSM. First **assignment** due 9/4.

Week 4 (9/20): Read Chapters 1 and 2 in https://www.economics.utoronto.ca/osborne/igt/. Second **assignment** due 9/11.

Week 5 (9/27): Read 3.3 and do 3.6 in https://web.stanford.edu/~hastie/ISLRv2_website.pdf . **Final project** assigned / parameters agreed on.

Week 6 (10/4): Read 4.1, 4.2, 4.3 in https://web.stanford.edu/~hastie/ISLRv2_website.pdf.

Week 7 (10/11): Read 5.1, 5.2, and do 5.3 in https://web.stanford.edu/~hastie/ISLRv2_website.pdf.
No class Oct 13 (Fall Break).

Week 8 (10/18): Read 6.1, 6.2, and 6.3 in https://web.stanford.edu/~hastie/ISLRv2_website.pdf and https://glmnet.stanford.edu/articles/glmnet.html.

Week 9 (10/25): Read 7.1 in https://web.stanford.edu/~hastie/ISLRv2_website.pdf . Also, this class will feature the very evil **practicum** on social choice theory featuring the final exam parameters; read Munger and Munger "Choosing in Groups" in preparation (will be emailed to class).

Week 10 (11/1): Read 8.1, 8.2, and do 8.3 in https://web.stanford.edu/~hastie/ISLRv2_website.pdf .

Week 11 (11/8): Read 12.1, 12.2, and do lab 12.5 in https://web.stanford.edu/~hastie/ISLRv2_website.pdf and watch https://www.youtube.com/watch?v=P5mlg91as1c&t=92s.

Week 12 (11/15): TBA. This depends on progress in the class – we may still be covering the core material or we may move on to advanced topics. Typically, we will either do a deep dive into neural networks or NLP (especially given the current generation of LLMs).

Week 13 (11/22): Thanksgiving break.

Week 14 (11/29): TBA. **Final projects** due.

Week 15 (12/6): TBA.