Machine Learning for Political Science

PLSC 597
Fall 2023
Time: Monday 9:00am–12:00pm
Location: Sparks 006

Bruce A. Desmarais
bdesmarais@psu.edu

Office Hours: Monday, 1:30–2:30, Wednesday, 9–10& by appointment
Office Location: Sparks B001

Course Overview: Political science research is now regularly conducted using data that is larger and more complex than the data for which conventional statistical tools were designed. Examples of such data include population-scale information on individual-level consumer and political behavior, data streams collected from social media, and archives of electronic government records. There are three fundamental ways in which fine-grained, voluminous, and high-dimensional data require a set of methods that are more flexible than the conventional toolkit of quantitative social science. First, the data is inherently more complex, making it difficult to specify an adequate statistical model from theory alone. Second, the data is high dimensional, meaning there are more variables than one can include in conventional statistical models. Third, the data contains adequate information to make accurate predictions about unseen data (e.g., forecasts). These three features demand a statistical toolkit that is capable of learning model structure, selecting variables, and producing accurate predictions, which are all capabilities of foundational machine learning methods. In this course, we will cover foundational machine learning, with a focus on application to problems in political science. .

Course Objectives: The broad objectives in this course are that students will develop:

1. Fluency in the language of machine learning; an in-depth understanding of the concepts that have proven most useful in the study of politics.
2. Awareness regarding the research objectives that are best-suited to investigation with machine learning.
3. Command of machine learning software.
4. Understanding of how to explore and describe data using machine learning tools.
5. Practical experience in conducting research using machine learning.

Books: The primary book for this course is (Raschka, Liu, Mirjalili, and Dzhulgakov, 2022). Additional helpful references include Hastie, Tibshirani, and Friedman (2009) and Rhys (2020), which are both available through the PSU library website, or on the authors' website in the case of ESL.

Prerequisites: Students in this course should have background in basic descriptive and inferential statistics. This includes an understanding of descriptive statistics, hypothesis testing, regression analysis, and some experience with a scripting-based statistical software.

Computing: All computing will be conducted in either R or Python. Most class sessions will include a detailed programming tutorial covering the methods from that week of the course. Students are encouraged to use Roar colab, and its capabilities to use RStudio and Jupyter in interactive mode, as the primary computing setup in this course.

Problem Sets: There will be at least one problem set covering each of the top-level topics listed in the course schedule. Problem sets are worth 40% of the final grade.

Methods Tutorial: Each student will be responsible for presenting a detailed tutorial of one of the methods covered in the class. Worth 20% of grade.

Application Review: Each student will be responsible for writing a review of, and leading discussion for, one of the application papers. Worth 10% of grade.

Research paper: Students are required to complete an original research paper, and present the paper during the final meeting of the course. The research paper and presentation are worth 30% of the final grade.

Grading Scale.

Grade	Lower	Upper
A	93	101
A-	90	93
B+	88	90
B	82	88
B-	80	82
C+	78	80
C	72	78
C-	70	72
D+	68	70
D	62	68
D-	60	62
F	0	60

Course Schedule: The schedule below gives the required reading. The readings listed for a particular day should be read before class time that day. The full citations for the readings can be found below in the references section.

1. **Section One:** Introduction and basic principles of Machine Learning

   08/21: Introduction to Machine Learning

   - Raschka et al. (2022), Ch. 1
   - Applications
     * Mitts, Phillips, and Walter (2022)

   08/28: Explanation vs. Prediction

   - (Shmueli et al., 2010)
   - Applications
     * Toft and Zhukov (2012)
     * Cranmer and Desmarais (2017)

2. **Section Two:** Classification and Prediction

   09/11: Classifying with logistic regression

   - Raschka et al. (2022), Ch. 3, pp. 53-75
   - Applications

    * Chenoweth and Ulfelder (2017)
    * Kim (2017)

9/18: SVM, Trees/forests, KNN

  - Raschka et al. (2022), Ch. 3, 75–98
  - Applications
    * ARGYLE and Barber (2022)
    * Streeter (2019)

9/25: Regression for machine learning

  - Raschka et al. (2022), Ch. 9
  - Applications
    * Golder, Golder, and Siegel (2012)
    * Beauchamp (2017)

10/02: Data wrangling and splitting for ML

  - Raschka et al. (2022), Ch. 4
  - Applications
    * Ward, Siverson, and Cao (2007)
    * Cordell, Clay, Fariss, Wood, and Wright (2022)

10/09: Model comparison and tuning

  - Raschka et al. (2022), Ch. 6
  - Applications
    * Menninga and Prorok (2021)
    * Rodriguez and Spirling (2022)

3. **Section Three:** Clustering and Dimension Reduction

10/16: Dimensionality Reduction with PCA

  - Raschka et al. (2022), Ch. 5, pp.139-153
  - Applications
    * West (2017)
    * Michaud, Carlisle, and Smith (2009)

10/23: Clustering
Raschka et al. (2022), Ch. 10

- Applications
  - Franko and Witko (2023)
  - Rudra (2007)

- **Section Four:** Deep learning

  10/30: Neural networks
  Raschka et al. (2022), Ch. 11 & 12
  - Applications
    - Beck, King, and Zeng (2000)

  11/06: Convolutional Neural Nets and Image Classification
  - Raschka et al. (2022) Ch. 14
  - Applications
    - Torres and Cantú (2022)
    - Cantú (2019)

  11/13: Recurrent Neural Nets
  - Raschka et al. (2022) Ch. 15
  - Applications
    - Malone (2022)
    - Chang and Masterson (2020)

  11/27 Transformers and Large Language Models
  - Raschka et al. (2022) Ch. 16
  - Applications
    - Bestvater and Monroe (2023)
    - Argyle, Busby, Fulda, Gubler, Rytting, and Wingate (2023)

  12/04: Project presentations

Student Illness Policy Attendance does not factor into the grade for this course. If a student is feeling ill, they are encouraged to stay home and/or seek care. Just get in touch with the instructor as soon as they are able to do so to make a plan for making up missed material/work. Reasonable accommodation will be made to assure the student's success in the course.

Disability Accommodation Statement Penn State welcomes students with disabilities into the University?s educational programs. Every Penn State campus has an office for students with disabilities. Student Disability Resources (SDR) website provides contact information for every Penn State campus (http://equity.psu.edu/sdr/disability-coordinator). For further information, please visit Student Disability Resources website (http://equity.psu.edu/sdr/).

In order to receive consideration for reasonable accommodations, you must contact the appropriate disability services office at the campus where you are officially enrolled, participate in an intake interview, and provide documentation: See documentation guidelines (http://equity.psu.edu/sdr/guidelines). If the documentation supports your request for reasonable accommodations, your campus disability services office will provide you with an accommodation letter. Please share this letter with your instructors and discuss the accommodations with them as early as possible. You must follow this process for every semester that you request accommodations.

Academic Integrity Statement Academic integrity is the pursuit of scholarly activity in an open, honest and responsible manner. Academic integrity is a basic guiding principle for all academic activity at The Pennsylvania State University, and all members of the University community are expected to act in accordance with this principle. Consistent with this expectation, the University's Code of Conduct states that all students should act with personal integrity, respect other students? dignity, rights and property, and help create and maintain an environment in which all can succeed through the fruits of their efforts.

Academic integrity includes a commitment by all members of the University community not to engage in or tolerate acts of falsification, misrepresentation or deception. Such acts of dishonesty violate the fundamental ethical principles of the University community and compromise the worth of work completed by others.

Counseling and Psychological Services Statement Many students at Penn State face personal challenges or have psychological needs that may interfere with their academic progress, social development, or emotional wellbeing. The university offers a variety of confidential services to help you through difficult times, including individual and group counseling, crisis intervention, consultations, online chats, and mental health screenings. These services are provided by staff who welcome all students and embrace a philosophy respectful of clients? cultural and religious backgrounds, and sensitive to differences in race, ability, gender identity and sexual orientation.

Counseling and Psychological Services at University Park (CAPS) (http://studentaffairs.psu.edu/counseling/): 814-863-0395

Counseling and Psychological Services at Commonwealth Campuses (http://senate.psu.edu/faculty/counseling-services-at-commonwealth-campuses/)

Penn State Crisis Line (24 hours/7 days/week): 877-229-6400 Crisis Text Line (24 hours/7 days/week): Text LIONS to 741741

Educational Equity/Report Bias Statements Consistent with University Policy AD29, students who believe they have experienced or observed a hate crime, an act of intolerance, discrimination, or harassment that occurs at Penn State are urged to report these incidents as outlined on the University?s Report Bias webpage (http://equity.psu.edu/reportbias/)

References

ARGYLE, L. P. and M. Barber (2022). Misclassification and bias in predictions of individual ethnicity from administrative records. *American Political Science Review*, 1–9.

Argyle, L. P., E. C. Busby, N. Fulda, J. R. Gubler, C. Rytting, and D. Wingate (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis 31*(3), 337–351.

Beauchamp, N. (2017). Predicting and interpolating state-level polls using twitter textual data. *American Journal of Political Science 61*(2), 490–503.

Beck, N., G. King, and L. Zeng (2000). Improving quantitative studies of international conflict: A conjecture. *American Political science review 94*(1), 21–35.

Bestvater, S. E. and B. L. Monroe (2023). Sentiment is not stance: Target-aware opinion classification for political text analysis. *Political Analysis 31*(2), 235–256.

Cantú, F. (2019). The fingerprints of fraud: Evidence from mexico's 1988 presidential election. *American Political Science Review 113*(3), 710–726.

Chang, C. and M. Masterson (2020). Using word order in political text classification with long short-term memory models. *Political Analysis 28*(3), 395–411.

Chenoweth, E. and J. Ulfelder (2017). Can structural conditions explain the onset of nonviolent uprisings? *Journal of Conflict Resolution 61*(2), 298–324.

Cordell, R., K. C. Clay, C. J. Fariss, R. M. Wood, and T. M. Wright (2022). Disaggregating repression: Identifying physical integrity rights allegations in human rights reports. *International Studies Quarterly 66*(2), sqac016.

Cranmer, S. J. and B. A. Desmarais (2017). What can we learn from predictive modeling? *Political Analysis 25*(2), 145–166.

Franko, W. W. and C. Witko (2023). Unions, class identification, and policy attitudes. *The Journal of Politics 85*(2), 553–567.

Golder, M., S. N. Golder, and D. A. Siegel (2012). Modeling the institutional foundation of parliamentary government formation. *The Journal of Politics 74*(2), 427–445.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Kim, I. S. (2017). Political cleavages within industry: Firm-level lobbying for trade liberalization. *American Political Science Review 111*(1), 1–20.

Malone, I. (2022). Recurrent neural networks for conflict forecasting. *International Interactions 48*(4), 614–632.

Menninga, E. J. and A. K. Prorok (2021). Battles and bargains: escalation, commitment, and negotiations in civil war. *International studies quarterly 65*(2), 406–422.

Michaud, K. E., J. E. Carlisle, and E. R. Smith (2009). The relationship between cultural values and political ideology, and the role of political knowledge. *Political Psychology 30*(1), 27–42.

Mitts, T., G. Phillips, and B. F. Walter (2022). Studying the impact of isis propaganda campaigns. *The Journal of Politics 84*(2), 1220–1225.

Raschka, S., Y. H. Liu, V. Mirjalili, and D. Dzhulgakov (2022). *Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python.* Packt Publishing Ltd.

Rhys, H. I. (2020). *Machine Learning with R, the tidyverse, and mlr*. Manning Publications.

Rodriguez, P. L. and A. Spirling (2022). Word embeddings: What works, what doesn't, and how to tell the difference for applied research. *The Journal of Politics 84*(1), 101–115.

Rudra, N. (2007). Welfare states in developing countries: Unique or universal? *the Journal of Politics 69*(2), 378–396.

Shmueli, G. et al. (2010). To explain or to predict? *Statistical science 25*(3), 289–310.

Streeter, S. (2019). Lethal force in black and white: Assessing racial disparities in the circumstances of police killings. *The Journal of Politics 81*(3), 1124–1132.

Toft, M. D. and Y. M. Zhukov (2012). Denial and punishment in the north caucasus: Evaluating the effectiveness of coercive counter-insurgency. *Journal of Peace Research 49*(6), 785–800.

Torres, M. and F. Cantú (2022). Learning to see: Convolutional neural networks for the analysis of social science data. *Political Analysis 30*(1), 113–131.

Ward, M. D., R. M. Siverson, and X. Cao (2007). Disputes, democracies, and dependencies: A reexamination of the kantian peace. *American journal of political science 51*(3), 583–601.

West, E. A. (2017). Descriptive representation and political efficacy: Evidence from obama and clinton. *The Journal of Politics 79*(1), 351–355.