Machine Learning for Political Science

PLSC 597
Spring 2021
Time: Tuesday 9:00am–12:00pm
Location: Zoom

Bruce A. Desmarais
bdesmarais@psu.edu
Office Hours: Tuesday, 2–3, Monday 1–2& by appointment
Office Location: Zoom

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: Students are not required to purchase any books for this course. The main references used in the course are, 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 the `R` statistical software. We will use add-on packages, most commonly `mlr` (Bischl, Lang, Kotthoff, Schiffner, Richter, Studerus, Casalicchio, and Jones, 2016). The course will include an instruction in the use of `R` for machine learning.

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

   1/19: Introduction to Machine Learning
   - Rhys (2020), Ch. 1
   - Applications
     * Rheault, Rayment, and Musulan (2019)

   1/26: Explanation vs. Prediction
   - (Shmueli et al., 2010)
   - Applications
     * Toft and Zhukov (2012)
     * Cranmer and Desmarais (2017)

2. **Section Two:** Classification

   2/2: Classifying with logistic regression
   - Rhys (2020)Ch. 4
   - Applications

    * Chenoweth and Ulfelder (2017)
    * Rossini, Stromer-Galley, and Zhang (2020)

2/16: Naive Bayes & SVM

- Rhys (2020) Ch. 6
- Applications
    * Pan (2019)

2/23: Decision trees & Random Forests

- Rhys (2020) Ch. 7 & 8
- Applications
    * Streeter (2019)
    * Gohdes (2020)

3/2: Neural Networks

- Günther and Fritsch (2010) & Hastie et al. (2009) Chapter 11.
- Applications
    * Available electronically through PSU library, Lagazio and Russett (2004)

3/9: Model comparison and selection

- Hastie et al. (2009) Chapter 7.
- Applications
    * Bagashka (2008)
    * Harden and Desmarais (2011)

3. **Section Three:** Regression for machine learning

3/16: Linear and nonlinear regression

- Rhys (2020)Ch. 9 & 10
- Applications
    * Golder, Golder, and Siegel (2012)
    * Shorrocks and Grasso (2020)

3/23: Regularization

- Rhys (2020) Ch. 11
- Applications
    * Wilf (2016)

    * Mitts (2019)

3/30: Complex Models for Continuous Outcomes

- Rhys (2020) Ch. 12
- Applications
    * Mohnen, Rotteveel, Doornbos, and Polder (2020)

4. **Section Four:** Clustering and Dimension Reduction

4/6: PCA

- Rhys (2020) Ch. 13
- Applications
    * Peters (2015)
    * Michaud, Carlisle, and Smith (2009)

4/13: Clustering

- Rhys (2020) Ch. 16 & 17
- Applications
    * Harris (2015)
    * Schrodt and Gerner (2000)

4/20: Text embedding

- Spirling and Rodriguez (ming)
- Applications
    * Rodman (2020)

4/27: Project presentations

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

Bagashka, T. (2008). Invisible politics: Institutional incentives and legislative alignments in the russian duma, 1996–99. *Legislative Studies Quarterly 33*(3), 415–444.

Bischl, B., M. Lang, L. Kotthoff, J. Schiffner, J. Richter, E. Studerus, G. Casalicchio, and Z. M. Jones (2016). mlr: Machine learning in r. *The Journal of Machine Learning Research 17*(1), 5938–5942.

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

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

Gohdes, A. R. (2020). Repression technology: Internet accessibility and state violence. *American Journal of Political Science*.

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.

Günther, F. and S. Fritsch (2010). neuralnet: Training of neural networks. *The R journal 2*(1), 30–38.

Harden, J. J. and B. A. Desmarais (2011). Linear models with outliers: Choosing between conditional-mean and conditional-median methods. *State Politics & Policy Quarterly 11*(4), 371–389.

Harris, R. C. (2015). State responses to biotechnology: Legislative action and policymaking in the us, 1990–2010. *Politics and the Life Sciences 34*(1), 1–27.

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

Lagazio, M. and B. Russett (2004). A neural network analysis of militarized disputes, 1885-1992: Temporal stability and causal complexity. *The Scourge of War: New Extensions on an Old Problem. Ann Arbor, MI: University of Michigan Press (28–60)*.

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. (2019). From isolation to radicalization: anti-muslim hostility and support for isis in the west. *American Political Science Review 113*(1), 173–194.

Mohnen, S. M., A. H. Rotteveel, G. Doornbos, and J. J. Polder (2020). Healthcare expenditure prediction with neighbourhood variables–a random forest model. *Statistics, Politics and Policy 11*(2), 111–138.

Pan, J. (2019). How chinese officials use the internet to construct their public image. *Political Science Research and Methods 7*(2), 197–213.

Peters, M. E. (2015). Open trade, closed borders immigration in the era of globalization. *World Politics 67*(1), 114–154.

Rheault, L., E. Rayment, and A. Musulan (2019). Politicians in the line of fire: Incivility and the treatment of women on social media. *Research & Politics 6*(1), 2053168018816228.

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

Rodman, E. (2020). A timely intervention: Tracking the changing meanings of political concepts with word vectors. *Political Analysis 28*(1), 87–111.

Rossini, P., J. Stromer-Galley, and F. Zhang (2020). Exploring the relationship between campaign discourse on facebook and the publics comments: A case study of incivility during the 2016 us presidential election. *Political Studies*, 0032321719890818.

Schrodt, P. A. and D. J. Gerner (2000). Cluster-based early warning indicators for political change in the contemporary levant. *American Political Science Review*, 803–817.

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

Shorrocks, R. and M. T. Grasso (2020). The attitudinal gender gap across generations: support for redistribution and government spending in contexts of high and low welfare provision. *European Political Science Review 12*(3), 289–306.

Spirling, A. and P. Rodriguez (Forthcoming). Word embeddings: What works, what doesnt, and how to tell the difference for applied research. *Journal of Politics*.

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.

Wilf, M. (2016). credibility and distributional effects of international banking regulations: evidence from us bank stock returns. *International Organization*, 763–796. LASSO.