Machine Learning for Policy Analysis

[WWS 586A] Princeton University, Spring 2018

- ljanastas@princeton.edu
- http://scholar.princeton.edu/ljanastas
- https://github.com/ljanastas/WWS586A-Machine-Learning-Policy-Analysis

Prerequisites

A course in probability and a course in statistical inference. Programming experience will be helpful but not necessary. Machine learning algorithms and data pre–processing will be implemented in `R` but you are free to use `Python` or other languages if you choose.

Course Overview and Objectives

This course will provide an introduction to the theory and applications of machine learning algorithms with a focus on policy applications and issues.

The goals of this course include:

- Developing a basic understanding of the statistical theory underlying common supervised and unsupervised machine learning algorithms.
- Developing the programming skills necessary to train and assess the performance of the most popular machine learning algorithms.
- Gaining an understanding of when and how to apply different types of machine learning algorithms to policy issues.

Required Texts

Hastie, Tibshirani and Friedman. 2013. *The Elements of Statistical Learning* (2nd ed), 7th Printing. Springer Series in Statistics. Available for free here:
https://web.stanford.edu/ hastie/Papers/ESLII.pdf.
Referred to in the schedule as **HTF**.

James, Witten, Hastie and Tibshirani. 2015. *An Introduction to Statistical Learning with Applications in R*. Springer Science. Available for free here:
http://www-bcf.usc.edu/ gareth/ISL/.
Referred to in the schedule as **JWHT**.

Monogan III, James E. 2015. *Political Analysis Using R*, Springer.
http://link.springer.com/book/10.1007%2F978-3-319-23446-5.
Referred to in the schedule as **M3**.

Salganik, Matthew. 2017. *Bit by Bit: Social Research in the Digital Age*, Princeton University Press.
Referred to in the schedule as **SM**.

In addition to these books assigned readings will be available on our blackboard site.

Attendance and Participation

The most important content from this class will come from the lectures and group assignments during lecture time. Because of this and the technical nature of this class, attendance and participation in class is important.

Computer, Tablet and Cell Phone Use Policy

Laptop computers and tablets may be used during class sessions for note taking and programming exercises. Cell phones and other electronic devices must remain off and stored out of sight at all times during class.

Academic Honesty and Integrity

As a student at Princeton University , you have agreed to abide by the University's academic honesty policy. Lack of knowledge of the academic honesty policy is not a reasonable explanation for a violation. Questions related to course assignments and the academic honesty policy should be directed to the instructor.

Special Accommodations

Students with disabilities who require reasonable accommodations in order to participate in course activities or meet course requirements should contact the instructor and work with the Office of Disability Services (https://ods.princeton.edu/) to develop an accommodation plan.

Problem Sets

There a total of five problem sets during the semester covering materials discussed in lectures and in the readings. The format of problem set assignments will vary but will invariably involve a combination of math problems and programming. Unless specifically noted on the problem set, these are **individual** assignments so students will need to show independent work. More information about each assignment will be provided in class the week before it is due.

Course Project

Working together in groups or individually, students will propose a course project which either (1) applies one or more of the machine learning algorithms covered to a substantive problem in your relevant discipline or; (2) proposes a method to improve the performance of a machine learning algorithm for a given problem domain. You will be asked to put together a course project proposal halfway through the academic year and the final course project will be due at the end of the semester.

The final course project will contain two components:

(1) A final presentation of the project to the class and others at the end of the semester and;

(2) A paper which will be handed in for a grade. The paper should resemble a policy report.

Grades

Attendance and participation	5%
Problem Sets	50%
Final Project Presentation	10%
Final Project Paper	35%

Overview of Topics

- Introduction to programming, APIs and web scraping with R.
- The role of machine learning and AI in policy analysis and governance.
- Introduction to statistical learning theory and machine learning.
- Supervised learning:
    - nearest neighbors
    - naive Bayes
    - decision trees
    - regression
    - neural networks
    - support vector machines.
- Regularization, model selection and inference.
- Unsupervised learning:
    - K-Means Clustering.
    - Hierarchical Clustering.
    - EM Algorithm.
    - Hidden markov models.
- Machine learning and causal inference.
    - Synthetic controls and Bayesian structural time series.
    - Recurrent neural networks.
- Network analysis

Tentative Schedule

Preliminaries: Introduction to Programming in R

- Using Github.
- Introduction to programming in R.
- Writing in **R** notebooks and **R** markdown.
- APIs and webscraping in **R**.

Readings

- **M3** Chapters 1, 2, 10, 11.1-11,4.

Introduction to machine learning for policy analysis

- What is machine learning?
- Supervised & unsupervised learning.
- Inference versus prediction for policy analysis.
- Big and small data: research design in the digital age.

Readings

- Kleinberg, J., Ludwig, J., Mullainathan, S. and Obermeyer, Z., 2015. Prediction policy problems. *American Economic Review*, 105(5), pp.491-95. PDF.
- Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S., 2017. Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1), pp.237-293. https://academic.oup.com/qje/article/133/1/237/4095198.
- **JWHT** – Introduction, pp 1-15.
- **SM** – Chapter 1.

Optional Readings

- Mullainathan, S. and Obermeyer, Z., 2017. "Does machine learning automate moral hazard and error?". *American Economic Review*, 107(5), pp.476-80. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5540263/.

Introduction to statistical learning theory

- Training, testing and cross–validation.
- Assessing model accuracy.
- Overfitting.
- Regression vs. classification problems.
- The Bias–Variance tradeoff.

Readings

∗ **JWHT** – Statistical Learning, pp 15-37, 176–184.
∗ **SM** – 2.1, 2.2,2.3.

SUPERVISED LEARNING

Nearest Neighbors

- kNN algorithm.
- *Policy application*: H1–B Visa Certification. H1–B Application Data

Readings

∗ **JWHT** – pp 39–42.

Naive Bayes

- Review of probability and Bayes rule.
- Learning with naive Bayes.
- *Policy application*: Terrorist threat assessment I. Global Terrorism Database

Readings

∗ Collins on Naive Bayes
http://www.cs.columbia.edu/ mcollins/em.pdf

Decision Trees

- Decision tree–based methods.
- Model assessment, information gain and "white box" methods.
- *Policy application*: Terrorist threat assessment II. Global Terrorism Database

Readings

- **JWHT** – Chapter 8.
- Athey, S. and Imbens, G., 2016. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences*, 113(27), pp.7353-7360. http://www.pnas.org/content/early/2016/06/29/1510489113.full

Regression I: Prediction

- Inference vs. prediction.
- Linear regression as a machine learning algorithm.
- Logistic regression as a machine learning algorithm.
- Parameter estimation via gradient descent.
- *Policy application*: Preventative policing: pre–crime targeting and detection I. NYC Stop and Frisk Data: 2003–2016.

Readings

- **JWHT** – Chapter 3.
- Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.https://arxiv.org/pdf/1609.04747v1.pdf.

Regression II: Model Selection and Inference

- Feature selection.
- Regularization.
- Shrinkage methods: ridge regression, LASSO, Bayesian LASSO.
- *Policy application*: Preventative policing: pre–crime targeting and detection II. NYC Stop and Frisk Data: 2003–2016.

Readings

- **JWHT** – Chapter 5.1, 203-243.

Neural Networks

- Overview of neural networks.
- Fitting neural networks with backpropagation.
- *Policy application*: Analyzing political images with convolutional neural networks.

Readings

- Daume. Neural Networks.

Support Vector Machines and Classifier Choice

- Comparing classifiers: performance & interpretability.

Readings

- **JWHT**. 337-366.

UNSUPERVISED LEARNING

Unsupervised Learning I: Clustering methods

- Learning labels from data.
- K-Means Clustering
- Hierarchical Clustering.
- Distance metrics.

Unsupervised Learning II: Latent Variables and Time–Series Analysis.

- EM Algorithm.
- Markov models for time series analysis.
- *Policy application:* Identifying policy regimes from time–series data with hidden Markov models.

Unsupervised Learning III: Machine learning and causal inference.

- Introduction to causal inference.
- Inferring counterfactuals: synthetic controls and Bayesian structural time series.
- Recurrent neural networks.

NETWORK ANALYSIS

Introduction to network analysis and centrality measurment.

- Building network graphs: nodes and links.
- Centrality metrics and pagerank.