POLS 507: Advanced Topics in Political Methodology II
Nonparametric Models and Machine Learning
126 Herzstein Hall, Thursday 2:30p-5:20p

Dr. Justin Esarey
Assistant Professor of Political Science
Rice University

Office: 108 Herzstein Hall
Office Hours: T F 2:00 – 3:00 PM
E-mail: justin@justinesarey.com
Phone: 678-383-9629

Version: 1/26/2017

Course Objectives and Learning Outcomes

This course introduces students to advanced statistical techniques and their application to problems in political science.

Students will be able to:

1. explain, implement, interpret, and diagnose problems with nonparametric and semi-parametric regression models
2. conduct and interpret non-parametric hypothesis tests (e.g., with bootstrapping, permutation/randomization tests, rank tests)
3. explain classification problems, differentiate them from other forms of prediction, and perform classification using parametric (e.g., logit) and non-parametric (e.g., neural network) techniques
4. Discover latent common dimensions in data using singular value decomposition, principal components analysis, and factor analysis.
5. Understand, explain, and compare all of the following techniques for prediction and explanation:
    a. Classification and Regression Trees
    b. Neural Networks
    c. Bayesian Networks
    d. Instrumental Variable Models
    e. Matching Algorithms
    f. Support Vector Machines
6. Apply all of the models noted to data analysis in the R programming environment.

Grading Policies and Assignment Details

Grade Components:

- Homework: 25%
- Exam 1: 25%
- Exam 2: 25%
- Final Exam: 25%

Grading Scale:

100%-97%: A+	76.9%-73%: C
96.9%-93%: A	72.9%-70%: C-
92.9%-90%: A-	69.9%-67%: D+
89.9%-87%: B+	66.9%-63%: D
86.9%-83%: B	62.9%-60%: D-
82.9%-80%: B-	>59.5%: F
79.9%-77%: C+	

Exams: There will be three exams in this class, two midterms and a final. All exams are cumulative, but will focus on material learned since the last exam. You must complete each exam within twenty-four hours of receipt, and must submit a typed LaTeX answer sheet (see research paper section above for details on software). The exams are open book and open note, but you may not consult anyone for advice on the exam. The rough timing of the exams is indicated on the course outline, and specific times will be scheduled in consultation with the class.

Homework: Homework problem sets will be distributed during class. I encourage collaborative work on problem sets: the goal of a homework problem set is to help you learn the material and enable you to perform well on the (non-collaborative!) research paper. With that said, simply copying another student's homework answers is not permitted and will be treated as academic dishonesty.

All homeworks must be typed in LaTeX.

Attendance: Attendance is mandatory in this class, and as graduate students I expect that attendance will not be a problem for you. Every class you fail to attend (without an acceptable excuse—see below) will result in a 2.5 percentage point deduction from your final grade. (I expect that this will never happen.)

Attendance penalties may be waived in the event of death in the immediate family (parent, spouse, sibling, or child) within 2 weeks before the due date, in the event of an unforeseeable medical emergency affecting yourself, your spouse, or your child, or if you are participating in a pre-approved academic activity (e.g., a conference). Penalty waivers are at the discretion of the instructor. I may require supporting documentation.

COURSE POLICIES

Late Work: Assignments are due at the date and time I specify for the assignment. Late homeworks will be marked off at 5 percentage points for the first 24 hours late, and an additional 10 percentage points for every subsequent 24 hours late. For exams, the first hour late incurs a 5 percentage point penalty and each additional hour incurs a 10 percentage point penalty.

Late work penalties may be waived in the event of death in the immediate family (parent, spouse, sibling, or child) within 2 weeks before the due date, or in the event of an unforeseeable medical emergency affecting yourself, your spouse, or your child. Penalty waivers are at the discretion of the instructor. I may require supporting documentation.

Honor Code/Academic Misconduct: All forms of academic misconduct will be handled according to the Rice University Honor Code. Details on the Honor Code are available at http://honor.rice.edu/honor-system-handbook/.

If you ever have any questions about what you should do to stay within the honor code on a particular assignment, PLEASE contact me with your question and I can assist you. I cannot guarantee a timely response unless you contact me at least 24 hours in advance of the time the assignment is due.

Students with Disabilities: If you have a disability and require accommodation in this class, please contact me as soon as possible (within the first two weeks of class) to discuss these accommodations. You will also need to contact the Disability Support Services Office (telephone extension: 5841) in the Allen Center.

Syllabus Change Policy: The policies of this syllabus (other than absence policies) may be changed by Prof. Esarey with advance notice.

Course Materials

Required Texts:

- Christopher Bishop. 2006. *Pattern Recognition and Machine Learning*. Springer.
- Stephen Marsland. 2009. *Machine Learning: An Algorithmic Perspective*. CRC/Chapman and Hall.

Other readings are available on the Canvas website.

Software: This course will teach material primarily through R. R is free and available from http://cran.r-project.org/.

All homeworks, papers and presentation slides must be typed in LaTeX, but free software exists to make this easy to accomplish. I suggest using LyX (http://www.lyx.org/) in combination with MiKTeX on Windows (http://miktex.org/), MacTeX on Macintosh (http://www.tug.org/mactex/) or TeXLive on Linux (http://www.tug.org/texlive/).

All students must have a valid Rice e-mail address and login (and access to the Canvas website) to participate in this course.

COURSE OUTLINE AND ASSIGNED READINGS

1) January 19: Basic Concepts and Methods of Nonparametric Inference

Questions: What is nonparametric statistics? What is "smoothing?" How can we estimate distributions, bivariate relationships, and multivariate relationships with non-parametric techniques? How do these techniques compare to parametric or structural modeling techniques, and when would we favor one over the other?

Skills and Concepts: kernel density estimation and univariate smoothing; procedures for optimal bandwidth selection; uncertainty in estimates; nonparametric regression.

Readings:

- Hardle, Muller, Sperlich, and Werwatz. 2004. *Nonparametric and Semiparametric Models.* Springer.
    - Chapters 2-4: "Histogram," "Nonparametric Density Estimation," and "Nonparametric Regression."
- Bishop, Chapter 1 Sections 1.1-1.4 and Chapter 6: "Introduction" and "Kernel Methods."

2) January 26: Bootstrapping, Randomization Testing, and other forms of Nonparametric Inference

Questions: How do we quantify our uncertainty about estimated quantities when we cannot write down an analytic form (or, when we suspect that this analytic form is a misspecification)? How do we apply nonparametric estimates of uncertainty to perform statistical hypothesis tests? How can we apply non-parametric regression techniques to more complex data sets?

Skills and Concepts: Mann-Whitney/Wilcoxon rank-sum tests; parametric and non-parametric bootstrapped confidence intervals and p-values; randomization tests for grouped data.

Readings:

- Givens and Hoeting. 2005. *Computational Statistics*. Wiley-Interscience.
    - Chapter 9: "Bootstrapping."
- Rader, Kelly. "Randomization Tests and Inference with Grouped Data." URL: https://www.dropbox.com/s/fmj52w9msu4moj6/randtest-paper.pdf?dl=0
- Davison and Hinkley. 1997. *Bootstrap Methods and their Application.* Cambridge.
    - Chapters 2, 4, and 5: "Basic Bootstraps," "Tests," and "Confidence Intervals."

3) February 16: Ensemble Models and Bayesian Model Averaging

Questions: Are two (or more) models better than one? How can we combine the results of multiple models in order to achieve more accurate predictions?

Skills and Concepts: boosting; Bayesian model averaging; the `ada`, `gbm`, `flexmix`, and `BMA` packages in R

Readings:

- Bishop, Chapter 14 ("Combining Models"), Sections 14.1-14.3 ("Bayesian Model Averaging," "Committees," and "Boosting").
- Marsland, Chapter 7 ("Decision by Committee: Ensemble Learning").
- Hoeting et al. 1999. "Bayesian Model Averaging." *Statistical Science* 14: 382-401. Available on-line at http://www.stat.colostate.edu/~jah/papers/statsci.pdf.

4) February 23: Kernel-Regularized Least Squares

Questions: How can least-squares concepts be merged with non-parametric regression concepts to achieve an estimate of statistical relationships that is more determined by data and less determined by model assumptions?

Skills and Concepts: radial basis functions, RBF networks, kernel-regularized least squares

Readings:

- Marsland, Chapter 4: "Radial Basis Functions and Splines."
- Hainmueller and Hazlett. 2014. "Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach." *Political Analysis* 22(2): 143-168. URL: http://pan.oxfordjournals.org/content/22/2/143.abstract.

*************** **CUTOFF FOR EXAM 1** ***************

5) March 2: "Causal Inference" and Matching

Questions: What does it mean to draw a "causal inference"? When are models capable of drawing causal inferences, and are some models more robust than others? What is matching, and how does it statistically reconstruct a quasi-experimental design? What are the Average Treatment Effect, the Average Treatment Effect on the Treated, and the Intention-to-Treat Effect and how do they differ? How do we get these effects out of parametric and structural models?

Skills and Concepts: the Neyman-Rubin causal model; matching methods and techniques; balancing and other metrics of matching model assessment; calculating and interpreting the ATE, ATT, and ITT from various models; propensity score matching and coarsened exact matching.

Readings:

- Morgan and Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge.
    - Chapters 2-4: The Counterfactual Model," "Estimating Causal Effects by Conditioning," and "Matching Estimators of Causal Effects."
- Iacus, Stefano M., Gary King, and Giuseppe Porro. 2012. "Causal Inference Without Balance Checking: Coarsened Exact Matching." *Political Analysis* 20(1): 1-24. URL: http://pan.oxfordjournals.org/content/20/1/1

6) March 9: Instrumental Variable Models

Questions: How can we use instrumental variable models to overcome problems of endogeneity and confounding to estimate a causal relationship?

Skills and Concepts: inferential problems stemming from endogeneity; solving confounding and endogeneity problems with instrumental variables; two stage least squares analysis

Readings:

- Gujarati and Porter, *Basic Econometrics*: Chapter 13, Section 13.5; Chapter 20
- Angrist, Joshua D. and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Chapter 4: "Instrumental Variables in Action," pp. 113-220.

7) March 16: Causality, Bayesian Networks, and Recovering Bayes' Nets from Observational Data

Questions: To what extent can causal inferences be derived from observational data? How does the concept of conditional (in)dependence enable some causal inferences to be derived?

Skills and Concepts: Pearl's structural causality model; the "back door" and "front door" criteria for causal inference; using the `bnlearn` library in R

Readings:

- Judea Pearl, "The Structural Theory of Causation." Chapter 33 in Illari, Russo, and Williamson. 2011. *Causality in the Sciences*. Oxford.
- Bishop, Chapter 8 ("Graphical Models").
- Marsland, Chapter 15, Section 15.1 ("Bayesian Networks").
- Mario Scutari. 2010. "Learning Bayesian Networks with the bnlearn R Package." *Journal of Statistical Software* 35(3): http://www.jstatsoft.org/v35/i03/paper.

8) March 23: Assessing the Fit of Discrete Choice Models

Questions: How (predictively) accurate are categorizations produced from a discrete choice model? How does predicting probabilistic outcomes differ from categorization?

Skills and Concepts: logistic classification; classification and probability assessment techniques (ROC, PCP, H-L, heat mapping); using the `heatmapFit` library in R

Readings:

- Hosmer, David, and Stanley Lemeshow. 2000. *Applied Logistic Regression, 2nd ed.* Chapter 5: "Assessing the Fit of the Model," pp. 143-202.
- Esarey, Justin, and Andrew Pierce. 2012. "Assessing Fit Quality and Testing for Misspecification in Binary Dependent Variable Models." *Political Analysis* 20(4): 480-500.

*************** **CUTOFF FOR EXAM 2** ***************

9) March 30: Identification and Measurement of Latent Continuous Characteristics

<u>Questions</u>: How do we identify the important (and latent) characteristics of a group using observational data?

<u>Skills and Concepts</u>: factor analysis/principal component analysis; singular value decomposition; probabilistic PCA; using `prcomp`, `factanal`, and the `pcaMethods` library in R

<u>Readings</u>:

- Bishop, Chapter 12 ("Continuous Latent Variables").
- Marsland, Chapter 10 ("Dimensionality Reduction").
- Stacklies et al. 2007. "pcaMethods—a bioconductor package providing PCA methods for incomplete data." *Bioinformatics* 23(9): 1164-1167.

10) April 6: Identifying Classes with Neural Networks

<u>Questions</u>: How can neural network models be used to reveal hidden categorical structures in data?

<u>Skills and Concepts</u>: theory of neural networks, including the single and multi-layer perceptron; fitting and interpreting neural networks; using the `neuralnet` library in R

<u>Readings</u>:

- Marsland, Chapters 1-3 ("Introduction," "Linear Discriminants," and "The Multi-Layer Perceptron").
- Bishop, Chapter 5 ("Neural Networks").
- Kutner, Nachtsheim, Neter, and Li. *Applied Linear Statistical Models, 5th ed.* Chapter 13, Section 6 ("Introduction to Neural Network Modeling").
- Gunther and Fritsch. 2010. "neuralnet: Training of Neural Networks." *The R Journal* 2(1): 30-38. Available on-line at http://journal.r-project.org/archive/2010-1/RJournal_2010-1_Guenther+Fritsch.pdf.

11) April 13: Tree-based Methods for Discrete and Continuous Classification

Questions: How are data used to develop a sequential decision-based prediction algorithm? How do the results of this procedure compare to more parametric (viz., GLM) and less parametric (viz., kernel regression) methods?

Skills and Concepts: decision trees; CART methodology; conditional inference trees; the `tree`, `randomforest`, and `party` libraries in R

Readings:

- Marsland, Chapter 6 ("Learning with Trees").
- Bishop, Chapter 14, Section 14.4 ("Tree-based Models").
- Veables and Ripley. 2002. *Modern Applied Statistics with S, 4th ed.* Chapter 9 ("Tree-Based Methods").
- Torsten Hothorn, Kurt Hornik, and Achium Zeileis. "Unbiased Recursive Partitioning: A Conditional Inference Framework." Available on-line at http://statmath.wu-wien.ac.at/~zeileis/papers/Hothorn+Hornik+Zeileis-2006.pdf.
- Andy Liaw and Matthew Wiener. "Classification and Regression by randomForest." *R News*, December 2002: pp.18-22. Available on-line at http://goo.gl/DOWIV.

12) April 20: Support Vector Machines

Questions: What are support vector machines, and how can they be used to improve the classification of complex data?

Skills and Concepts: support vector machines

Readings:

- Bishop, Chapter 7, part 7.1 ("Maximum Margin Classifiers")