POLS 8500 Applied Machine Learning

Professor L. Jason Anastasopoulos
ljanastas@uga.edu
http://scholar.harvard.edu/janastas
Office: Baldwin Hall 414
Office Hours: 5:15–6:15 Friday or by appointment

Class: 3:30–6:15pm Thursday
118 Journalism Building

Prerequisites

A course in linear algebra, at *least* two courses on probability and statistical inference. Programming experience will be helpful but not necessary. All of the machine learning algorithms and data pre-processing will be implemented in `R` but you are free to use `Python` if you choose.

Course Overview and Objectives

This course will provide an introduction to the theory and applications of some of the most popular machine learning algorithms with a focus on applications to social science data and problems.

The goals of this course include:

- Developing a basic understanding of the statistical theory underlying popular 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 for numerical, text and image data.

Required Texts

Hastie, Tibshirani and Friedman. 2013. *The Elements of Statistical Learning* (2nd ed), 7th Printing. Springer Series in Statistics. Available for free here: http://statweb.stanford.edu/ tibs/ElemStatLearn/printings/ESLII_print10.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**.

Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008. http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf.
Referred to in the schedule as **MRS**.

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**.

In addition to these books assigned readings will be available here:
http://scholar.harvard.edu/janastas/teaching
or as links in the course syllabus.

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 extremely important. If you cannot attend a lecture you must present me with a valid excuse at least 24 hours prior to the start of class unless the situation you encountered was an emergency. Either way, absence requires explanation and documentation if you do not want want points taken off of your final grade.

Computer, Tablet and Cell Phone Use Policy

Laptop computers and tablets may be used during class sessions for note taking ONLY. ANY instance of unapproved use of laptop computers or tablets in the classroom will result in your laptop/tablet privileges being revoked for the remainder of the semester. 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 University of Georgia student, you have agreed to abide by the University's academic honesty policy, "A Culture of Honesty," and the Student Honor Code. All academic work must meet the standards described in "A Culture of Honesty" found at: www.uga.edu/honesty. 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 Disability Resource Center (http://drc.uga.edu/students/register-for-services) to develop an accommodation plan. The student is responsible for providing a copy of that plan to the instructor.

Make-up exams and Incomplete or "I" grades are permitted in only extremely rare circumstances. The instructor has the right to (1) require documentation and proof of the need for the make-up exam or "I" grade (2) require the completion of different versions of assignments missed and/or (3) impose a grade penalty for a missed exam or Incomplete grade in the course. Please let the instructor know as soon as you see a problem developing. Any students wishing to withdraw from the course must follow the University's course withdrawal procedures.

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 poster version of the project that will be presented to the class and others in the university towards the end of the semester and;

(2) A paper which will be handed in for a grade. The paper should resemble a polished draft that you would submit to an academic conference.

Grades

Attendance and participation	5%
Problem Sets	50%
Final Project Proposal	10%
Final Project Poster	10%
Final Project Paper	25%

Overview of Topics

- Review of probability theory, statistics and linear algebra.
- Introduction to statistical learning theory.
- Linear regression, optimization with gradient descent and stochastic gradient descent.
- Model selection, regularization and cross-validation.
- Logistic regression, linear discriminant analysis and naive Bayes.
- Applications: Classifying text data I – Intro to NLP and text-as-data.
- Applications: Classifying text data II – Logistic regression and naive Bayes.
- Support vector machines.
- Applications: Classifying text data III – Support vector machines.
- Neural networks.
- Applications: Text and Image Data – Recurrent and Convolutional Neural Networks.
- Dirichlet processes and topic modeling.
- EM algorithm, K-Means Clustering, Hierarchical Clustering.
- Principal Components, Multidimensional Scaling, Google PageRank Algorithm.

Tentative Schedule

All readings and topics are tentative and may change during the semester.

01.05 **Course Overview**

Topics

What is machine learning?

Supervised & unsupervised learning.

Inference versus prediction.

Readings

O'Connor, Brendan. 2008. "Statistics v. Machine Learning, Fight!" Link

Breiman, Leo. 2001. "Statistical Modeling: The Two Cultures." *Statistical Science*. Link.

JWHT – Introduction, pp 1-15.

01.12 **Probability, Statistics and Linear Algebra Review, Introduction to R**

Topics

Review of probability theory and linear algebra.

Intro to programming in R.

Useful APIs and databases for political science research, JSON.

Readings

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

Christopher M. Bishop. 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA. Chapters 1.2.1-1.2.4, 2.1-2.2, 2.4,2.5.1.

Additional resources and documentation

`rsunlight` - a package for extracting data from Sunlight APIs (OpenStates, Congress and Capitol Words.
https://cran.r-project.org/web/packages/rsunlight/rsunlight.pdf.

`httr` - useful package for extracting data from APIs in JSON format using url's.
https://cran.r-project.org/web/packages/httr/httr.pdf.

Openstates.
https://openstates.org/

01.19 **Introduction to Statistical Learning Theory**

Topics

What is statistical learning?

Assessing model accuracy.

Bias-Variance Trade-Off.

Readings

JWHT – Statistical Learning, pp 15-37; *Optional*, Introduction to R Lab, pp 42-49 (*strongly* recommended for those without `R` experience.

Optional, In-Depth Treatment Bousquet, O., Boucheron, S. and Lugosi, G., 2004. Introduction to statistical learning theory. In Advanced lectures on machine learning (pp. 169-207). Springer Berlin Heidelberg.
Link.

Supervised Learning

01.26 **Linear Regression**

Topics

Prediction with linear regression.

Gauss-Markov.

Optimization: gradient descent, stochastic gradient descent and normal equations.

Readings

HTF – pp 43-56.

Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.
https://arxiv.org/pdf/1609.04747v1.pdf.

Optional, In-Depth Treatment Bottou, L., 2010. Large-scale machine learning with stochastic gradient descent. In Proceedings of COMPSTAT'2010 (pp. 177-186). Physica-Verlag HD.
http://leon.bottou.org/publications/pdf/compstat-2010.pdf.

02.02 **Linear Model Selection, Regularization and Cross Validation**

Topics

Feature selection.

Shrinkage methods: ridge regression, the LASSO, tuning parameters.

Dimensionality reduction methods.

High dimensional data.

Readings

JWHT – Chapter 5.1, 203-243.

02.09 **Logistic Regression, Linear Discriminant Analysis & Näive Bayes**

Topics

Logistic regression.

Linear discriminant analysis and näive Bayes.

Readings

JWHT – 127-151.

Ng, Andrew and Michael Jordan. 2002. On discriminative vs. generative classifiers: A comparison of logistic regression and näive bayes. Advances in neural information processing systems, 14, p.841.
http://machinelearning.wustl.edu/mlpapers/paper_files/nips02-AA28.pdf.

02.16 **Applications: Classifying Text Data I - Intro to NLP and Text As Data**

Topics

Text as data and intro to natural language processing.

Text preprocessing.

Text tagging with hidden markov models (HMMs).

Vector space model and tf-idf weighting.

Readings

MRS. Chapter 2.
http://nlp.stanford.edu/IR-book/pdf/02voc.pdf

MRS. Chapter 6.
http://nlp.stanford.edu/IR-book/pdf/06vect.pdf.

Basic Text Mining in R. Do examples from "pre-processing" and "explore your data" sections.
https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html

Additional Resources

Stanford Univ., CS 276: Information Retrieval and Web Search.
http://web.stanford.edu/class/cs276.

02.23 **Applications: Classifying Text Data II - Logistic regression and näive Bayes**

Topics

Classification of political texts with logistic regression and näive Bayes.

Readings

MRS. Chapter 3.
http://nlp.stanford.edu/IR-book/pdf/13bayes.pdf

Xiaojin Zhu, *Text Categorization with Logistic Regression.*
http://pages.cs.wisc.edu/ jerryzhu/cs838/LR.pdf

03.02 **Support Vector Machines**

Readings

JWHT. 337-366.

03.09 **No class, spring break.**

03.16 **Applications: Classifying Text Data III - Support Vector Machines**

MRS. Chapter 15.
http://nlp.stanford.edu/IR-book/pdf/15svm.pdf.

03.16 **Neural Networks**

Readings

KJ. 141-145.

HTF. 389-414.

- **Additional resources**

  Hamed Hashemina. Neural Networks in R Tutorial.

03.23 **Applications: Classifying Text and Image Data - Recurrent and Convolutional Neural Networks**

Readings

Karpathy, Andrej. Convolutional Neural Networks for Visual Recognition.

Karpathy, Andrej. The unreasonable effectiveness of recurrent neural networks.

Optional, Advanced Readings. Goodfellow, Benigo and Courville. *Deep Learning*. 2016. MIT Press. Chapter 9: Convolutional Neural Networks.

Optional, Advanced Readings. Goodfellow, Benigo and Courville. *Deep Learning*. 2016. MIT Press. Chapter 10: Recurrent and Recursive Nets.

Unsupervised Learning

03.30 **Dirichlet Processes and Topic Modeling**

Topics

Introduction to non-parametric Bayesian methods and the Dirichlet processes.

Topic modeling with the latent Dirichlet allocation (LDA).

Structural topic models.

Readings

Blei, D.M., 2012. Probabilistic topic models. Communications of the ACM, 55(4), pp.77-84. https://www.cs.princeton.edu/ blei/papers/Blei2012.pdf.

Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), pp.993-1022.

Teh, Y.W. Dirichlet Processes.

Roberts, et. al., 2014. Structural Topic Models for OpenEnded Survey Responses. American Journal of Political Science, 58(4), pp.1064-1082.

04.06 **EM Algorithm, K-Means Clustering, Hierarchical Clustering**

Topics

The EM algorithm in general and its relationship to K-Means.

K-Means and hierarchical clustering.

Readings

HTF – Chapter 8.5.

JWHT – Chapters 10.3, 10.5

04.13 **Principal Components, Multidimensional Scaling, Google PageRank**

- **Readings**

  HTF – Chapters 14.5, 14.7-14.10.

04.20 **Final project poster session.**

05.01 **Final project paper due.**