Machine Learning
PSCI 8358
Summer 2017
MTWR 1:10-4:00 PM
Location: Commons 363

Jonathan Klingler
301E Commons
jonathan.d.klingler@vanderbilt.edu

Office Hours
4:30-5:30 PM Wednesdays
or by appointment

Course Description

Perception and reasoning by intelligent systems requires the processing of inputs into a reasonable mapping of the world, learning from experience, and management of the uncertainty that pervades the world around us. This course will provide a firm introduction to machine learning and the design of intelligent systems capable of handling these three tasks. The course will engage with machine learning by developing the theoretical and mathematical underpinnings of a variety of algorithms, and apply successful algorithms in several contexts to witness their limitations and strengths. This course is appropriate for graduate students and computer science undergraduates comfortable with programming who have basic training in probability theory and familiarity with linear algebra and calculus.

Objectives

- To acquaint students with the fundamentals of machine learning
- To introduce students to key concepts and research in the design of intelligent systems
- To provide students with the ability to write programs applying a variety of machine learning algorithms to relevant social science phenomena
- To develop students' critical thinking and analytical skills through class discussion and assessment of the strengths and weaknesses of machine learning algorithms across applications

Required Texts
Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning.*
Guller, Mohammed. *2015. Big Data Analytics with Spark.*

Suggested Resources
Kriesel, David. A Brief Introduction to Neural Networks
http://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf
Andrew Ng. *Machine Learning.* www.coursera.org/learn/machine-learning
Mueller, John Paul, and Luca Massaron. *2016. Machine Learning for Dummies.*
Bishop, Christopher. 2007. *Pattern Recognition and Machine Learning*

Grading
Participation 10%; Short Assignments 30%, Cumulative Project 60%.

Reading and any additional assignments should be completed before class on the day assigned. All required texts are held at the Science Library on two hour reserve. Pay careful attention to the syllabus and to any adjustments that may occur. You are responsible for material presented in lectures and discussed in class regardless of your attendance. Notes will not be supplied for students who have missed class, so plan to make other arrangements. Late and makeup assignments will be allowed only with a doctor's note or other equally serious documented reason for the absence. Athletes and other students who will miss class due to extra- or co-curricular travel should inform me of the relevant dates early in the semester. Students with any other special needs should meet with me early in the semester so that proper accommodations can be made.

Class participation and attendance count for 10% of your final grade. Obviously those who do not attend class cannot participate in class discussion. Attendance in class will be taken regularly throughout the semester. Because the exchange of ideas is an important part of learning, you are encouraged to frequently ask questions and share your informed opinions during regular class. In order to maintain an atmosphere conducive to learning, students should use professional language in class discussions and written work. No offensive slang or profanity is permitted. Disagree without being disagreeable.

Four short assignments will be used to calculate 30% of your final grade. These short assignments consist of problem sets focused on the mathematical underpinnings of the principles discussed in class. Students will receive ample notice if preparatory work outside of class is required, as may occasionally be the case. The lowest short assignment grade will be automatically dropped. Missed short assignments cannot be made up and will be graded as zeros.

The cumulative project constitutes 60% of your final grade. Students will select a dataset from several available on the course website, and over the semester are responsible for analyzing these data in two cumulative sections. In the first section, a supervised learning technique will be used to predict a phenomenon in the dataset. In the second section, unsupervised learning will be used to describe the data. Proposals for the cumulative project must be submitted on July 18, and the two sections of the cumulative project are due on August 1st and August 11th. A presentation of both sections of the cumulative project must be given to the class on August 11th.

Academic honesty is expected. Turn in material that you have completed yourself. Absolutely no cheating or plagiarism (using someone else's words or ideas without proper citation) will be tolerated. Both cheating and plagiarism are serious offenses that will be reported for disciplinary action. Please refer to the University Handbook and/or speak with the instructor if you have any questions in this area.

Course Outline

Mathematical and Technical Foundations

July 11
Introduction and Course Overview
Hastie et al. Chapters 1 and 2.

July 13
Software Tutorial
Mueller and Messaron. Chapters 4-7.

July 14
Linear Algebra Review
Univariate Linear Regression

July 17

Multivariate Linear Regression
Hastie et al., Chapter 3. Linear Methods for Regression.

July 18 – Project Proposals Due
Logistic Regression
Hastie et al. Chapter 4. Linear Methods for Classification.

Supervised Learning

July 20
Regularization
Hastie et al. Chapter 5. Basis Expansions and Regularization.

July 21
Best Practices in the Application of Machine Learning
DeTurck, Dennis. "The 1936 Literary Digest Poll"
www.math.upenn.edu/~deturck/m170/wk4/lecture/case1.html
Wolpert, David. "The Lack of A Priori Distinctions between Learning Algorithms"
Ng. Advice for Applying Machine Learning.
Scikit-learn Users Guide Chapter 3: http://scikit-learn.org/stable/model_selection.html

July 24
Neural Networks: Representation
Kriesel Chapters 1-3

July 25
Neural Networks: Learning
Hastie et al. Chapter 11. Neural Networks.

July 27
Support Vector Machines
Hastie et al. Chapter 12. Support Vector Machines and Flexible Discriminants.

July 28
Improving Model Performance
Hastie et al. Chapter 13. Prototype Methods and Nearest-Neighbors.
Ng. Machine Learning Systems Design.

July 31
Ensemble Learning
Hastie et al. Chapter 15 and 16.

Unsupervised Learning

August 1 - Project Section #1 Due
Unsupervised Machine Learning
Hastie et al. Chapter 14, pp. 485-528

August 3

Principal Components Analysis
Hastie et al. Chapter 14. pp. 529-577

Applications and Practical Considerations

August 4
Recommendation Algorithms
MovieLens Dataset: http://grouplens.org/datasets/movielens

August 7
Working with Images
Scikit Image Tutorial (http://scipy-lectures.github.io/packages/scikit_image)
Eigenfaces Tutorial (http://scikit-learn.org/stable/auto_examples/decomposition/plot_faces_decomposition.html)

August 8
Managing Big Data
Guller Chapters 1-5, and 8

August 10 – Project Section # 2 Due
Presentations