Introduction to Machine Learning

CptS 437

Fall 2022

Tuesday / Thursday, 9:10-10:25, Cleveland 30

Course Overview

Machine learning is the study of computer algorithms and models that learn automatically from data. It is a key area of artificial intelligence and has applications in many domains, including biology, social science, statistics, and image processing. This introductory course covers key topics in machine learning, including linear models for regression and classification, decision trees, support vector machines and kernel methods, neural networks and deep learning, ensemble methods, unsupervised learning and dimension reduction.

Course Instructor

Instructor: Diane Cook

Office hours: Tu/Th 10:30-11:00, EME 121

Contact: djcook@wsu.edu

Teaching assistant: Ramesh Sah

Office hours: M/Tu 2:30-4:00, EME 136

Contact: ramesh.sah@wsu.edu

Prerequisites

Required: CptS 223 or CptS 233 or CptS 215 (or equivalent). In addition, students are expected to have some familiarity with basic linear algebra (vectors, matrices, matrix-vector computations, vector and matrix norms, linear independence), multivariate calculus (derivatives of univariate functions, derivatives of multivariate functions, chain rule), and basic probability and statistics (discrete and continuous probability distributions, sum rule, product rule, marginal probability distributions, conditional probability distributions, joint probability distributions, independence and conditional independence, Bayes Theorem, variance and covariance, expectation).

Required Instructional Material

Required textbook: Hal Daumé, A Course in Machine Learning, 2017. Available for download from http://ciml.info/. Additional online materials may also be recommended for individual lectures, see course schedule.

Specific Course Learning Outcomes and Assessments

Following completion of this course, students should (1) have an understanding of major supervised, unsupervised and reinforcement learning techniques, (2) have a basic understanding of evaluation methodologies, (3) have a working knowledge of how to apply machine learning technologies to real-world datasets, and (4) have gained experience designing and applying machine learning techniques in team settings.

This class provides a unique opportunity to strengthen skills in each of the WSU Seven Learning Goals and Outcomes: 1) Critical and Creative Thinking, 2) Quantitative Reasoning, 3) Scientific Literacy, 4) Information Literacy, 5) Communication, 6) Diversity, and 7) Depth, Breadth, and Integration of Learning. The methods and measures for each goal is summarized in the table.

WSU Learning Outcome	Goal (by end of course)	Course topics that address the learning outcome	Evaluation
Critical and Creative Thinking	Understand the method and applicability of alternative machine learning strategies	• Decision trees, nearest neighbors, k-means cluster, neural network, linear regression, logistic regression, SVMs	• Homework assignments • Exams • Project
Quantitative Reasoning	Grasp properties involved in algorithm assessment	• Decision boundaries, margin, performance measures, validation	• Homework assignments • Exams • Semester project
Scientific Literacy	Be aware of and understand state-of-the-art research in machine learning	• Guest lectures on dep learning, generative adversarial networks, tensor flow	• Exams
Information Literacy	Be able to access and utilize literary resources to understand a machine learning challenge	• Research projects	• Semester project
Communication	Present the results of a research project and service learning orally and in writing	• Research project	• Project poster presentation • Project demonstration
Diversity	Be aware of ethical issues related to machine learning	• Lectures on supervised and unsupervised learning	• Exams • Semester project
Depth, Breadth, and Integration of Learning	Understand issues related to practical application of machine learning technologies	• Multi-disciplinary research project	• Semester project

Course Requirements

(1) *Homework Assignments (35%)*. You will be assigned six homework assignments to complete. All assignments will have written components and programming components. The homework assignments will expose you to the machine learning methods we discuss in class and data from a diversity of applications that illustrate how the methods can be used. All programs will be written in Python. They will be assigned and submitted using Google's Collaboratory online Python programming environment. Completed homework assignments are due, through Canvas, by 11:59pm on the due date.

(2) *Three Midterm Exams (40%)*. Three online (through Canvas) exams will be given during the semester. The exams will cover all class material up to the lecture prior to the exam date.

(3) *Semester Project (25%)*. To obtain experience designing, applying, and evaluating machine learning techniques, you will complete a semester project. This project will take the place of a final exam. See the Semester Project section below for additional details.

Semester Grades

A	A-	B+	B	B-	C+	C	C-	D	F
≥93%	90%	87%	83%	80%	77%	73%	70%	60%	≤60%

Semester Project

A requirement for this class is that you design and complete a machine learning project (graded out of 100 points). Each project will include implementation of a machine learning technique not described in class or enhancement of a described technique, with application to a real-world dataset or problem. Students are encouraged, but not required, to work in teams consisting of 2-3 students. Due dates related to the project are listed below.

- October 6: Project ideas and requirements will be summarized in class.
- November 15: Project proposal due as part of Homework #5 (see assignment for details).
- December 15: Project Presentations, Code, and Demo due. Project teams will give a short (5-8 minute) talk about their project. The presentation should clearly state the project objectives and hypothesis, datasets used or created, methods, and results with goals for next steps. Teams will post a link to their video presentation on the discussion board by the end of the day. Also, provide a link to working code with instructions on running it (if the code is self-contained in Colab) and/or a video demonstrating how to run the code.
- December 16: Optional feedback on projects due by the end of the day.

Grades are determined based on:

- Proposal (10 points)
- Presentation (30 points)
- Well structured, working code with demo and documentation (40 points)
- Scope of the project: depth, difficulty, creativity (20 points)

Exams

Exams will be available on Canvas at the beginning of class on the exam date. All students must be on Zoom during the entire exam period, with their cameras turned on. Exams will not be graded for students who are not present during the exam period. Canvas will give you 90 minutes to complete the exam (this gives you extra time to deal with any technical difficulties that arise). Canvas includes a link for you to optionally upload a file (pdf format only) to show your work. The pdf document will not be read during grading but will be there if you need to justify your answers.

Policy Regarding Late Work

Assignments should be uploaded by 11:59pm on the due date. After that, 15% will be deducted per day for the first two days. Assignments turned in more than two days late will not be graded.

Students with Disabilities

Reasonable accommodations are available for students with a documented disability. If you have a disability and may need accommodations to fully participate in this class, please either visit the Access Center (Washington Building 217) or call 509-335-3417 to make an appointment with an Access Advisor. All accommodations MUST be approved through the Access Center.

Academic Integrity Policy

Academic integrity is the cornerstone of higher education. As such, all members of the university community share responsibility for maintaining and promoting the principles of integrity in all activities, including academic integrity and honest scholarship. Academic integrity will be strongly enforced in this course. Students who violate WSU's Academic Integrity Policy (identified in Washington Administrative Code (WAC) 504-26-010(3) and -404) will fail the assignment, will not have the option to withdraw from the course pending an appeal, and will be reported to the Office of Student Conduct.

Cheating includes, but is not limited to, plagiarism and unauthorized collaboration as defined in the Standards of Conduct for Students, WAC 504-26-010(3). You need to read and understand all of the definitions of cheating: http://app.leg.wa.gov/WAC/default.aspx?cite=504-26-010. If you have any questions about what is and is not allowed in this course, you should ask course instructors before proceeding. If you wish to appeal a faculty member's decision relating to academic integrity, please use the form available at conduct.wsu.edu.

Safety Information

Washington State University is committed to maintaining a safe environment for its faculty, staff, and students. Safety is the responsibility of every member of the campus community and individuals should know the appropriate actions to take when an emergency arises. In support of our commitment to the safety of the campus community the University has developed a Campus Safety Plan, http://safetyplan.wsu.edu. It is highly recommended that you visit this web site as well as the University emergency management web site at http://oem.wsu.edu/ to become familiar with the information.

Course Calendar

Black = in person
Blue = holiday

Date	Topic	Reading	Due
8/23	Syllabus, Introduction	Daumé Chapter 1	HW #1 assigned
8/25	Python / Colab overview		
8/30	Decision trees	Mitchell Chapter 3 [1]	
9/1	Limits of learning, inductive bias, underfit/overfit	Daumé Chapter 2	
9/6	Geometry and nearest neighbors, decision boundaries	Daumé Chapter 3	
9/8	K-means clustering, curse of dimensionality, sklearn		
9/13	Perceptron	Daumé Chapter 4	HW #1 due HW #2 assigned
9/15	Linear separability, margin, averaged perceptron		
9/20	Practical issues, class imbalance	Daumé Chapter 5 Supplemental material [2]	
9/22	Exam 1 (Canvas / Zoom)		
9/27	Ranking		HW #2 due HW #3 assigned
9/29	Evaluating model performance		
10/4	Significance testing, confidence intervals, bootstrapping, multi-class classification	Daumé Chapter 6, Supplemental material [3]	
10/6	Linear regression	Daumé Chapter 7	
10/11	Loss functions, regularization, bias and fairness	Daumé Chapter 8	HW #3 due HW #4 assigned
10/13	Naïve Bayes classifier, text mining	Daumé Chapter 9	
10/18	Logistic regression		
10/20	Support vector machines	Supplemental material [4]	
10/25	Exam 2 (Canvas / Zoom)		
10/27	Neural networks, backpropagation	Daumé Chapter 10, Supplemental material [5-7]	
11/1	Deep networks, CNN, image processing		HW #4 due HW #5 assigned
11/3	TensorFlow, Pytorch		
11/8	RNN, autoencoder, GAN		
11/10	Ensemble methods	Daumé Chapter 13	
11/15	K-means++, dimensionality reduction	Daumé Chapter 15	HW #5 due HW #6 assigned

11/17	PCA	Supplemental material [8,9]	
11/22	Thanksgiving		
11/24	Thanksgiving		
11/29	Exam 3 (Canvas / Zoom)		
12/1	Reinforcement learning	Supplemental material [10]	
12/6	Reinforcement learning, conclusions		HW #6 due
12/8	Extended office hours		Presentation videos due
12/15	Project demo and code due	Final code, results, demo due	

[1] http://www.cs.princeton.edu/courses/archive/spr07/cos424/papers/mitchell-dectrees.pdf
[2] http://cs229.stanford.edu/materials/ML-advice.pdf
[3] http://cs229.stanford.edu/notes2020spring/cs229-notes1.pdf (Part I, Section 1)
[4] http://cs229.stanford.edu/notes2020spring/cs229-notes1.pdf (Part II, Section 5)
[5] https://www.researchgate.net/publication/285164623_An_Introduction_to_Convolutional_Neural_Networks
[6] https://ip.cadence.com/uploads/901/cnn_wp-pdf
[7] https://cs.stanford.edu/~quocle/tutorial2.pdf
[8] http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf
[9] http://incompleteideas.net/book/bookdraft2017nov5.pdf
[10] https://arxiv.org/abs/1812.02849