CS 4/5780: Intro to Machine Learning

Instructor: Kilian Q. Weinberger

Contact: cs4780staff@gmail.com

Course staff office hours: Calendar link

Office hours: Mondays 9:00 - 10:00 am (Booking Link) in 410 Gates Hall

Lectures: Tuesday and Thursday from 1:00 pm to 2:15 pm in Uris Hall G01.

Course overview: The course provides an introduction to machine learning, focusing on supervised learning and its theoretical foundations. Topics include regularized linear models, boosting, kernels, deep networks, generative models, online learning, and ethical questions arising in ML applications.

Prerequisites: probability theory (e.g. BTRY 3080, ECON 3130, MATH 4710, ENGRD 2700), linear algebra (e.g. MATH 2940), calculus (e.g. MATH 1920), and programming proficiency (e.g. CS 2110).

Course logistics: For enrolled students the companion Canvas page serves as a hub for access to Ed Discussions (the course forum), Vocareum (for course projects), Gradescope (for HWs), and paper comprehension quizzes. If you are enrolled in the course you should automatically have access to the site. Please let us know if you are unable to access it.

Homework, projects, and exams

Your grade in this course is comprised of three components: homework, exams, and projects. Please also read through the given references in concert with the lectures.

Homework:
There will be a number of homework assignments throughout the course, typically made available roughly one to two weeks before the due date. The homework primarily focuses on theoretical aspects of the material and is intended to provide preparation for the exams. Homework may be completed in groups of up to three. The assignments themselves will be made available via Gradescope (through Canvas). You are allowed two slip days per homework.

Projects
To provide hands on learning with the methods we will discuss in class there are a number of programming projects throughout the course. The projects may be completed solo or in a

group of two. They are accessed, submitted, and graded using Vocareum. You are allowed two slip days per project.

Paper comprehension

Students enrolled in this course at the graduate level (i.e., enrolled in 5780) are required to read assigned research papers and complete the associated online quiz. Papers will be assigned roughly once every two to three weeks. You are allowed two slip days per quiz.

Exams:

There will be two exams for this class, an evening prelim and a final exam. The location and times for both are **to be determined.**

Midterm exam: March 16, 2023
Final exam: **TBD**

Grading

Final grades are based on homework assignments, programming projects, and the exams. For the 5780 level version of the course, the research comprehension quizzes will also factor in.

For CS 4780 your final grade consists of:
Exams: 48%
Homework: 15%
Projects: 37%

For CS 5780 your final grade consists of:
Exams: 45%
Homework: 10%
Projects: 35%
Paper comprehension: 10%

Undergraduates enrolled in 4780 may choose to do the paper comprehension assignments; if completed you will receive the higher of your two grades between the above schemes.

Schedule

A tentative schedule follows, and includes the topics we will be covering, relevant reference material, and assignment information. It is quite possible the specific topics covered on a given day will change slightly. This is particularly true for the lectures in the latter part of the course, and this schedule will be updated as necessary. **Please note that the due dates here are mostly correct, but may change. Check Canvas for any changes to assignment due dates.**

Date	Topic	References	Notes/assignments
1/24/23	Introduction	PML: 1.1; ESL: Ch. 1; and PPA: Ch. 1	
1/26/23	ML Basics	PML: 1.2, and ESL: 2.1 and 2.2.	html pdf handwritten
1/31/23	K Nearest Neighbors and the curse of dimensionality	PML: 16.1	html pdf handwritten 5780: Cover and Hart 1967
2/2/23	The Perceptron	Wikipedia article	html pdf handwritten
2/7/23	Clustering: K-means	ESL: 14.3.6 and 14.3.7, and PML: 21.3	Project 0 due html handwritten
2/9/23	Principal Component Analysis	PML: 20.1, ESL: 14.5.1 and 14.5.2	html handwritten
2/14/23	MLE and MAP	Nice Youtube video for MLE and MAP. Ben Taskar's lecture notes. Tom Mitchell's book chapter on MLE and MAP ESL: 8.2.2-8.3	html pdf Homework 1 due; Project 1 due
2/16/23	MLE and MAP continued		Cover and Hart reading quiz due
2/17/23	Naive Bayes	ESL: 6.6.3, and Tom Mitchell's book chapter	P1 Due html pdf
2/21/23	Naive Bayes	ESL: 4.4, and PML: 10.1, 0.2, and 10.3	html pdf
2/23/23	Logistic Regression and Gradient descent	PML: 8.1, 8.2, and 8.3 Tom Mitchell's book chapter on Naive Bayes and Logistic Regression;	Homework 2 due Project 2 due Eiganfaces Paper Reading Quiz 2 due html pdf html pdf

Date	Topic	References	Notes/assignments
2/28/23	**February break, no class**		
3/2/23	Newton's method. AdaGrad	PML: 8.1, 8.2, 8.3, and 8.4 (specifically, see PML 8.4 for SGD)	
3/7/23	Linear regression	PML 11.1, 11.2,11.3 and ESL 3.2	Project 3 due Homework 3 due html pdf
3/9/23	Support Vector Machine		NB for Spam Classification Paper Reading Quiz 3 due html pdf
3/14/23	Midterm Review		Homework 4 due
3/16/23	**Midterm**	Midterm Jeopardy	**Location**: Kennedy Hall 116 **Time**: 7:30pm
3/21/23	Empirical Risk Minimization	PML 4.3, 5.4	html pdf
3/23/23	Bias and Variance Tradeoff		html pdf
3/28/23	Bias and Variance Tradeoff and Model Selection		Project 4 due html pdf
3/30/23	Kernels, part 1	PML: 17.1	html pdf
4/4/23	Spring Break	Woohooo!!	
4/6/23	Spring Break	Woohooo!!	
4/11/23	Kernels, part 2	PML: 17.3	html pdf slides Kernel Ridge Regression Demo Project 5 due
4/13/23	Classification and regression trees, part 1		Homework 5 due html html pdf

Date	Topic	References	Notes/assignments
4/18/23	Classification and regression trees, part 2		Project 6 due html html pdf Classification Tree Demo Regression Tree Demo
4/20/23	Ensemble Methods: Bagging & random forest		Homework 6 due html pdf
4/25/23	Ensemble Methods: Boosting		html pdf
4/27/23	Neural Network		pdf
5/2/23	Neural Network: backpropagation, convolution	PML: 14.1, 14.2, 14.3,15.4, 15.5	
5/4/23	Neural networks: Transformers	Transformer Algorithm Transformers explained Formal Algorithm	Project 8 due Homework 7 due Kaggle due Bias-Variance Tradeoff Paper Reading Quiz due pdf
5/9/23	AI in Human Society		pdf
5/14/23	Final Exam		**Location:** TBD **Time:** 2:00pm

References

While this course does not explicitly follow a specific textbook, there are several that are very useful references to supplement the course.

Books

We will not be explicitly following any single textbook in this course. Nevertheless, the books by Golub and Van Loan, and Trefethen and Bau collectively cover the material for the course and are recommended. Most suggested readings are assigned out of these two texts. Three additional texts are provided that complement these texts and are useful for further study (or to gain another perspective).

Probabilistic Machine Learning: An Introduction, by Murphy
We will provide section numbers to this text alongside many of the lectures (abbreviated as PML in the schedule). This text is available digitally through the Cornell University Library and a draft version is available directly from the author. [Book website]

The Elements of Statistical Learning by Hastie, Tibshirani, and Friedman
This text provides a comprehensive introduction to statistical learning and provides in-depth discussion of many of the topics in this course (abbreviated as ESL in the schedule). The book is available directly from the authors. [Book website]

Additional references

An Introduction to Statistical Learning by James, Witten, Hastie, and Tibshirani
This book provides a good overview of some methods in statistical learning, some of which we will discuss. The book is available online through the books website and via the Cornell Library. [Book website]

Patterns, Predictions, and Actions by Hardt and Recht
A very nice new book that covers many of the topics we do in this class (abbreviated as PPA in the schedule). The book is available directly from the authors. [Book website]

Fairness and Machine Learning by Barocas, Hardt, and Narayanan
While a work in progress, this text provides insight into fairness as a central tenet of machine learning. In particular, it highlights ethical challenges that arise in the practice of machine learning. The current version of this book is available directly from the authors. [Book website]

Background references

Linear Algebra by Khan Academy
Relive the basics of linear Algebra. Everybody loves Khan Academy. [Linear Algebra (Khan Academy)]

Linear algebra course by Strang
Portions of this course will utilize your knowledge of linear algebra. If you feel you need additional preparation, or would like to revisit the topic, you may find Gilbert Strangs linear algebra course quite useful. [MIT Open Courseware]

Matrix Methods in Data Analysis, Signal Processing, and Machine Learning by Strang
A subsequent course to the above by Strang covers some of the same topics we will (particularly for the linear algebra part of the course) and you may find the videos a useful additional resource. [MIT Open Courseware]

Software

Python
NumPy
PyTorch

Course policies

Inclusiveness

You should expect and demand to be treated by your classmates and the course staff with respect. You belong here, and we are here to help you learn and enjoy this course. If any incident occurs that challenges this commitment to a supportive and inclusive environment, please let the instructors know so that the issue can be addressed. We are personally committed to this, and subscribe to the Computer Science Department's Values of Inclusion. [Statement reproduced with permission from Dan Grossman.]

Mental health resources

Cornell University provides a comprehensive set of mental health resources and the student group Body Positive Cornell has put together a flyer outlined the resources available.

Participation

You are encouraged to actively participate in class. This can take the form of asking questions in class, responding to questions to the class, and actively asking/answering questions on the online discussion board.

Collaboration policy

Students are free to share code and ideas within their stated project/homework group for a given assignment, but should not discuss details about an assignment with individuals outside their group. The midterm and final exam are individual assignments and must be completed by yourself.

Academic integrity

The Cornell Code of Academic Integrity applies to this course.

Accommodations

In compliance with the Cornell University policy and equal access laws, we are available to discuss appropriate academic accommodations that may be required for student with disabilities. Requests for academic accommodations are to be made during the first three weeks of the semester, except for unusual circumstances, so arrangements can be made. Students are encouraged to register with Student Disability Services to verify their eligibility for appropriate accommodations.

COVID-19 considerations

While many aspects of this course are built with flexibility in mind, if situations arise that may require additional accommodations please reach out to the instructors to discuss potential arrangements.

Built using: Bootstrap