Cornell Bowers CIS
College of Computing
and Information Science

CS 3780*/CS 5780 Introduction to Machine Learning

Cornell Bowers CIS Department of Computer Science

Fall 2024 DRAFT Syllabus—Subject to Change

Course website: TBA

Please note: *This course was previously numbered CS 4780 for undergraduate students. The course content and difficulty remain unchanged. It will also continue to fulfill requirements for the CS major and the same categories for the IS and ISST majors that CS 4780 fulfilled.*

General Information

Faculty Name: Sarah Dean & Thorsten Joachims
Faculty Email: sdean@cornell.edu & tj36@cornell.edu
Faculty Office Hours: TBA by the start of the Fall 2024 semester.

Course Staff and Course Staff Office Hours: Please see the course staff calendar on Canvas for course staff information, including their office hours' schedule.

Credits and Credit Hour Options:

- **Credits:** 4.0, of which 1.0 credit is earned through independent programming projects.
- **Grading:** Student Option: Students may elect to take the course for a letter grade or S/U (Satisfactory/Unsatisfactory). Audit is not permitted.
- **Co-meets:** CS 3780 and CS 5780 co-meet. This is a combined syllabus for both courses; graduate and professional students must register for CS 5780. Differences between the two courses are outlined herein.

Time and Location: This course meets on Tuesdays and Thursdays from 1:25 - 2:40 p.m. for a total of 28 sessions. The classroom will be announced by the Office of the University Registrar and posted on the Fall 2024 Course Roster and in Student Center.

There will also be one evening prelim and a final exam. The dates and times for the evening prelim and the final will be scheduled and announced by the Office of the University Registrar; the final will be held during the university's final exam period.

Prerequisites/Corequisites: Probability theory (e.g., BTRY 3080, ECON 3130, MATH 4710, ENGRD 2700, CS 2800), linear algebra (e.g., MATH 2210, MATH 2940, MATH 2310), single-variable calculus (e.g. MATH 1910, MATH 1110), and programming proficiency (e.g., CS 2110).

Forbidden Overlaps: ECE 3200 (previously ECE 4200), ORIE 3741 (previously ORIE 4741), STSCI 3740 (previously STSCI 4740).

Enrollment Information and Questions: If you have questions about enrolling in this course, please review the Bowers CIS enrollment policies and waitlist information available on the Bowers CIS Courses Help website. If you can't find the answer to your question, please submit a ticket on that same page for assistance.

Course Description

Machine learning is concerned with the question of how to make computers learn from experience. The ability to learn is not only central to most aspects of intelligent behavior, but machine learning techniques have become key components of many software systems. For example, machine learning techniques are used to build search engines, to recommend movies, to understand natural language and images, and to build autonomous robots. This course will introduce the fundamental set of techniques and algorithms that constitute machine learning as of today. The course will not only discuss individual algorithms and methods, but also tie principles and approaches together from a theoretical perspective. In particular, the course will cover the following topics:

- Supervised Batch Learning: model, decision theoretic foundation, model selection, model assessment, empirical risk minimization
- Instance-based Learning: K-Nearest Neighbors, collaborative filtering
- Decision Trees: TDIDT, attribute selection, pruning and overfitting
- Linear Rules: Perceptron, logistic regression, linear regression, duality
- Support Vector Machines: Optimal hyperplane, margin, kernels, stability
- Deep Learning: multi-layer perceptrons, deep networks, stochastic gradient
- Generative Models: generative vs. discriminative, naive Bayes, linear discriminant analysis

Course Objectives

The goal of CS 3780/5780 is to introduce students to the fundamentals of Machine Learning. A student should gain knowledge the main concepts and principles that underlie most machine learning algorithms, such as discriminative and generative learning, parametric and non-parametric models, generalization, statistical inference, the basics behind neural networks, and kernel methods. The class is a prerequisite for subsequent classes on reinforcement learning, deep learning, computer vision, natural language processing.

Learning Outcomes

By the end of this course, students will be able to:

- Interpret the basic machine learning setup and principles.
- Develop the ability to implement basic machine learning models, apply them to real-world data sets, and evaluate their performance.
- Obtain an understanding of what assumptions different machine learning methods make and for which settings they apply
- Review the bias variance tradeoff and apply it to debug machine learning algorithms in practical settings by avoiding over- and underfitting.
- ***For CS 5780 only:*** Read and describe original research papers in the field.

Method of Assessing Student Achievement

Basis of Grade Determination

The final grades are based on homework assignments, programming projects, and a midterm (prelim) and final exam. *For the 5000-level version of the course, there are also quizzes that test the comprehension of original research papers, which factor into the final grade.*

Specifically, for CS 3780 the basis of grade determination is:

- 60% exams
- 40% independent programming assignments and homework

For CS 5780 the basis of grade determination is:

- 55% exams
- 35% independent programming assignments and homework
- 10% paper comprehension

Undergraduates enrolled in CS 3780 may choose to do the paper comprehension assignments; if completed they will receive the higher of their two grades from the above schemes.

Independent Programming Projects

More information about the independent programming projects, including topics, expectations, etc., will be shared in class and on the course website. There is an expectation that your project time will include additional contact with course staff and/or the instructor. In total, your hours on the project should be commensurate with 1.0 credit of out-of-class work (~40 hours).)

Grading Scale

This class adheres to a letter grading system, which is the official grading system at Cornell University. Grades will be assigned on the following scale:

A+	98-100%	4.3
A	93-97%	4.0

A-	90-92%	3.7
B+	88-89%	3.3
B	83-87%	3.0
B-	80-82%	2.7
C+	78-79%	2.3
C	73-77%	2.0
C-	70-72%	1.7
D	60-69%	1.3
F	Below 60%	0.0

Course Management

Inclusion

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.]

Participation

Students 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.

Academic Integrity

The Cornell Code of Academic Integrity applies to this course. Absolute integrity is expected of every Cornell student in all academic undertakings. Integrity entails a firm adherence to a set of values, and the values most essential to an academic community are grounded on the concept of honesty with respect to the intellectual efforts of oneself and others. Academic integrity is expected not only in formal coursework situations, but in all University relationships and interactions connected to the educational process, including the use of University resources. Consult the Cornell Code of Academic Integrity for further information: http://cuinfo.cornell.edu/aic.cfm.

Student Disability Accommodations

Students with Disabilities: Your access in this course is important to us. Please request your accommodation letter early in the semester, or as soon as you become registered with Student Disability Services (SDS), so that we have adequate time to arrange your approved academic accommodations.

- Once SDS approves your accommodation letter, it will be emailed to both you and me. Please follow up with the course staff to discuss the necessary logistics of your accommodations.
- If you experience any access barriers in this course, such as with printed content, graphics, online materials, or any communication barriers, reach out to the course staff or SDS.
- If you need immediate accommodation, please speak with me after class or send an email message to me and SDS at sds_cu@cornell.edu.

If you have, or think you may have a disability, please contact Student Disability Services for a confidential discussion: sds_cu@cornell.edu or visit the SDS website to learn more.

Mental Health and Wellbeing

Your health and wellbeing are important to us. There are services and resources at Cornell designed specifically to bolster undergraduate, graduate, and professional student mental health and well-being. Remember, your mental health and emotional well-being are just as important as your physical health. If you or a

friend are struggling emotionally or feeling stressed, fatigued, or burned out, there is a continuum of campus resources available to you. Help is also available any time day or night through Cornell's 24/7 phone consultation (607-255-5155). You can also reach out to me, your college student services office, your resident advisor (if applicable), or Cornell Health for support.

Important Note

Please note that this syllabus is subject to change. Any changes will be announced in class and/or on the course website.