

Machine Learning, Spring, 2020

Course Number: COMPSCI 589

Time: MW / 2:30-3:45 PM

Room: Thompson Hall room 102

Course Final: Wed May 6, 2020, 3:30PM - 5:30PM, Totman Phys. Ed. Bldg. Gym

Instructor: [Justin Domke](#)

Staff Email: Please use Piazza

Course Website: Detailed materials for the course will be hosted on Moodle. Syllabus (this page) is at

<http://people.cs.umass.edu/domke/courses/cs589/>

Instructor Office Hours: TBD

TA Hours: TBD

Course Description: This course will introduce core machine learning models and algorithms for classification, regression, clustering, and dimensionality reduction. On the theory side, the course will focus on understanding models and the relationships between them. On the applied side, the course will focus on effectively using machine learning methods to solve real-world problems with an emphasis on model selection, regularization, design of experiments, and presentation and interpretation of results.

Update: (Nov 7, 11am) Due to high demand, we are increasing the capacity of the course in section 589-01. These seats should be available on SPIRE soon. Any further changes will be announced here. **Second update:** (Nov 7 7:30pm) These seats are now almost all allocated.

Override questions:

- I constantly get questions about override requests for this course, so I apologize that I cannot answer your question individually.
- If you'd like to take this course but cannot register, please submit an override request through the [online system](#).
- Above all, please describe your background in **linear algebra**, **probability theory**, and **basic multivariate calculus**.
- Please list any courses you've taken either *in* those topics, or *using* those topics or any other relevant training or experience you might have.
- If you have taken a course elsewhere that you wish to substitute for a prerequisite, include a link to the syllabus of this course.
- Out of fairness, emailing me directly will not result in any preferential access. Note that it is likely there will be a substantial waitlist for the course, so I can't guarantee anyone entry to the course from the waitlist regardless of preparation.

Textbooks: This semester, the course will have **no mandatory textbook**. We will have optional readings from two open textbooks:

- [The Elements of Statistical Learning](#) by Hastie, Tibshirani and Friedman
- [An Introduction to Statistical Learning](#) by James, Witten, Hastie and Tibshirani

Grading Scheme:

| | |
|----------------------|-----|
| Homework | 50% |
| Final | 30% |
| Quizzes | 15% |
| Course Participation | 5% |

Homework: There will be five homework assignments.

Quizzes: There will be approximately 5-6 quizzes, each taking place in-class.

Preliminary Schedule

1. Introduction and Overview

(Unit 1: Regression)

2. Linear Regression, Ridge, and Lasso
3. KNN Regression, Regression Trees, and Feature Selection
4. Support Vector and Neural Network Regression
5. KOLS and Gaussian Process Regression

(Unit 2: Classification)

6. K-Nearest Neighbors and Decision Trees
7. Naive Bayes, LDA, and Logistic Regression
8. Overfitting, Regularization, and Cross-Validation
9. Support Vector Machines Basis Expansion, and Kernels
10. Neural Networks and Deep Learning
11. Ensembles and Classification

(Unit 3: Kernels)

12. Kernel Ridge Regression
13. Support Vector Machines

(Unit 4: Bayesian Methods)

14. Bayesian Methods 1
15. Bayesian Methods 2
16. Markov Chain Monte Carlo
17. Generative and Discriminative Methods

(Unit 4: Unsupervised Learning)

18. Hierarchical Clustering
19. K-Means
20. Mixture Models
21. Linear Dimensionality Reduction and SVD
22. Principal Components Analysis

(Final)

23. Final Review

What is the difference between CMPSCI 589 and CMPSCI 689?:

589 has been designed to focus on understanding and applying core machine learning models and algorithms, while 689 focuses on the mathematical foundations of machine learning. While both courses require a background in multivariate calculus, linear algebra, and probability; 689 is more theoretically focused and will use more of this background material than 589. In particular, 589 will not focus on deriving learning or optimization algorithms.

Should I take CMPSCI 589 or CMPSCI 689?: 589 is appropriate as an introductory machine learning course for senior undergraduate students, masters students, and MS/PhD students interested in applying machine learning in their research. Note that 589 can count for credit for MS/PhD students, but it does not satisfy an AI core requirement. Graduate students who intend to pursue research in machine learning or who need a course to satisfy the AI core requirement should take 689.

Required Background: While this course has an applied focus, it still requires appropriate mathematical background in probability and statistics, calculus and linear algebra. The official prerequisites for undergrads are CMPSCI 383 and MATH 235 (CMPSCI 240 provides sufficient background in probability and Math 131/132 provide sufficient background in calculus). Graduate students can check the descriptions for these courses to verify that they have sufficient mathematical background for 589. The course will also use Python as a programming language including the numpy, scipy, and scikit-learn. Some familiarity with Python will be helpful, but senior CS students should be able to learn Python during the course if needed. Graduate students from outside computer science with sufficient background are also welcome to take the course. The following references can provide a useful review:

- [Probability Theory](#)
- [Linear Algebra and Matrix Calculus](#)
- [Matrix Cookbook](#)
- [Matrix Identities](#)

Course Policies

- **Pass Fail / Audit.** This class has extreme demand for seats. Thus, I

will not approve any students to take the class pass/fail or as audit. Depending on what degree program you are in, you still may be able to register for the class in this way, and I will of course respect the university policy in that regard. However, I will not sign any forms approving a student to take the course pass/fail. Many students also ask to switch to pass/fail several weeks into the semester. I will not approve any such requests. *If you want to be able to switch to pass/fail, you should not register for this class.*

- **Quiz Solutions.** For logistical reasons, we are not able to provide solutions for quizzes. However, you can (and are encouraged to!) go over your quiz answer with a TA in office hours.
- **Homework Submission:** Homework assignments will generally consist of developing machine learning systems in Python, evaluating the systems, and producing written reports. Both the code and report must be submitted by the due date for a submission to be considered on time.
- **Late Homework:** To allow some flexibility to complete assignments given other constraints, you have a total of five free late days. You will be charged one late day for handing in an assignment within 24 hours after it is due, two late days for handing in an assignment within 48 hours after it is due, etc. Your assignment is considered late if either the written or code portions are submitted late. The late homework clock stops when both the written and code portions are submitted. After you have used up your late days, late homework will not count for credit except in special circumstances (ie: illness documented by a doctors note). If you do not hand in an assignment at all, this will count as using all five late days.
- **Homework Collaboration:** You are encouraged to *verbally discuss* assignments and course material with other students in person or on the course forums. However, you must show that you fully understand the solution to any homework problem arising from such collaboration by writing your own code, running your own experiments, and producing your own write-up for the problem. You should not at any time look at the solutions of another student. Showing your solutions to another student is also considered cheating.
- **Academic Honesty Policy:** You are required to list the names of anyone you discuss problems with on the first page of your solutions. This includes teaching assistants or instructors. Copying any solution materials from external sources (books, web pages, etc.) or other students is considered cheating. To emphasize: *no detectable* copying is acceptable, even, e.g., copying a single sentence from an outside source. Sharing your code or solutions with other students is also considered cheating. Any detected cheating will result in a grade of -100% on the assignment for all students involved (negative credit), and potentially a grade of F in the course.
- **Re-grading Policy:** Errors in grading of assignments and exams can occur despite the best efforts of the course staff. If you believe you've found a grading error, complete the online re- grade request form. Re-grade requests must be submitted no later than one week after the assignment is returned. Note that re-grading may result in your original grade increasing or decreasing as appropriate.

