



# ELEG/CISC/FSAN 817 Large Scale Machine Learning

Credits: 3

Semester: Fall, Year: 2023

Meeting Days, Times, Location and Room: MWF 4:10–5:05 pm

Office Hours: MW 5:15–5:45 pm, Evans Hall, Room 306

## Table of Contents

[1. Instructor Information](#)

[2. Course Description](#)

[3. Learning Outcomes](#)

[4. Learning Resources](#)

[5. Learning Assessment](#)

[6. Course Calendar](#)

[7. Course Policy Document](#)

- [Course-Specific Policies](#)
- [UD Standard Course Policies](#)

# 1. Instructor Information and Introduction

## Instructor Contact Information

Instructor name: Austin J. Brockmeier (preferred address: Dr. Brockmeier)  
Instructor preferred pronouns: he/his/him  
Email address: [ajbrock@udel.edu](mailto:ajbrock@udel.edu)  
Faculty Profile: <https://www.cis.udel.edu/people/faculty-profile/?id=264>  
Website: <https://www.eecis.udel.edu/~ajbrock/>  
Office location: 306 Evans Hall  
Office hours: MW 5:15–5:45

# 2. Course Description

## Description

Introduction to the analysis and processing of massive and/or high-dimensional data. Large-scale machine learning problems can involve growth in the number of data points, features, target variables, or related prediction tasks. Approaches to address these cases rely on concepts from optimization theory, statistics, and artificial neural networks. Computational and statistical scaling from both theoretical and practical perspectives are covered.

## Prerequisites

Students should have taken a previous graduate-level machine learning or statistical estimation course.

## Course Delivery

Course will meet for in person lectures, discussions, and assessments in person in a classroom setting.

# 3. Learning Outcomes

At the completion of this course an actively engaged student will be able to

1. frame data science problems by mathematically formulating machine learning tasks, with precisely stated objectives, assumptions, constraints, and the mathematical characteristics of input and output.
2. appropriately choose regularization methods, model selection criterion, and a valid experimental design (including hyperparameter selection) to ensure generalizability and reproducibility.
3. list the challenges, errors, and uncertainties inherent with large-scale data.
4. select and justify appropriate algorithms (based on relaxations, randomization, greedy approaches, and distributed approaches), data structures (sparse matrices, low-rank matrices and tensors, tree structured representations), and neural network architectures for large-scale problems such that computation can be successfully executed with an understanding of the trade-off between approximation and complexity in memory and space.
5. summarize and critique descriptions of machine learning methods, experimental design, result discussions (including statistical tests) in a peer review setting with constructive feedback.

## 4. Learning Resources

Class slideshows as PDFs will typically be posted to Canvas the same day as lecture either shortly before or shortly after each lecture.

PDFs and links to papers will be posted on Canvas.

A list of suggested supplementary books available from the UD Library will be updated below as the topics arise.

### Technology

Appropriate programming environment: Projects and homework will require access and familiarity to a computer programming environment suitable for data processing, statistics, numerical processing, and machine learning. Suggested languages (alphabetic order): Julia, MATLAB/ Octave, python, R, ...

**Canvas**: In this class, Canvas, UD's online learning management system, will be used for all course activities and communication channels. All assignments will be posted through the Canvas course site unless otherwise directed. Information on how to use Canvas is available through the [Canvas Student Guide](#). Canvas can also be accessed via [MyUD](#).

**Poll Everywhere**: We will use this online polling platform to collect real-time responses from everyone attending class. Access to Poll Everywhere is possible by using any computer or mobile device with internet access. Poll Everywhere is a UD licensed tool and there is no cost to you to use this tool. Please read over the [UD student guide](#) that will demonstrate how to log in and how to respond to polls. As a reminder, the [UD academic honesty policy](#) applies to the use of educational technology tools such as Poll Everywhere.

### Office Hours

Please ask questions that are relevant to the entire class during the lecture. Class will be best if the audience is interactive.

Please come by office hours to discuss topics that are specific to your project, or your learning outcome performance. (Also I am open to field questions about careers related to machine learning and data science based on my colleagues' experiences outside of academia.)

**Since office hour timing may not match your schedule, email is also great if you don't imagine the discussion will be much of a dialogue.**

**Email me to schedule a Zoom meeting if you cannot make office hours but want to discuss something that won't be relevant to the rest of the class.**

## 5. Learning Assessment

### Final Grade Breakdown

The final course grade will be calculated using the following components:

Course Component	Percentage of Total
<b>Participation</b>	5
<b>Assigned readings summary</b>	10
<b>Homework assignments</b>	20
<b>First in-class exam</b>	10
<b>Second in-class exam</b>	10
<b>Project submissions</b>	35
<b>Peer feedback on project submission</b>	10

## Participation

There will be points for participation in discussions on Canvas, interactive polls in class for reviews and quizzes, and course evaluation completion.

## Assigned Readings Summary

Weekly assigned readings require summaries that show you can comprehend, summarize, critique, extract insight, and catalog resources. A short description 4–8 sentences in your own words, a discussion of insights from reading the paper (2–5 sentences), a list of any resources (algorithms, data sets, experimental designs, proof techniques, statistical tests, theorems, visualizations) you would find useful for your own or future research, and a perspective on how it can be applied to your own research or project.

## Project Submissions

The project will be broken into a series of assessments due in sequence. These intermediate results are crucial for feedback. Allow time for each.

- Initial abstract (2%)
- Revised abstract (3%)
- Precise formulation and detailed experimental design (10%)
- Revised formulation and experimental design, preliminary results initial results, and plan for remainder (5%)
- Final report (15%)

**Plagiarism includes failing to cite sources or presenting others' work as your own. These have serious academic implications as described in the Academic Integrity portion of the Course Policy Document below.**

## Peer Feedback on Project Submissions

Peer feedback on three other students project-related submissions in the style of a conference peer review cycle.

- Peer feedback project abstract (2%)
- Peer feedback on formulation, experimental design, and results plan (3%)
- Peer feedback on final report (5%)

## Grading Scale

Students will be assigned the following letter grade based on the calculation coming from the course assessment section.

Grade	Interval
A	$[90, \infty)$
A-	$[85, 90)$
B+	$[80, 85)$
B	$[75, 80)$
B-	$[70, 75)$
C+	$[65, 70)$
C	$[60, 65)$
⋮	⋮

## 6. Course Calendar

### Important Dates

- 8/30 (W) First day of class.
- 9/4 (M) No class (Labor Day)
- 10/13 (F) First in-class midterm examination
- 10/27 (F) Project abstract due
- 11/3 (F) Revised project abstract due
- 11/10 (F) Peer feedback on project abstract due
- 11/17 (F) Project formulation and experimental design due
- 11/20–11/24 Classes suspended due to Thanksgiving Holiday and Native American Heritage Day
- 12/1 (F) Second in-class midterm examination; Revised project formulation, experimental design, preliminary results, and plan for remainder due
- 12/6 (W) Peer feedback on initial project formulation and experimental design due
- 12/11 (M) Last class
- 12/13 (W) Final project report due
- 12/18 (M) Peer feedback on final project report due; Deadline for any late work or extra credit

Date	Lecture	Theme/Topic	Readings/Assignments Due
8/30	1	Welcome and introduction, definition of machine learning	9/1 Participate in Introduction Discussion
9/1	2	The distinct spaces in a machine learning formulation, vector spaces, convexity	9/6 Participate in Discussion on Michael Jordan “Artificial Intelligence”
9/6	3	Norms in high dimensional spaces	Readings: “Distances in high-dimensional spaces”

9/8	4	Multivariate statistics, estimators, bias, variance, regularization, bootstrap, cross-validation, principal component analysis, Johnson Lindenstrauss	
9/11	5	Statistical machine learning, least squares formulation, covariance, ridge regression	
9/13	6	Machine learning and basic optimization	Readings: "Sparsity for feature selection" and "bootstrapping"
9/15	7	Linear programs, dual norms, Lagrangian dual formulations for constrained optimization	
9/18	6	Lasso	Homework 1 due
9/20	7	Consistency of Lasso, Bootstrap Lasso	Readings: "Big Data Bootstrap"
9/22	8	Shrinkage, hyper-parameter selection, CV	
9/25	9	Polytopes, atomic sets	
9/27	10	Frank-Wolfe algorithm, dual representation for linear models	Readings: "Frank-Wolfe Algorithm"
9/29	11	Greedy feature selection	
10/2	12	Decision trees, random forests	
10/4	13	Gradient boosting, additive losses, reweighted risk minimization	Readings: "Gradient boosting"
10/6	14	Clustering, k-means, multitask learning	
10/9	15	Dual form of linear models	Homework 2 due
10/11	16	Review for Exam 1	
10/13	—	In-class midterm examination (Exam 1)	
10/16	17	Exam review and introduce project	
10/18	18	The kernel trick and kernel methods	Readings: "Large scale (non)-linear models"
10/20	19	Support vector machines	
10/23	20	Kernel regression	Readings: "Nyström method"
10/25	21	Gaussian processes	
10/27	22	Scaling kernel methods, random Fourier features	Project abstract due
10/30	23	Bayesian optimization for hyper-parameter optimization	Readings: "Practical Bayesian Optimization"
11/1	24	Spectral clustering, Kernel PCA	
11/3	25	Statistical Tests; Two-sample Tests	Revised project abstract due

11/6	26	Divergence, introduction to generative adversarial networks (GANs), (sliced) Wasserstein distance, Fréchet distance, Sinkhorn divergence	Readings: "Generative Adversarial Networks"
11/8	27	GANs: the saddle point optimization problem	Peer feedback on project abstract
11/10	28	Learning representations; auto-encoders	Homework 3 due
11/13	29	Independence, dependence, and mutual information; distance correlation and centered kernel alignment	
11/15	30	Metric learning, non-linear dimensionality reduction for visualization (t-SNE)	Readings: "t-SNE"
11/17	31	Scaling softmax computation during learning; tree-structured softmax; Negative sampling for embedding optimization	Project formulation due
11/27	32	Introduction to contrastive learning, self-supervised learning, multimodal learning, data augmentation	Readings: "SimCLR"
11/29	33	Review for Exam 2	12/1
12/1	—	In-class final examination (Exam 2)	Revised project formulations, experimental results, preliminary results, and plan for remainder due
12/4	34	Guest lecture/emerging topics	
12/6	35	Guest lecture/emerging topics	Peer feedback on formulation and experimental design due;
12/8	36	Guest lecture/emerging topics	
12/11	37	Last class: Review and project report QA	
12/13	—	—	Final project report due
12/18	—	—	Peer feedback on final project; Any late work, resubmissions, etc.

## 7. Course Policy Document

### Course-Specific Policies

Efforts will be made to record the sessions. Nonetheless, please try to attend and participate in class discussion and ask questions.

There will be some options of resubmission for homework assignments.

Please turn in all project assignments, especially those with a peer-reviewing component, on time. Dates are specifically chosen to ensure regular feedback from the instructor. Likewise do not delay your peer feedback.

We will welcome the comments and questions of everyone and be respectful in our discussions. Everyone should be able to feel a sense of belonging.

Ethical considerations: Access to data and computational automation are powerful tools. Machine learning requires responsible use and ethical considerations.

When giving peer feedback consider the following list adopted from [NeurIPS 2021](#)

- Be thoughtful. This may be the author's first machine learning project and you don't want to crush their spirits.
- Be fair. Do not let personal feelings affect your review.
- Be useful. A good review is useful to all parties involved: authors, other reviewers and instructor.
- Be specific. Do not make vague statements in your review, as they are unfairly difficult for authors to address.
- Be timely. Please respect the deadlines. If you cannot complete your review on time, please let the instructor know as soon as possible.

### Attendance

Absences on religious holidays listed in university calendars are recognized as an excused absence. Nevertheless, students are urged to remind the instructor of their intention to be absent on a particular upcoming holiday. Absences on religious holidays not listed in university calendars, as well as absences due to athletic participation or other extracurricular activities in which students are official representatives of the university, shall be recognized as excused absences when the student informs the instructor in writing during the first two weeks of the semester of these planned absences for the semester.

### Communication

Content-related questions should be discussed in class or during office hours. Individual performance, learning and career development, or detailed individual project questions are best discussed in office hours. Please limit the use of email to times when these options are not suitable, as questions sent by email may not be answered before the next class, especially questions whose response could benefit the entire class.

### Student Mental Health & Wellbeing

In addition to impacting your overall well being, diminished mental health can interfere with optimal academic performance. If this course is causing or contributing significant mental or emotional stress, then please reach out to me directly. However, problems with other parts of your life can also contribute to decreased academic performance. UD's Center for Counseling & Student Development (CCSD) provides cost-free and confidential mental health services to help you manage personal challenges that threaten your emotional or academic well-being. Remember, getting help is a smart and courageous thing to do -- for yourself and for those who care about you.

- Contact me
  - If you are struggling with this class, please check-in during office hours or contact me by email at [ajbrock@udel.edu](mailto:ajbrock@udel.edu)



- Check-in with your advisor
  - If you are struggling in multiple classes, unsure whether you are making the most of your time at UD, or unsure what academic resources are available at UD.
- UD's Center for Counseling & Student Development and UD Helpline
  - CCSD is open and available remotely, and 24/7 mental health support remains available on the UD Helpline at 302-831-1001 for any student in need of someone to talk to. Visit [CCSD's website](#) for additional information and resources.
- UD's Crisis Text Line
  - Text "UDTEXT" or "STEVE" at 741741 connect with a professional who specializes in supporting students of color via a confidential text message.
- Division of Student Life
  - Explore the Student Life's [Wellbeing webpage](#) for a comprehensive listing of well-being resources, activities and services available to all students.

## UD Academic Policies

### Academic Integrity

Please familiarize yourself with UD policies regarding academic dishonesty. To falsify the results of one's research, to steal the words or ideas of another, to cheat on an assignment, to re-submit the same assignment for different classes, or to allow or assist another to commit these acts corrupts the educational process. Students are expected to do their own work and neither give nor receive unauthorized assistance. View the [university's academic integrity policies and procedures](#). Office of Student Conduct, 218 Hullahen Hall, (302) 831-2117. E-mail: [student-conduct@udel.edu](mailto:student-conduct@udel.edu)

### Harassment, Discrimination, and Sexual Misconduct

The University of Delaware works to promote an academic and work environment that is free from all forms of discrimination, including harassment and sexual misconduct. As a member of the community, your rights, resource and responsibilities are reflected in the Non-Discrimination, Sexual Misconduct, and Title IX policy. Please familiarize yourself with this policy at the [University's Office of Equity & Inclusion's website](#). You can report any concerns to the University's Office of Equity & Inclusion, at 305 Hullahen Hall, (302) 831-8063 or you can report anonymously through UD Police (302) 831-2222 or the [EthicsPoint Compliance Hotline](#).

- Read the [full policy](#)
- [File a report](#)

### Faculty Statement on Disclosures of Instances of Sexual Misconduct

If, at any time during this course, I happen to be made aware that a student may have been the victim of sexual misconduct (including sexual harassment, sexual violence, domestic/dating violence, or stalking), I am obligated by federal law to inform the university's Title IX Coordinator. The university needs to know information about such incidents to, not only offer resources, but to ensure a safe campus environment. The Title IX Coordinator will decide if the incident should be examined further. If such a situation is disclosed to me in class, in a paper assignment, or in office hours, I promise to protect your privacy--I will not disclose the incident to anyone but the Title IX Coordinator.

For more information on Sexual Misconduct policies, where to get help, and reporting information, please refer to [www.udel.edu/sexualmisconduct](http://www.udel.edu/sexualmisconduct). At UD, we provide 24/7/365 crisis assistance and victim advocacy and counseling. Contact 302-831-1001 to get in touch with a sexual offense support advocate, as well as confidential and anonymous counseling services for other concerns.

### Accommodations for Students with Disabilities

Any student who may need an accommodation based on a disability should contact the Office of Disability Support Services (DSS) office as soon as possible. For more information, please visit [Getting Registered at DSS](#). Contact DSS by

phone: 302-831-4643; fax: 302-831-3261; website: [www.udel.edu/dss](http://www.udel.edu/dss); email: [dsoffice@udel.edu](mailto:dsoffice@udel.edu); or visit 240 Academy Street, Alison Hall Suite 130 during business hours (8-5 M-F).

## Non-Discrimination

The University of Delaware does not discriminate against any person on the basis of race, color, national origin, sex, gender identity or expression, sexual orientation, genetic information, marital status, disability, religion, age, veteran status or any other characteristic protected by applicable law in its employment, educational programs and activities, admissions policies, and scholarship and loan programs as required by Title IX of the Educational Amendments of 1972, the Americans with Disabilities Act of 1990, Section 504 of the Rehabilitation Act of 1973, Title VII of the Civil Rights Act of 1964, and other applicable statutes and University policies. The University of Delaware also prohibits unlawful harassment including sexual harassment and sexual violence.

For inquiries or complaints related to non-discrimination policies, please contact: Office of Equity & Inclusion-  
[oei@udel.edu](mailto:oei@udel.edu), 305 Hullahen Hall Newark, DE 19716 (302) 831-8063

For complaints related to Section 504 of the Rehabilitation Act of 1973 and/or the Americans with Disabilities Act, please contact: Office of Disability Support Services, [dsoffice@udel.edu](mailto:dsoffice@udel.edu), Alison Hall, Suite 130, Newark, DE 19716 (302) 831-4643 OR contact the [U.S. Department of Education - Office for Civil Rights](#)