



Machine Learning and Optimization (MLOPT)

CSCI 6962/4140

4 credit hours (both CSCI6962 and CSCI4140)

Fall 2024, RPI

Lally 104, Mondays and Thursdays, 2pm–3:50pm ET

Instructor:

Name: Alex Gittens* (he/him)

*Either Professor Alex or Professor Gittens are acceptable forms of address.

Email: gittea at rpi.edu

Office: Lally 316

Phone: (518)-276-6476

Office Hours: Mondays 4pm–5pm ET and Thursday 10am–11am ET in my office

TA:

Name: Vijay Sadashivaiah (he/him)

Email: sadasv2 at rpi.edu

Office Hours: Tuesday 3pm–4pm via WebEx at <https://rensselaer.webex.com/meet/sadasv2>

Course website:

<http://www.cs.rpi.edu/~gittea/old-site/teaching/fall2024/mlandopt/mlandopt.html>

Course discussion site:

<https://piazza.com/rpi/fall2024/csci41406962/home>

Course Submittity:

<https://submittity.cs.rpi.edu/courses/f24/csci4140>

Prerequisites CSCI 2300 (Introduction to Algorithms) AND (CSCI 2210 (MFML - Mathematical Foundations of Machine Learning MFML) OR MATH 4100 (Linear Algebra)). Students may not receive credit for both the 4000 level and 6000 level versions of this course.

Course synopsis. Modern (i.e. large-scale, or “big data”) machine learning and data science typically proceeds by formulating the desired outcome as the solution to an optimization problem, then using suitable algorithms to solve these problems efficiently.

The first portion of this course introduces the probability and optimization background necessary to understand the randomized algorithms that dominate applications of ML and large-scale optimization, and surveys several popular randomized and deterministic optimization algorithms, placing the emphasis on those widely used in ML applications.

The second portion of the course introduces architectures used in modern machine learning because of the proven effectiveness of their inductive biases, and presents common regularization techniques used to mitigate the issues that arise in solving the nonlinear optimization problems ubiquitous in modern machine learning.

The homeworks involve hands-on applications and empirical characterizations of the behavior of these algorithms and model architectures. A project gives the students experience in critically reading the research literature and crafting articulate technical presentations.

Required Texts. There is no official text for the course, but supplementary reading material will be added to the course website as we go along. I strongly recommend the following free online resources if you find you have difficulty with the linear algebra or probability used in the course, and for learning the PyTorch framework we use for the deep learning assignments:

Linear Algebra:	Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares. Vandenberghe and Boyd.
Probability:	Introduction to Probability, Statistics, and Random Processes. Hossein Pishro-Nik. Jeff Erickson's notes on discrete probability. Jeff Erickson.
PyTorch/Deep Learning:	PyTorch official tutorial.

Shared Learning Outcomes (CSCI 6962 and CSCI 4140) Upon successful completion of this course, each student:

- (S1) will be able to apply key concepts from probability appropriately to the analysis of randomized optimization algorithms
- (S2) will be able to formulate standard supervised and unsupervised machine learning tasks as deterministic and stochastic optimization problems,
- (S3) will be able to apply key concepts relating to convex functions, convex sets, and optimality conditions for convex optimization,
- (S4) will be able to implement first-order and stochastic first-order solvers for convex optimization problems,
- (S5) will be able to implement second-order solvers for convex optimization problems,
- (S6) will be able to discuss the trade-offs inherent in using first- and second-order solvers for optimization problems arising in machine learning, including adaptive solvers,
- (S7) will be able to implement state-of-the-art classes of architectures used in deep learning in standard problem domains, including computer vision and NLP, and discuss the relevant inductive biases,
- (S8) will be able to apply machine learning techniques including deep learning methods to real-world problems and evaluate their performance.

Additional Learning Outcomes (CSCI 6962) Upon successful completion of this course, each student:

- (G1) will be able to evaluate and discuss properties relevant to the selection and performance of optimization algorithms, including strong convexity, smoothness, and non-convexity,
- (G2) will demonstrate the ability to read and critically evaluate the contemporary literature on large-scale machine learning by discussing the theoretical guarantees and empirical performance of novel methods, and effectively sharing their conclusions in oral and written form,
- (G3) will demonstrate the ability to theoretically analyze the performance of stochastic optimization algorithms used in large-scale machine learning.

Course Assessment Measures

(CSCI 4140) The assessment mechanisms consist of six homeworks (equally weighted), a project, and weekly participation exercises (equally weighted). The homeworks for CSCI 4140 will test the portion of the shared learning outcomes relevant to the material covered on that homework. The project will be conducted in groups of at most four other undergraduates, and will consist of students implementing and conducting fresh empirical evaluations of a standard modern ML method we did not have time to cover in class (e.g. normalizing flows as an alternative to GANs), and giving an in-class presentation.

(CSCI 6962) The assessment mechanisms consist of six homeworks (equally weighted), a project, and weekly participation exercises (equally weighted). The homeworks for CSCI 6962 consist of the homeworks for CSCI 4140 with additional questions intended to test the additional learning outcomes for CSCI 6962. The project will be conducted in groups of at most two graduate students, and will consist of students reading the ML research literature, implementing and conducting fresh empirical evaluations of a recent contribution to the ML research literature, implementing and evaluating a novel modification to the method, giving an in-class presentation explaining the method, their modification, and the results, and submitting a written report on their project formatted as an ICLR workshop submission.

There are no makeup participation scores or homeworks. Special circumstances will be handled case-by-case,

if the student presents an institute letter requesting it and if the instructor deems the request reasonable. Regrading is available upon request: start by contacting the TA, and have them escalate to the instructor if you are not satisfied with the resolution.

Course Calendar. Nota bene, the actual material covered may vary as I adjust the course pace to facilitate the achievement of the course learning objectives. Planned lectures may be merged or elided.

August 29	Logistics; supervised ML models
September 2	Labor Day, no class
September 3	Empirical risk minimization and probability theory
September 5	Probability continued: importance sampling and approximate probabilities
September 9	Conditional probability, Bayes optimal estimators, and regression functions
September 9	Homework 1 assigned; due September 23
September 12	Decomposition of the ERM risk and the opportunity for approximate optimization
September 16	Regularized ERM and practical considerations: train/test/validation splits, cross-validation, hyperparameter optimization, regularization
September 19	Convex sets and convex/concave functions, consequences for optimization, first and second-order conditions, Jensen's inequality
September 23	Examples of convex sets, functions, convex optimization problems
September 23	Homework 2 assigned; due October 7
September 26	Optimality conditions for smooth and nonsmooth, constrained and non-constrained convex optimization; gradient descent
September 30	Subdifferentials and subgradients; rules for subdifferential calculations, and examples
October 3	The importance of curvature, backtracking line search, Method of steepest descent, Newton's method
October 7	Augmented Lagrangian method and Alternating Direction Method of Multipliers
October 7	Homework 3 assigned; due October 21
October 10	Probability meets optimization: (projected) stochastic gradient descent, convergence rates
October 14	Indigenous People's Day, no class
October 17	Adaptive gradient descent methods
October 21	Nonparametric ML, kernel methods
October 24	Approximate kernel methods
October 24	Homework 4 assigned; due November 8
October 28	Neural networks, inductive biases
October 31	Backpropagation/the chain rule
November 4	Autoencoders, regularizations
November 7	Convolutional neural networks
November 7	Homework 5 assigned; due November 22
November 11	Practical NN concerns: overfitting, vanishing/exploding gradients, combined algorithm and hyperparameter selection. Dropout, Batch and Layer Normalization
November 14	Data augmentation, skip connects and residual blocks, various architectures
November 18	NN architectures for sequential data: recurrent neural networks, (truncated) backpropagation through time, tokenization
November 21	Long short term memories, teacher forcing
November 21	Homework 6 assigned; due December 6
November 25	Bidirectional and deep RNNs; sequence-to-sequence models; encoder-decoder architectures
November 28	Thanksgiving, no class
December 2	Attention in RNNs, transformer architectures
December 5	Project presentations
December 9	Project presentations

Grading Criteria

CSCI 6962:

Homeworks	Project	Weekly Participation
50%	45%	5%

Threshold	90%	85%	80%	75%	70%	<70%
Grade	A	B+	B	C+	C	F

CSCI 4140:

Homeworks	Project	Weekly Participation
50%	35%	15%

Threshold	90%	85%	80%	75%	70%	65%	60%	<60%
Grade	A	B+	B	C+	C	D+	D	F

Other than the weekly participation assignments which will be used to elicit feedback from students, the assignments will target the learning objectives as follows.

	6962 and 4140	Additional for 6962
Homework 1	(S1)(S2)	
Homework 2	(S1)(S2)(S3)(S8)	(G1)
Homework 3	(S3)(S4)(S5)(S6)(S8)	(G1)(G3)
Homework 4	(S4)(S6)(S7)(S8)	(G2)(G3)
Homework 5	(S7)(S8)	(G2)(G3)
Homework 6	(S7)(S8)	(G2)(G3)
Project	(S7)(S8)	(G2)(G3)

The grading rubric and deadlines for the project are as follows.

Task	Due date	% of grade
Project selection	Lecture 8	10
Progress report	Lecture 20	35
Deliverables	Lecture 27	20
Presentations	Lecture 27 & 28	35

Students will be able to assess their performance by monitoring their grades which will be posted to Submitty as each assignment is graded. Additionally, CSCI 4140 students will receive two EWS updates during the semester.

Academic Integrity. Student-teacher relationships are built on trust. For example, students must trust that teachers have made appropriate decisions about the structure and content of the courses they teach, and teachers must trust that the assignments that students turn in are their own. Acts that violate this trust undermine the educational process. The Rensselaer Handbook of Student Rights and Responsibilities and The Graduate Student Supplement define various forms of Academic Dishonesty and you should make yourself familiar with these. In this course, all assignments that are turned in for a grade must represent the student's own work. In cases where help was received, or teamwork was allowed, a notation on the assignment should indicate your collaboration.

Plagiarism or collaborative writing of code, program outputs, or written answers will be considered cheating on any assignments other than those relating to the group project. In cases of academic dishonesty, the minimum penalty is a course grade of F. Violations of academic integrity may also be reported to the appropriate Dean (Dean of Students for undergraduate students or the Dean of Graduate Education for graduate students, respectively).

The use of artificial intelligence (AI) tools– including models using Transformer or diffusion architectures and foundation models of any type–to generate complete or partial solutions for assignments is prohibited unless instructor permission is obtained in advance. Submission of work generated by AI without (i) prior

instructor permission, and (ii) citation of the tool used and how it was used to complete the assignment will be considered a violation of academic ethics.

If you have any question concerning this policy before submitting an assignment, please ask for clarification. In addition, you can visit the following site for more information on our Academic Integrity Policy: Student Rights, Responsibilities, and Judicial Affairs.

Academic Accommodations. Rensselaer Polytechnic Institute is committed to providing equal access to our educational programs and services for students with disabilities. If you anticipate or experience academic barriers due to a disability, please contact the Office of Disability Services for Students (DSS) (dss@rpi.edu; 518-276-8197) to establish reasonable accommodations. Once you have been approved for accommodations, please provide your Faculty Memorandum (a letter provided to students by DSS) to the instructor of this course. Please provide this at the very beginning of the semester.