Machine Learning and Optimization (MLOPT)
CSCI 6962/4962

4 credit hours (both CSCI6962 and CSCI4962)
Fall 2023, RPI
JEC 5119, Tuesdays and Fridays, 10am–11:50am ET

Instructor:
Name: Alex Gittens* (he/him)
*Either Prof./Dr. Alex or Prof./Dr. Gittens are acceptable forms of address.
Email: `gittea` at rpi.edu
Office: Lally 316
Phone: (518)-276-6476

Office Hours: T 8:30am–9:30am ET and F 12pm–1pm ET in my office or by appointment

TA:
Name: Dong Hu (he/him)
Email: `hud3` at rpi.edu
Office Hours: M/Th 8:30-9:30am ET in Lally 09 (basement level)

Course website:
http://www.cs.rpi.edu/∼gittea/teaching/fall2023/mlandopt.html
Course discussion site:
https://piazza.com/rpi/fall2023/csci49626962/home
Course Submittys:
https://submitty.cs.rpi.edu/courses/f23/csci4962, and
https://submitty.cs.rpi.edu/courses/f23/csci6962

Prerequisites CSCI 2300 (Introduction to Algorithms), MATH 2010 (Multivariable Calculus and Matrix Algebra). Students may not receive credit for both the 4000 level and 6000 level versions of this course.

Course synopsis.

As a second course in machine learning, this course focuses on the optimization algorithms used in machine learning and the neural architectures used in modern deep learning.

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 4962) Upon successful completion of this course, each student:

(S1) will be able to formulate standard supervised and unsupervised machine learning tasks as deterministic and stochastic optimization problems,

(S2) will be able to apply key concepts relating to convex functions, convex sets, and optimality conditions for convex optimization,

(S3) will be able to implement first-order and stochastic first-order solvers for convex optimization problems,

(S4) will be able to implement second-order and quasi-Newton solvers for convex optimization problems,

(S5) 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,

(S6) 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,

(S7) 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 4962) The assessment mechanisms consist of six homeworks (equally weighted), a project, and weekly participation exercises (equally weighted). The homeworks for CSCI 4962 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 4962 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 elided or skipped.

Lecture 1	Logistics; supervised ML models
Lecture 2	Empirical risk minimization and probability theory
Lecture 3	Probability continued: importance sampling and approximate probabilities
Lecture 4	Conditional probability, Bayes optimal estimators, and regression functions
Lecture 4	Homework 1 assigned; due Lecture 8
Lecture 5	Decomposition of the ERM risk and the opportunity for approximate optimization
Lecture 6	Regularized ERM and practical considerations: train/test/validation splits, cross-validation, hyperparameter optimization, regularization
Lecture 7	Convex sets and convex/concave functions, consequences for optimization, first and second-order conditions, Jensen's inequality
Lecture 8	Examples of convex sets, functions, convex optimization problems
Lecture 8	Homework 2 assigned; due Lecture 12
Lecture 9	Optimality conditions for smooth and nonsmooth, constrained and non-constrained convex optimization; gradient descent
Lecture 10	Subdifferentials and subgradients; rules for subdifferential calculations, and examples
Lecture 11	The importance of curvature, backtracking line search, Method of steepest descent, Newton's method
Lecture 12	Augmented Lagrangian method and Alternating Direction Method of Multipliers
Lecture 12	Homework 3 assigned; due Lecture 16
Lecture 13	Probability meets optimization: (projected) stochastic gradient descent, convergence rates
Lecture 14	Adaptive gradient descent methods
Lecture 15	Nonparametric ML, kernel methods
Lecture 16	Approximate kernel methods
Lecture 16	Homework 4 assigned; due Lecture 20
Lecture 17	Neural networks, inductive biases
Lecture 18	Backpropagation/the chain rule
Lecture 19	Autoencoders, regularizations
Lecture 20	Convolutional neural networks
Lecture 20	Homework 5 assigned; due Lecture 24
Lecture 21	Practical NN concerns: overfitting, vanishing/exploding gradients, combined algorithm and hyperparameter selection. Dropout, Batch and Layer Normalization
Lecture 22	Data augmentation, skip connects and residual blocks, various architectures
Lecture 23	NN architectures for sequential data: recurrent neural networks, (truncated) backpropogation through time, tokenization
Lecture 24	Long short term memories, teacher forcing
Lecture 24	Homework 6 assigned; due Lecture 28
Lecture 25	Bidirectional and deep RNNs; sequence-to-sequence models; encoder-decoder architectures
Lecture 26	Attention in RNNs, transformer architectures
Lecture 27	Project presentations, first day
Lecture 28	Project presentations, second day

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 4962:

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 4962	**Additional for 6962**
Homework 1	(S1)	
Homework 2	(S1)(S2)(S7)	(G1)
Homework 3	(S2)(S3)(S4)(S5)(S7)	(G1)(G3)
Homework 4	(S3)(S5)(S6)(S7)	(G2)(G3)
Homework 5	(S6)(S7)	(G2)(G3)
Homework 6	(S6)(S7)	(G2)(G3)
Project	(S6)(S7)	(G2)(G3)

The grading rubric and deadlines for the project are as follows (again, the timings are approximate).

Task	**Due date**	**% of grade**
Project selection	Lecture 8	10
Progress report	Lecture 20	35
Deliverables	Lecture 26	20
Presentations	Lecture 26	35

The project selections and progress reports will be submitted in Submitty. The deliverables will be posted to a public Github repo, and a 20 minute recorded presentation will be posted to the RPI Box.

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 4962 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).

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.