Machine Learning and Optimization (MLOPT)
CSCI 6961/4961

4 credit hours (both CSCI6961 and CSCI4961)
Fall 2021, RPI
Sage 3101 (in person), Mondays and Thursdays, 10am–11:50am ET

Instructor:
Name: Alex Gittens*
*Either Professor Alex or Professor Gittens are acceptable forms of address, and my pronouns are he/him.
Email: `gittea` at rpi.edu
Office: Lally 316
Phone: (518)-276-6476
Office Hours: Mon, Thur 12pm–1pm ET via Campuswire or by appointment

TA:
Name: Ian Bogle (he/him)
Email: `boglei` at rpi.edu
Office Hours: TWed 1-2pm ET via Campuswire

Course website:
http://www.cs.rpi.edu/~gittea/teaching/fall2021/mlandopt.html
Course discussion site:
https://campuswire.com/c/GE81BDE9B/feed
Course Submittys:
CSCI 6961: https://submitty.cs.rpi.edu/courses/f21/csci6961
CSCI 4961: https://submitty.cs.rpi.edu/courses/f21/csci4961

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. 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 6961 and CSCI 4961) Upon successful completion of this course, each student:

(S1) demonstrate competence with probability theory/statistics needed to formulate and solve machine learning problems
(S2) can formulate standard supervised and unsupervised machine learning tasks as optimization problems,
(S3) can recall key definitions relating to convex functions, convex sets, and convex optimization
(S4) can implement first-order and stochastic first-order solvers for convex optimization problems,
(S5) can implement Newton's method and L-BFGS solvers for convex optimization problems,
(S6) can identify the trade-offs inherent in using first-order vs second-order solvers for optimization problems arising in machine learning,
(S7) is familiar with the state-of-the-art classes of architectures used in deep learning, and their mapping onto standard problem domains, including computer vision and NLP
(S8) has experience in applying machine learning techniques, including deep learning methods, to real-world problems

Additional Learning Outcomes (CSCI 6961) Upon successful completion of this course, each student:

(G1) can identify convex optimization problems, their optimality conditions, and properties relevant to the selection and performance of optimization algorithms such as strong convexity and smoothness.
(G2) can read the contemporary literature on randomized approaches to problems in large-scale machine learning, evaluate the theoretical guarantees and empirical performance of novel methods, and effectively share their conclusions.
(G3) can understand arguments guaranteeing the performance of stochastic optimization algorithms used in large-scale machine learning.

Course Assessment Measures

The assessment mechanisms consist of six homeworks, a project, and weekly participation exercises. The homeworks for CSCI 4961 will test the portion of the shared learning outcomes relevant to the material covered on that homework. The homeworks for CSCI 6961 consist of the homeworks for CSCI 4961 with additional questions intended to test the additional learning outcomes for CSCI 6961, as appropriate for the material covered in that homework.

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 30	Logistics; supervised ML models
September 2	Empirical risk minimization and probability theory
September 7	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 13	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 20	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 27	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 4	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 14	Probability meets optimization: (projected) stochastic gradient descent, convergence rates
October 18	Adaptive gradient descent methods
October 21	Nonparametric ML, kernel methods
October 25	Approximate kernel methods
October 25	Homework 4 assigned; due November 8
October 28	Neural networks, inductive biases
November 1	Backpropagation/the chain rule
November 4	Autoencoders, regularizations
November 8	Convolutional neural networks
November 8	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 15	Data augmentation, skip connects and residual blocks, various architectures
November 18	NN architectures for sequential data: recurrent neural networks, (truncated) backpropogation through time, tokenization
November 22	Long short term memories, teacher forcing
November 22	Homework 6 assigned; due December 6
November 29	Bidirectional and deep RNNs; sequence-to-sequence models; encoder-decoder architectures
December 2	Attention in RNNs, transformer architectures
December 6	Project presentations, first day
December 9	Project presentations, second day

Grading Criteria

CSCI 6961:

Homeworks	Project	Weekly Participation
50%	**35%**	**15%**

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

CSCI 4961:

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.

	6961 and 4961	**Additional for 6961**
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)

Projects will be completed in assigned groups of at most five students. The grading rubric and deadlines for the project are as follows.

Task	**Due date**	**% of grade**
Project selection	October 4	10
Progress report	November 4	35
Deliverables	December 2	20
Presentations	December 6 & 9	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 4961 students will receive two EWS updates during the semester.

Collaboration and Academic Honesty. 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.

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.