

Optimization for Machine Learning

ENG EC 525 Fall 2023

Instructor: Ashok Cutkosky email: cutkosky@bu.edu office: PHO 420

Class meets Mondays and Wednesdays 12:20-2:05 in PSY B37

Course website: <https://optmlclass.github.io/>

Piazza signup link: <https://piazza.com/bu/fall2023/ec525>

Gradescope course ID: 607283 (<https://www.gradescope.com/courses/607283>) course code: EJ7DRE

Office Hours

Monday 4-5pm and Wednesday 5-6pm, or by appointment. In person PHO420 or virtually by request.

Course description:

Efficient algorithms to train large models on large datasets have been critical to the recent successes in machine learning and deep learning. This course will introduce students to both the theoretical principles behind such algorithms as well as practical implementation considerations. Topics include convergence properties of first-order optimization techniques such as stochastic gradient descent, adaptive learning rate schemes, and momentum. Particular focus will be given to the stochastic optimization problems with non-convex loss surfaces typically present in modern deep learning problems. Additional topics may include second-order methods and variance reduction. After completing this course, students should be able to read, understand, and implement optimization algorithms from current research.

Prerequisites:

Ability to program in Python. Some experience with linear algebra, calculus, and probability. Example concepts that should be familiar include gradients, eigenvectors, eigenvalues, Taylor series, and expectations.

Coursework:

There will be regular homework (approximately weekly) assignments, which may include both mathematical problems and programming exercises, as well as a take-home final exam.

- All homework solutions should be produced via some word-processing program such as latex. Latex is *strongly encouraged* in place of other methods such as microsoft word.
- <https://www.overleaf.com> is a good resource for producing documents with latex, and also has many helpful tutorials. The latex source for the homework assignments will be provided.
- <https://detexify.kirelabs.org/classify.html> is a good resource for finding how to make various symbols in latex.

Homework will be posted on Blackboard.
Homework submission will use gradescope.

Grading

Grades will be 90% based on homework, 10% based on final exam.

Collaboration Policy:

You may collaborate with anyone else taking or auditing the class on the homeworks. However, you must write up or code your answers on your own.

On the exam you may NOT collaborate with anyone.

Please do not try to look up homework problems on the internet, or ask questions on forums.
This is not allowed.

Late Work Policy

Homework is due at 11:59pm on the date listed in the assignment. Late work is not accepted, but we will drop the two lowest assignment scores for the semester.

List of Potential Topics:

- Stochastic Gradient Descent
- Momentum-based optimization, and accelerated gradient descent.
- Adaptive gradient methods, including AdaGrad and Adam.
- Normalized stochastic gradient descent, LARS and LAMB.
- Large batch size optimization
- Stochastic preconditioning.
- Memory-efficient optimization
- Learning rate scheduling.
- Hyperparameter tuning.
- Second-order optimization and hessian-vector products
- Variance reduction

Textbook:

There is no textbook for this course, although detailed course notes and links to relevant research papers will be published on the class website.

Example papers and algorithms that will be discussed include:

[AdaGrad](#)

[Adam](#)

[AMSGrad](#)

[LAMB](#)