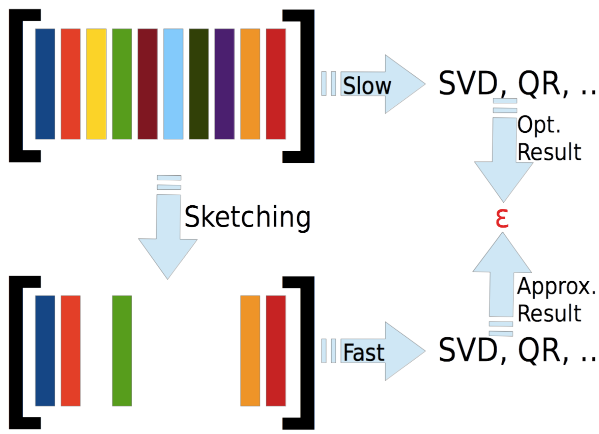
Large Scale Matrix Computation and Machine Learning

CSCI 6971/4971

Mondays and Thursdays, 2pm-3:50pm, Lally 02

Alex Gittens
gittea@rpi.edu
Lally 316
Office Hours: Mon, Thurs 4pm-5pm or by appointment
(518) 276-6476



Course Description: Modern machine learning routinely deals with millions of points in high-dimensional spaces. Classical numerical linear algebra algorithms can be prohibitively costly in such applications, as they aim at machine precision and scale super-linearly in the size of the input data. Fortunately, machine precision is seldom necessary, and recent developments have established that randomization can be used to bring the costs of matrix algorithms closer to linear in the size of the input data; this is done by sacrificing, in a principled manner, computational accuracy for increased speed. This course surveys modern randomized numerical linear algebra and its applications to machine learning, with the goal of providing a solid foundation for the development of new methods. Topics will include fast linear regression, fast approximate low-rank decompositions, fast k-means and spectral clustering, fast approximate kernel methods, and promising research directions in the field.

Course Website: `http://www.cs.rpi.edu/~gittea/teaching/spring2018/csci6971-and-4971.html` Note that this document supercedes any statements made on the website.
Course Text(s): None required. Lectures will be self-contained.

Course Objectives: At the end of the semester, students will understand how randomization allows us to efficiently tackle large numerical linear algebra and machine learning problems. To that end, they will be well-versed in algorithms for approximate matrix multiplication, randomized direct and iterative linear solvers, and large-scale kernel methods based on low-rank matrix approximations. They will be familiar with the analytical tools used in guaranteeing the performance of these algorithms, and be able to determine when randomized matrix algorithms are appropriate.

Grading Criteria:

Homeworks	50%
In-class Pop Quizzes	20%
Project	30%

Students are expected to have writing supplies on hand in each class to complete the in-class pop quizzes. If you are an athlete or for some other reason are not able to attend each class, make alternative arrangements with the instructor in the first two weeks of the course.

Letter grades will be computed from the semester average. Maximum lower bound cutoffs for A, B, C and D grades are 90%, 80%, 70%, and 60%, respectively. These bounds may be moved lower at the instructors discretion.

Homework Policy

All assignments must be typeset and are due, via email, at the start of class (the first 15 minutes). Late homeworks will be penalized and accepted at the discretion of the instructor. There will be no makeup homeworks.

Project Policy

A presentation project will be assigned during the second half of the semester. It will entail reading a paper, summarizing the theoretical results, implementing the main algorithm and appropriately empirically evaluating it, then delivering a 15 minute presentation before the class.

Academic Integrity

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 students own work. In cases where help was received, or teamwork was allowed, a notation on the assignment should indicate your collaboration.

Submission of any assignment that is in violation of this policy will result in an administrative grade of F for the course.

If you have any question concerning this policy before submitting an assignment, please ask for clarification.