

COMP 441: Large Scale Machine Learning

- **Instructor:** [Anshumali Shrivastava](#) (anshumali AT rice.edu)
- **Class Timings:** Mon/Wed/Fri 11am-11:50pm
- **Location:** AEB B209
- **Office Hours:** Monday and Wednesday 11:50am - 12:30pm, Duncan Hall 3118

Overview

Learning from large dataset is becoming a ubiquitous phenomena in all applications spanning robotics, medical decisions, internet, communication, biology, etc. A semester long project based course designed to give senior UG students a thorough grounding in the theory and algorithms needed for research and practical applications in machine learning for modern massive datasets. Topics draw from machine learning, classical statistics, algorithms and information theory.

Grading

- Project: 50%
- Assignments: 25%
- Quizzes: 15 %
- Lecture Scribing: 10%
- Participation and Discussion Forums: 5% (Extra)

Prerequisite

Familiarity with basics in linear algebra, probability, and machine learning is required.

Topics

Tentative List

Schedule

- 01/11 : Introduction, Logistics. ([slides](#))([scribe](#))
- 01/13 : Tail Inequalities and Reservoir Sampling. ([scribe](#))
- 01/15 : MapReduce and GPUs in 50 mins. ([slides](#))
- 01/20 : Bloom Filters.
- 01/22 : Count-Min Sketch.
- 01/25 : ([Assignment 1](#)) Due 12:00pm 8th Feb
- 01/25 : Johnson-Lindenstrauss Lemma.
- 01/27 : Locality Sensitive Hashing. ([slides](#))
- 01/29 : Randomized SVD.
- 02/01 : Convexity.
- 02/03 : Gradient Descent and Convergence.

- 02/05 : Variants of Gradient Descent.
- 02/08 : Linear Regression.
- 02/10 : ([Assignment 2](#)) Due 12:00pm 24th Feb
- 02/10 : MLE and Logistic Regression.
- 02/12 : SVMs.
- 02/15 : Constraint Optimization and Lagrangian Duality.
- 02/17 : Kernels.
- 02/19 : Kernel Features.
- 02/22 : More on SVMs and Kernels.
- 02/26 : PCA and Kernel PCA.
- 03/07 : POS Tagging and LogLinear Models.
- 03/09 : POS Tagging and Generative Models.
- 03/11 : Generative Model and Viterbi Algorithm.
- 03/14 : MidTerm Project Presentation.
- 03/16 : ([Assignment 3](#)) Due 12:00pm 30th March
- 03/16 : Structural SVMs.
- 03/18 : Quiz 1 and Structural SVMs contd.
- 03/21 : Recommender Systems Basic Neighborhood Models
- 03/23 : Recommender Systems Factorization Models
- 03/25 : Boosting (Ada-boost)
- 03/28 : Boosting Analysis and Statistical View
- 04/04 : ICA by Ryan Spring
- 04/06 : Decision Trees by Beidi Chen
- 04/08 : Online Learning and Weighted Majority Algorithm
- 04/11 : Multi-arm Bandits and UCB Algorithm
- 04/13 : ([Assignment 4](#)) Due 12:00pm 1st May
- 04/13 : Active Learning
- 04/15 : Machine Learning with Graphs
- 04/20 : Quiz 2 and Machine Learning with Graphs continued
- 04/22 : Final Project Presentations

Students with Disability

If you have a documented disability that may affect academic performance, you should: 1) make sure this documentation is on file with Disability Support Services (Allen Center, Room 111 / adarice@rice.edu / x5841) to determine the accommodations you need; and 2) meet with me to discuss your accommodation needs.