COMP 441: Large Scale Machine Learning

1 Instructor: Anshumali Shrivastava

2 Hours: Mon/Wed/Fri 11am - 11:50am

3 Description

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.

4 Prerequisite

Familiarity with basics in linear algebra, probability is required. Having done an Optimization/Big-Data course will be plus. Please take permission of the instructor if you don't have enough machine learning background.

5 Grading

Project: 50%
Assignments: 20%
In Class-Quizzes and Mid-Term Exam: 10 %
Lecture Scribing: 5%
Participation and Discussion Forums: 5%

6 Topics

Please check `http://www.cs.rice.edu/~as143/COMP441_Fall16/index.html` for more details.

Tentative Schedule:

Week 1: Fundamentals

- Logistics
- Probability Fundamentals
- System Fundamentals

Week 2. Sketching, Hashing and Data Streams

- Counting on Streams (Sketches and Bloom Filters).
- Locality Sensitive Hashing.

Week 3. Randomized Linear Algebra

- Matrix Sketching.
- Advanced Topics.

Week 4. Optimization for Machine Learning

- Convexity and Duality.
- Gradient Descent, Newtons Method, Projected Gradient Descent.
- Accelerated Gradient Descent and Distributed Optimization.

Week 5. Learning Models

- SVMs, Neural Network, Decision Trees.
- Bagging and Boosting (Ensembles).
- Advanced Topic.

Week 6. Kernels and Structured Prediction

- Kernel Trick and Kernel Learning, Hashing and Random Kitchen Sinks.
- Structured Predictions.

- Advanced Topic.

Week 7. Recommender Systems

- Matrix Factorization and Latent Space Modeling.
- Neighborhood Models and Ordinal Regression.
- Advanced Topic.

Week 8. Graphical Models and Variational Inference

- Graphical Models, Message Passing and EM.
- Gibbs Sampling and Variational Inference.
- Advanced Topic.

Week 9. Mining Massive Graphs

- Page Rank.
- Graph Kernels and Learning with Graphs.
- Advanced Topic.

Week 10. Large Scale Deep Learning

- Deep Architectures, CNNs.
- Large Scale Distributed Learning of Deep Networks.
- Advanced Topic.

Week 11. Online Learning and Multi-Arm Bandits

- Online Learning and Sequential Decision Making.
- Multi-arm Bandits and their variations.
- UCB and Thompson Sampling.

Week 12. Advanced Applications

- ML for Internet.
- ML for NLP.
- ML for Vision.