DS 222 Aug 3:1

Machine Learning with Large Datasets

Instructor

Partha P Talukdar
Email: ppt@iisc.ac.in

Teaching Assistant

Email:

Department: CDS

Course Time: TTh 2-3:30pm

Lecture venue: CDS 102

Detailed Course Page: https://sites.google.com/site/2017ds222/

Announcements

Brief description of the course

This course intends to teach students about scalable machine learning techniques. Students are expected to be

familiar with machine learning.

Prerequisites

Prior exposure to machine learning.

Syllabus

Streaming algorithms and Naive Bayes, fast nearest neighbor, parallel perceptrons, parallel SVM, randomized

algorithms, hashing, sketching, scalable SGD, parameter servers, graph-based semi-supervised learning,

scalable link analysis, large-scale matrix factorization, speeding up topic modeling, big learning and data

platforms, learning with GPUs.

Course outcomes

This course teaches students to develop scalable machine learning techniques, both in standalone and in

distributed settings. Students learn about design considerations in this area, available tools and algorithms, and

also about open problems. Students also get to learn about developments in the industry through various guest

lectures from the industry.

Grading policy

30% assignments, 10% midterm, 60% project

Assignments

Resources

* Mining of Massive Dataset. Jure Leskovec, Anand Rajaraman, Jeff Ullman

* Scaling up Machine Learning: Parallel and Distributed Approaches. Ron Bekkerman, Mikhail Bilenko, John Langford

* Foundations of Data Science. Avrim Blum, John Hopcroft, Ravi Kannan

* Research literature