

CS550: Massive Data Mining and Learning

Spring 2018

Class time: M/Th, 12:00pm – 1:20pm

Location: Allison Road Classrooms (ARC) 107

Instructor:

Yongfeng Zhang, PhD

Assistant Professor

Department of Computer Science

Rutgers, The State University of New Jersey

E-mail: yongfeng.zhang@rutgers.edu

Office Hours: Fridays, 10:00 – 12:00 or by appointment



TA: Sepehr Janghorbani

Email: sj620@scarletmail.rutgers.edu

Course Website: <http://yongfeng.me/teaching/s2018>

I. Course Description:

This class is a foundational class for the newly created Professional Master in Data Science within the Computer Science Department. Students will learn algorithms to store, process, mine, analyze, and synthesize streaming data, or data at rest that does not fit in Random Access Memory. The material covered here equips students with the main backend algorithms and infrastructure necessary for the completion of the required Capstone Project for MSDS graduation.

II. Prerequisites

CS 512 or CS 513 (Fundamental Algorithms)

Linear Algebra, Basic Probability (Moments, Typical Distributions, MLE)

Programming Languages: C++/Java

Infrastructure: Hadoop Cluster

III. Course Goals

Students will learn algorithms to store, process, mine, analyze, and synthesize streaming data, or data at rest that does not fit in Random Access Memory. Students will also get understandings of various massive data applications and their algorithms including recommendation system, social networks, stream data mining and machine learning, etc.

IV. Textbooks and Readings

(LRU) Mining Massive Data Sets by J. Leskovec, A. Rajaraman, J. D. Ullman.
(<http://www.mmids.org/>)

V. Homework, Midterm and Final Project

Homework: 4 homework assignments (10% each, 40% total)

Midterm: Thu, Mar 8, 12:00pm – 1:15pm (25%)

Final Project: Complete as a team of at most 3 students, choose between an assigned project or a self-proposed project, and provide a 10 min presentation (35%, and 5% bonus for self-proposed projects)

The midterm is closed-book, but you are allowed to bring one US letter-sized page of note that you have prepared by yourself.

Homework assignments are posted on Fridays, and you have 2 weeks to finish the homework, due dates are on Fridays at 11:59pm. A total of 2 late days are granted.

VI. Tentative Schedule

Note that the schedule may be subject to change (e.g., due to snow or campus close). Please check the course website for the latest schedule.

Class #	Date	Topics	Reading
1	1/18	Introduction, Map Reduce I	Ch1, Ch 2.1-2.4
2	1/22	Map Reduce II	Ch 2.1-2.4
3	1/25	Association Rule Mining	Ch 6
4	1/29	Frequent Item Sets Mining	Ch 6
5	2/1	Locally Sensitive Hashing I	Ch 3.1-3.4
6	2/5	Locally Sensitive Hashing II	Ch 3.5-3.8
7	2/8	Clustering, similarity, k-means, BFR	Ch 7.1-7.4
8	2/12	Dimensionality Reduction, SVD	Ch 11.1-11.3
9	2/15	Dimensionality Reduction, CUR	Ch 11.4-11.5
10	2/19	Content-based Recommendation	Ch 9.1-9.2

11	2/22	Collaborative Filtering, Latent Factor Models	Ch 9.3-9.4
12	2/26	Deep Learning for RS, Project Description	
13	3/1	Link Analysis, Page Rank	Ch 5.1-5.3, 5.5
14	3/5	Web Spam, Trust Rank	Ch 5.4
15	3/8	Mid-term exam	
	3/12		
	3/15		
16	3/19	Social Networks, Community Detection	Ch 10.1-10.2, 10.6
17	3/22	Overlapping Communities	Ch 10.3-10.5, 10.7-10.8
18	3/26	Large-scale Machine Learning I	Ch 12
19	3/29	Large-scale Machine Learning II	Ch 12
20	4/2	Mining Data Streams I	Ch 4.1-4.3
21	4/5	Mining Data Streams II	Ch 4.4-4.7
22	4/9	Computational Advertising	Ch 8
23	4/12	Learning through Experimentations	
24	4/16	Optimizing Submodular Functions	
25	4/19	Case Study in Smart Cities	
26	4/23	Project Presentations I	
27	4/26	Project Presentations II	
28	4/30	Course Review	

VII. Course Policies

Plagiarism

It is expected that the work you submit in this course will be your own. Collaborations and discussions are encouraged, but the final work should be finished and submitted by yourself. If you have conducted any discussions or received any help from others, acknowledgement should be provided at the end of your work.

Special Needs

Students with disabilities who require special accommodations or other classroom modifications should notify the instructor and the University's Disability Services no later than the 2nd week of the term. Students may be asked to provide documentation of their disability to determine the appropriateness of the request. Rutgers Disability Services is located at Lucy Stone Hall, Suite A145, Livingston Campus, 54 Joyce Kilmer Avenue, Piscataway, NJ 08854-8045, and can be contacted at 848-445-6800 or dsoffice@echo.rutgers.edu. Students who must miss an exam or class due to religious observances must notify the instructor ahead of time and make alternative arrangements.