

COMP 640: Graduate Seminar In Machine Learning



Rice University

Anshumali Shrivastava

anshumali At rice.edu

22nd August 2016

- Instructor : Anshumali Shrivastava
- Email : anshumali AT rice.edu

- Class Timing: Monday 3pm to 4:30 pm
- Class Location : Duncan Hall 2014
- Office Hours : Monday 4:30pm - 5:30pm, Duncan Hall 3118
- Website: www.cs.rice.edu/~as143/COMP640_Fall16/index.html
- Canvas

Learn modern techniques for scaling up Machine Learning for Massive Datasets

- We will read some cool recent papers !
- Full Schedule on Website.

Broad Topics

- Sketching Algorithms
- Randomized Machine Learning
- Optimizations
- Advanced Sampling
- Differential Privacy
- Question Answering

How will it work ?



Read the suggested papers before coming to class, there will be a warm up quiz.

- We will discuss a set of (connected) papers every week.
(Webpage for complete list)

Read the suggested papers before coming to class, there will be a warm up quiz.

- We will discuss a set of (connected) papers every week. (Webpage for complete list)

Presentation Logistics

- Each one of you picks a paper from the list, starting 09/12, to present. (**Due by 8/29 next class**)
We will resolve conflicts by usually first come first served basis, so choose soon (I will set up a Google drive Link)
- A week before your scheduled presentation, you give a test run to me.
Example: If the presentation is on 09/12 then in office hours of 09/05 you give me a test run.
- One paper can be presented in a group of at most two.

For 1 credit

- One presentation
- Class participation

For 3 credits

- In addition, a semester long project. (In a group of at most 2)
- It will have two presentation, one proposal and one final presentation.

**Please read suggested papers
before coming to the class.**

Components

- Semester long
- In a group of at most 2. (For larger group ask me)
- Ideally it should have connections with data mining or machine learning. Ask me if you have confusions.

Timelines

- Sept 9th, Project Proposals due by email to me.
1-3 pages describing why its important (motivation), problem statement and why it is feasible.
- Oct 17th, 10-15 min mid-term project presentation in class (Proposal)
- Nov 28th, Final project presentation.

What can be a good ML project ?



- Take a well known algorithm and try to make it faster.
 - Propose a novel fast approximate version.
 - Identify bottlenecks and opportunities to parallelize in a novel way.
 - Take an interesting dataset and try to find something interesting using custom ML models.
 - Propose an alternative to well known models in some real environment.
 - Propose a ML (like deep learning) algorithm/model for a novel application with real data.
 - Theoretical analysis of some new properties of known or proposed algorithms.
-
- Ideally a good project should be publishable if the goals are met.
 - Project can be totally unrelated to topics covered in class.
 - **START EARLY.**

Important Dates to Remember



- 8/29 next class : Your paper preferences.
- 9/9 : Project Proposals due.
- 10/17 : 10 min mid term project presentation in class
- 11/28 : Final project presentation.

Warmup: Some Probabilistic Tricks (On Board)