COMP 640: Graduate Seminar In Machine Learning

RICE

Rice University

Anshumali Shrivastava

anshumali At rice.edu

24th August 2015

About

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

- Class Timing: Monday 3pm to 4:30 pm (Except on 28th Sept)
- Class Location : Duncan Hall 3076
- Office Hours : Monday 4:30pm - 5:30pm, Duncan Hall 3118
- Website: www.cs.rice.edu/~as143/COMP640_Fall15/index.html
- Piazza: https://piazza.com/class#fall2015/comp640

Our Focus

Learn modern techniques for scaling up Machine Learning for Massive Datasets

- We will read some cool papers !
- Some of these papers are best paper awards in recent topmost conferences.
- Some are classical and top cited papers in the field.

Three major Directions

- Use randomized algorithms for reducing the computation.
- Use of parallelizations to speed up machine learning.
- Delve more into Deep Learning.

Roadmap

Hashing Algorithms for Search and Learning

- Locality Sensitive Hashing for Sub-linear Search (8/31)
- Integrate Hashing with SVMs (9/14 and 9/28)
- Making Hashing Techniques Faster (9/21 and 10/5)
- Real Application (10/26)

Roadmap

Hashing Algorithms for Search and Learning

- Locality Sensitive Hashing for Sub-linear Search (8/31)
- Integrate Hashing with SVMs (9/14 and 9/28)
- Making Hashing Techniques Faster (9/21 and 10/5)
- Real Application (10/26)

Recent Advances in Deep Learning

- A Recent Successful Technique for Training Deep Networks (11/2)
- Theory for Deep Learning (11/09)

Rovadmap

Hashing Algorithms for Search and Learning

- Locality Sensitive Hashing for Sub-linear Search (8/31)
- Integrate Hashing with SVMs (9/14 and 9/28)
- Making Hashing Techniques Faster (9/21 and 10/5)
- Real Application (10/26)

Recent Advances in Deep Learning

- A Recent Successful Technique for Training Deep Networks (11/2)
- Theory for Deep Learning (11/09)

Topic Models and Scalable Inference

- Classical LDA and Variational Inference. (11/23)
- Scaling up LDA and faster Bayesian inference. (11/30)

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

- We will discuss two (connected) papers every week.
  (Webpage for complete list)

How will it work ?

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

- We will discuss two (connected) papers every week. (Webpage for complete list)

Presentation Logistics

- Each one of you picks a paper from the list, starting 09/14, to present. (**Due by 8/31 next class**)
  We will resolve conflicts by usually first come first served basis, so mail me you preference soon.

- A week before your scheduled presentation, you give a test run to me.
  Example: If the presentation is on 09/21 then in office hours of 09/14 you give me a test run.

- One paper can be presented in a group of at most two.

Grading Policies

For 1 credit

- One presentation
- Class participation

For 3 credits

- In addition, a semester long project. (In a group of at most 2)

Please read suggested papers before coming to the class.

Projects and Timelines

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 6th, Project Proposals due by email to me.
  1-3 pages describing why its important (motivation), problem statement and why it is feasible.
- Oct 19th, 10 min mid term project presentation in class
- Nov 30th, 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

- 8/31 next class : Your paper preferences.
- 9/6 : Project Proposals due.
- 10/19 : 10 min mid term project presentation in class
- 11/30 : Final project presentation.

Is 5:30pm - 7pm Fine ? Or any time except 3pm - 5pm.

Next Lecture : Locality Sensitive Hashing