CSCi 8980: Project Thoughts

~~Midterm~~

- Decided this is not the best use of our time
- Rather you spend energy and focus on projects
- We will have a few more papers
  - 1 more volunteer for Thursday?

Projects

- We’ve read papers, wrote critiques, now we get our hands dirty!
- Groups of any size (2-4 would be typical)
  - Brainstorming HIGHLY encouraged!
- Main goal: learn something new, do something interesting
- It is **your** job to propose a project idea
- I will work with you to refine the project as needed

Learn something

- Don’t do something you haven’t already done
- Use this opportunity to stretch yourself to try someone new
    - Risk is encouraged!
    - If it “fails”, you can tell us lessons learned

- 1 page proposal due 3/14

Constraints: Systems + ML

1. Relate to a system, application, or an application framework
   - Think about what domain interests you?
   - What is the metric of interest? performance, power, accuracy, reliability, etc., etc.
   - What is the question you are asking? Do you need source code? Can you simulate?

2. Must **employ** one or more ML or DL methods
   - Why is ML a good choice?
   - Compare to reasonable baseline(s)

I am very flexible on “system” and “ML” aspects

Got Data?

Elephant in the Room: getting “data”

- Search for any published logs, contact paper authors (I can help if necessary)
- It is fine to synthetically generate data based on valid assumptions => goal is to show potential

- Run a system on benchmarks to generate data
  - TPC-*: various DBs, Spark, etc.

- Github search?

Some Data Sources

- HPC:

  Failure logs: https://www.usenix.org/cfdr

  Parallel job traces: http://www.cs.huji.ac.il/labs/parallel/workload/

- Clusters/Clouds:

  Borg: https://github.com/google/cluster-data

  Azure VMs: https://github.com/Azure/AzurePublicDataset

  Alibaba: https://github.com/alibaba/clusterdata

- Big Data platforms:

  Many: https://alexpucher.com/blog/2015/06/29/cloud-traces-and-production-workloads-for-your-research/

Old Traces are fine

If stuck: Reproduce a Paper

- Take a paper and try to reproduce the idea (or a part therefore) and evaluate or extend
  - Github and/or contact the authors

- Look ahead to future paper abstracts to see if something looks interesting down the road

Remember Jeff Dean’s advice

Project Proposal

- You provide a 1 page writeup of the project by 3/14
  - domain + methods + problem you are addressing
  - justify why ML is (hopefully) good for this problem
- The project **must provide *some* evidence of feasibility**
  - You have already started on it; located software, data, etc.
  - Have the necessary resources
- The project **must contain a set of clear goals or objectives**
  - Ranked these from easy to hard
- Experimental plan
  - Rough sketch

Recap

- You **must propose an idea of your own** and develop a project around it
  - Do something that excites you and will challenge you

- You are welcome to run **your** ideas by me in office hours at any time