Course Overview
CSci 8980: ML at Large Scale and High Dimensions

Instructor: Arindam Banerjee

January 22, 2014
General Information

- Course Number: CSci 8980
- Class: Mon Wed 01:00-02:15 pm
- Location: Akerman Hall 227
- Instructor: Arindam Banerjee
- Office Hours: Keller Hall 6-213 Mon Wed 02:30 - 03:30 pm
- Web page: http://www-users.cselabs.umn.edu/classes/Spring-2014/csci8980-lshd/
- Email:
  - banerjee@cs.umn.edu
  - Mention 8980 in the title, so I know
Course Activities

- Please read the syllabus carefully
- Paper Reviews: 45% of total grade
  - Individual activity
  - Review 15 (out of 21) papers from the list
  - Each paper is 3% of total grade
- Paper Presentations: 20% of total grade
  - Present in paper specific groups of (at least) 2
  - Present (at least) 2 papers, each paper is 10%
- Project: 35% of total grade
  - Group activity, groups of 2
  - Proposal, progress report, final report, presentation
Course Activities: Paper Reviews

- Individual activity
- Choose 15 (of 21) papers you want to review
  - Do not have to submit the list to me upfront
- Each paper is 3% of total grade
- 2 page (max) writeups, including references
  - 11 pt font, 1 inch margin, in pdf
  - Latex template will be provided
- Guidelines:
  - Problem considered, related work
  - Main result, associated results
  - Key insights/ideas used or introduced
  - Potential limitations, if any, and future directions
- Due before class, in moodle
Course Activities: Paper Presentations

- Each student will do (at least) 2 paper presentations
- Each paper will be presented by (at least) 2 students
  - Group will be paper specific
  - Determined by preferences for presentation
- Submit preferences over all 21 papers:
  - Ranking (preference list) over all 21 papers
  - Due: **Tue, Jan 28** in moodle
- Assignments finalized by Thu, Jan 30
  - Each paper will be assigned to 2 students
- Plan a 45 min presentation
  - Work together on the presentation
Course Activities: Project

- Groups of 2 students, form groups by **Fri Feb 28**

**Project**
- Proposal: 1-page, due **Fri, Mar 14**
- Progress Report: 2-page, due **Fri, Apr 11**
- Final Report: 5-page + references, etc., due **Fri, May 09**
- Presentation: Last two weeks of the semester (see schedule)

**Scope of project has to be aligned with the course topics**
- High-dimensional statistics
- Large scale optimization
- Graphical model structure learning
- Structured sparsity, e.g., low ‘rank’ matrices/tensors, etc.
- Superposition or ‘dirty’ models, e.g., low rank + sparse, etc.
- ...

**Cannot simply apply a ML method to a dataset**
Grading

- Paper Reviews: 45% = 15 × 3%

- Paper Presentation: 20% = 2 × 10%

- Project: 35% = 5 + 5 + 15 + 10%  
  (Proposal + Progress Report + Final Report + Presentation)

- Grading is absolute: A = 90-100, A- = 85-90, B+ = 80-85, B = 70-80, B- = 65-70, C+ = 60-65, C = 50-60, F = less than 50.

Instructor: Arindam Banerjee
Overview

- **Sparsity**
  - Initiated in the 90s (studied even earlier)
  - Models with sparsity and structured sparsity
  - Focus on theoretical advances

- **Structure Learning**
  - Conditional independence for high-dimensional distributions
  - Provable recovery using sparse estimation

- **Optimization**
  - Large scale non-smooth optimization
  - Online and stochastic optimization
  - Methods for structured sparsity
Opt 1  John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Ambuj Tewari, Composite Objective Mirror Descent, Conference on Learning Theory (COLT), 2010.


Paper Presentation Schedule 2


Paper Presentation Schedule 5


