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/
- Moodle page: https://ay13.moodle.umn.edu/course/view.php?id=10099
- 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 $\times$ 3 %
- Paper Presentation: 20 % = 2 $\times$ 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.

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

Paper Presentation Schedule 1

Opt 1 John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Ambuj Tewari, Composite Objective Mirror Descent, Conference on Learning Theory (COLT), 2010.

Opt 2 Anatoli Juditsky and Arkadi Nemirovski, First-Order Methods for Nonsmooth Convex Large-Scale Optimization, I: General Purpose Methods, Optimization for Machine Learning, 2011.

Spt 1 Martin J. Wainwright, Sharp Thresholds for High-Dimensional and Noisy Sparsity Recovery Using L1-Constrained Quadratic Programming (Lasso), IEEE Transactions on Information Theory, 55:5, 2183-2202, 2009.

Spt 2 Peter J. Bickel, Ya'acov Ritov, and Alexandre Tsybakov, Simultaneous analysis of Lasso and Dantzig Selector, The Annals of Statistics, 37:4, 1705-1732, 2009

Paper Presentation Schedule 2

Opt 3 John Duchi, Elad Hazan, and Yoram Singer, Adaptive Subgradient Methods for Online Learning and Stochastic Optimization, Journal of Machine Learning Research, 12, 2121-2159, 2011.

Opt 4 Huahua Wang and Arindam Banerjee, Online Alternating Direction Method, International Conference on Machine Learning (ICML), 2012.

Spt 3 Sahand N. Negahban, Pradeep Ravikumar, Martin J. Wainwright, and Bin Yu, A Unified Framework for High-Dimensional Analysis of M-Estimators with Decomposable Regularizers, Statistical Science, 27:4, 538–557, 2012.

Spt 4 Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky, The Convex Geometry of Linear Inverse Problems, Foundations of Computational Mathematics, 12, 805-849, 2012.

Paper Presentation Schedule 3

Spt 5 Yaniv Plan and Roman Vershynin, Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach, IEEE Transactions on Information Theory, 59:1, 482-494, 2013.

Opt 5 John Duchi, Alekh Agarwal, and Martin Wainwright, Dual Averaging for Distributed Optimization: Convergence Analysis and Network Scaling, IEEE Transactions on Automatic Control, 57:3, 592-606, 2012.

Opt 6 Ohad Shamir and Tong Zhang, Stochastic Gradient Descent for Non-smooth Optimization: Convergence Results and Optimal Averaging Schemes, International Conference on Machine Learning (ICML), 2013.

Opt 7 Sebastian Bubeck and Nicolo Cesa-Bianchi, Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems, Foundations and Trends in Machine Learning, 5:1, 1-122, 2012.

Paper Presentation Schedule 4

Spt 6 Garvesh Raskutti, Martin J. Wainwright, and Bin Yu, Restricted eigenvalue properties for correlated Gaussian designs, Journal of Machine Learning Research, 11, 2241-2259, 2010.

Spt 7 Mark Rudelson and Shuheng Zhou, Reconstruction from anisotropic random measurements, IEEE Transactions on Information Theory, 59:6, 3434-3447, 2013.

Opt 8 Lei Yuan, Jun Liu, and Jieping Ye, Efficient Methods for Overlapping Group Lasso, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35:9, 2104-2116, 2013.

Opt 9 Andreas Argyriou, Rina Foygel, and Nathan Srebro, Sparse Prediction with the k-Support Norm, Advances in Neural Information Processing Systems (NIPS), 2012.

Paper Presentation Schedule 5

Opt 10(a) Ambuj Tewari, Pradeep Ravikumar, and Inderjit Dhillon, Greedy Algorithms for Structurally Constrained High Dimensional Problems, Advances in Neural Information Processing Systems (NIPS), 2011.

Opt 10(b) Martin Jaggi, Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization, International Conference on Machine Learning (ICML), 2013.

StL 1 Pradeep Ravikumar, Martin J. Wainwright, Garvesh Raskutti, and Bin Yu, High-dimensional covariance estimation by minimizing l1-penalized log-determinant divergence, Electronic Journal of Statistics, 5, 935-980, 2011.

StL 2 Tony Cai, Weidong Liu and Xi Luo, A Constrained L1 Minimization Approach to Sparse Precision Matrix Estimation, Journal of American Statistical Association, 106:494, 594-607, 2011.

Paper Presentation Schedule 6

StL 3(a) Han Liu, Fang Han, Ming Yuan, John Lafferty, and Larry Wasserman, High Dimensional Semiparametric Gaussian Copula Graphical Models, The Annals of Statistics, 40:4, 2293-2326, 2012.

StL 3(b) Lingzhou Xue and Hui Zou, Regularized Rank-Based Estimation of High-Dimensional Nonparanormal Graphical Models, The Annals of Statistics, 40:5, 2541-2571, 2012.

StL 4 Pradeep Ravikumar, Martin J. Wainwright and John Lafferty, High-Dimensional Ising Model Selection using L1-regularized Logistic Regression, The Annals of Statistics, 38:3, 1287-1319, 2010.