Machine Learning and Computational Statistics (DS-GA 1003)

David Rosenberg

New York University

January 28, 2015

Logistics

- Class webpage: `https://davidrosenberg.github.io/ml2015`
  - Syllabus on the website
- Piazza: `https://piazza.com/nyu/spring2015/dsga1003`
  - Ask questions here
- Required Lab Session
  - Thursdays: 6:10pm - 7pm (WWH 109)
  - **First session:** Guest lecture on differentiating w.r.t matrices and vectors
  - Lab instructor: TBD

Lab Sessions

- Supplemental Topics
- Mathematical topics (e.g. subgradients, matrix derivatives, SVD)
- Demonstrations or examples related to lecture
- Answer common questions from Piazza
- Midterm review
- Meeting with project advisors

Evaluation

- About 8 to 10 homeworks (60%)
- Midterm Exam (20%)
  - In 10th week of class, after spring break
- Project (20%)
- Extra Credit Opportunities
  - Up to (2%) for answering questions on Piazza
  - Optional problems or competitions on the homework

Homework (60%)

- First assignment out tonight – Due in one week.
- Submit with NYU Classes: `https://newclasses.nyu.edu`
- Late homework: Accepted up to 48 hours late with 20% penalty
- Collaboration is fine, but
  - Write up solutions and code on your own
  - List names of who you talked to about each problem
- Graders:
  - Hao Xu, Ran Bi, Prasoon Goyal

Midterm (20%)

- **Late midterm: 10th week of term** (April 8th)
- In class – during lecture

Projects (20%)

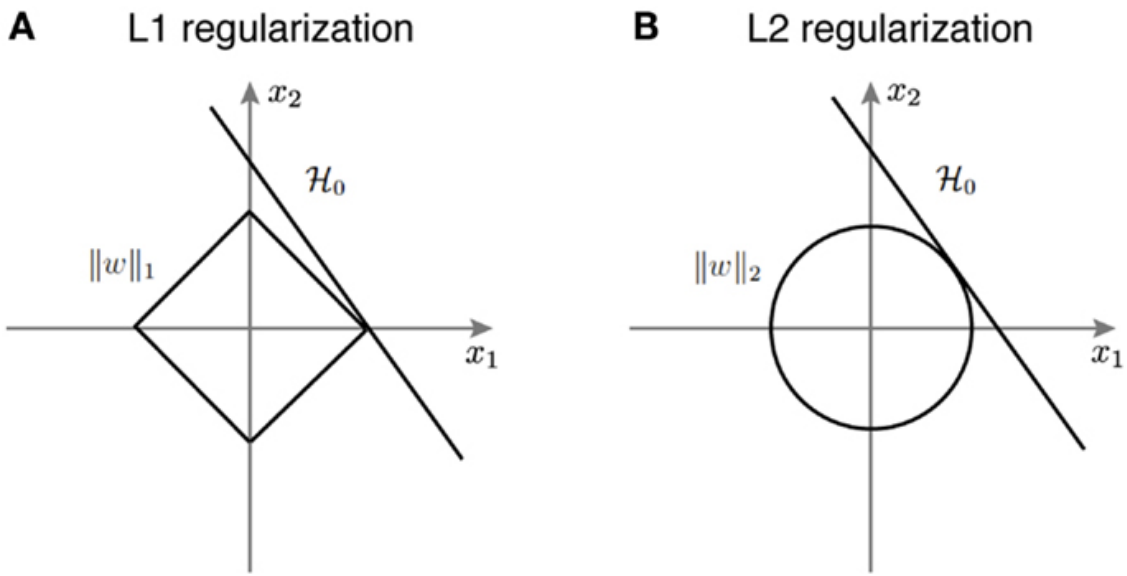
- Find some new data or new approach to old data
- Project philosophy the same as in these courses:
  - `http://cs.nyu.edu/~dsontag/courses/ml14/assignments/projects.html`
  - `http://web.stanford.edu/class/cs221/project.html`
- Logistics:
  - 2 students per group
  - First meeting with advisors **March 12th**
  - Project proposal due after Spring Break: **March 26th**
- Advisers:
  - Kurt Miller, Gideon Mann, plue one or two more TBD

Prerequisites

- Introduction to Data Science (DS-GA 1001)
- Math
  - Multivariate Calculus
  - Linear Algebra
  - Probability Theory
  - Statistics
- Python programming (numpy)

General Philosophy

- Mastery vs Performance
  - (understanding vs "getting the grade")
- Don't confuse "kind of understanding" with "actual understanding"
- Can you explain this picture?



Course Topics

- Frequentist Approaches
  - ERM, regularization, SVM, kernels, ensemble methods, neural networks
- Probabilistic Models
  - GLM, Bayesian networks, Gaussian mixture models, EM algorithm, HMM
- Bayesian Approaches
  - priors/posteriors, hierarchical models, sampling methods, Bayesian model selection
- Misc. and Advanced Topics
  - dimensionality reduction, structured prediction

Questions?

- What are you looking to get out of the course?
- Questions for me?