Thought Leaders in Data Science and Analytics
- Big Data Analytics

Ram Akella
akella@ischool.berkeley.edu
Cell 650-279-3078
754 SDH (Sutardja Dai Hall)
http://courses.ischool.berkeley.edu/i296a-dsa/s14/

January 22, 2014

Introduction

- Technology cycles
  - Analytics
  - Architecture/Infrastructure
  - Interaction
- A possible cut
  - Enterprise Analytics:
    - Enterprise databases (DB)and Business Intelligence (BI)
  - Web Analytics
    - Leading to Hadoop, Spark/Shark, Streaming + Analytics
  - Internet of Things
    - Continuous sensing and proactive response
    - What is new and different about it?

Data Science and Analytics: What is it?

- Components
  - Data collection, storage, and basic processing
    - Architecture and Infrastructure
  - Analytics
  - Domain
  - Business Needs
- To solve real Big Data problems, need expertise in some or all of these areas
- Need to form teams!

Seminar and Course Structure

- Set of broad ranging industry talks
  - Provide perspectives on
    - Domain and business needs
    - Infrastructure needs
    - Analytics needs
    - State of the art
- Potential projects, ideas, mentoring (including VCs) and possible data

Projects and Participation

- These are key to learning
- Forming teams is critical
- Need Analytics, infrastructure, business, domain
- We can help
  - Faculty
  - Staff
  - Other students
  - Industry
  - VCs

Background

- Basic
  - An Introduction to Statistical Learning

  (James, Witten, Hastie, Tibshirani)
  - R or equivalent
  - Data Mining, linear algebra, statistics, or equivalent
  - Additional (specialized):
    Field Experiments (Gerber, Green)

Advanced: To discuss

Action

- Sign up sheet
- Set up teams
- Provide CVs
- Start determining data sets and projects
- Meeting times, including Skype (beyond class times)
- Set up boot camp times for Infrastructure and Machine Learning/Data Mining
- Use Piazza!