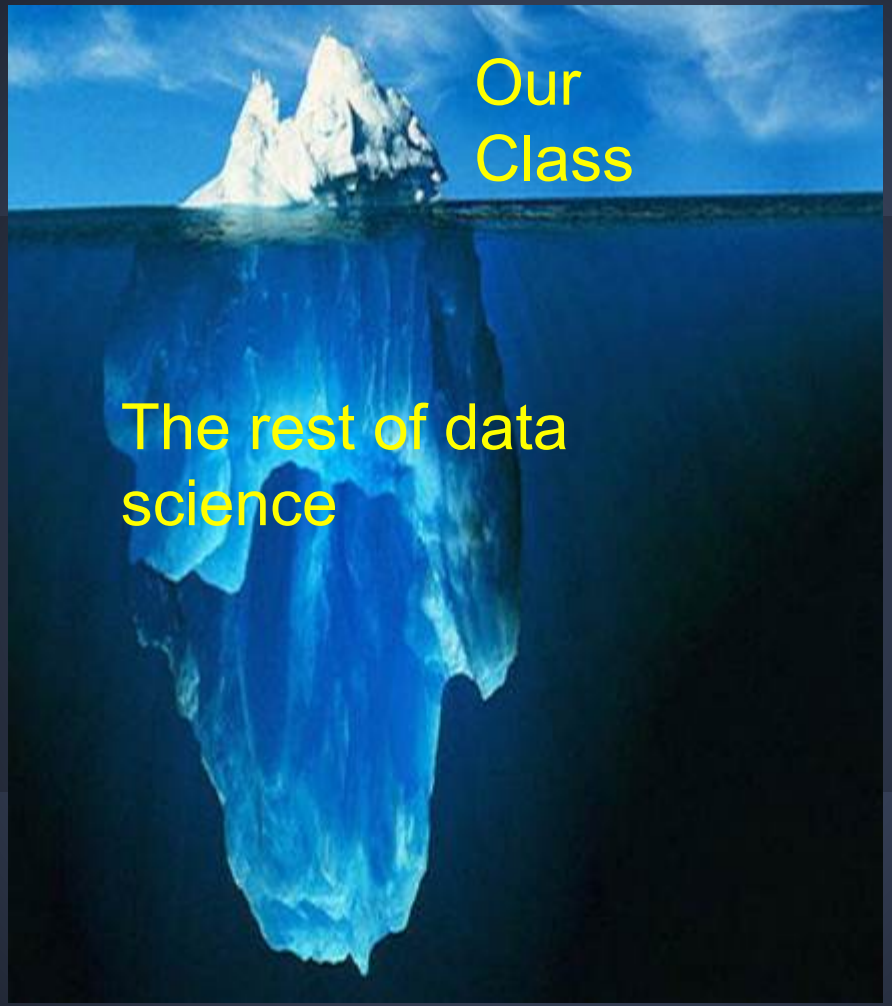
Next Steps After DAT12

Alex Chao


Our
Class
The rest of data
science

Data Science Roadmap

http://www.analyticschampion.com/wp-content/uploads/2013/07/RoadToDataScientist1.png

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21th century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing package e.g. R
- Databases SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop and Hive/Pig
- Custom reducers
- Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot or lattice
- Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization: customer tracking and on-site analytics: predictive analytics and econometrics: data warehousing and big data systems: marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing DISTILLERY

You guys are great!

- We covered a lot of material
- But hopefully now you have the confidence to apply data science to your own use cases

Where Do I Go From Here?

- Grad School in relevant field (CS, Stats, ML)
  - Berkeley (Go Bears!)
    - Two year online Masters in Data Science
      - 60k per year…. ouch
  - Stanford, CMU, MIT, etc.
  - Some offer certificate programs in specialized fields

Where Do I Go From Here?

Data Science for Social Good

http://dssg.io/

- 12 weeks in Chicago
- work in small teams on analytics project in partnership with nonprofits, local governments, or federal agencies
- full-time mentors and technical advisors

Where Do I Go From Here?

Startup.ML Fellowship http://www.startup.ml/blog/fellowship

- 4 months, project-focused
- Building a machine learning system from scratch
- Free and doesn't have hard requirements about having an advanced degree

Where Do I Go From Here?

- Another Bootcamp-like program
  - Galvanize www.galvanize.it/education
    - GalvanizeU - 1 Year Masters (Ask Alessandro!)
    - Galvanize Immersive - 12 weeks full time
    - 16k for immersive, 48k for Masters
  - Insight Data Science (For PhDs only) http://insightdatascience.com/
  - Insight Data Engineering (For anyone!) http://insightdataengineering.com/
    - 7-8 week program, Ramesh is an alum, FREE

Where Do I Go From Here?

The Data Incubator https://www.thedataincubator.com/

- Can do in NYC, Washington DC, SF, or electronically.
- 7 weeks
- Free, but primarily geared towards PhDs

Start Your Own Company!

- GA recently partnered with 500 Startups http://techcrunch.com/2015/03/25/general-assembly-and-500-startups-partner-for-accelerator-prep-program/#SUzvlP:4ESU
- 500 Startups will accept at least 10 teams that feature a GA alumni
- "Successful applicants will receive $8,000 in cash and a $2,000 tuition waiver from 500 Startups in exchange for a 2% equity stake. Those companies that are accepted will also be fast-tracked for the 500 Startups accelerator"
- Apply here!

Own Opinion

- For most of you, you already have the skills to continue building your data science toolkit yourself
- There are so many resources out there
- Bootcamps are expensive
  - They do have value though so decide on the tradeoff (esp. if they are full-time)

Ways to Stay in Touch/Collaborate

- LinkedIn/Facebook
- Slack
- Alessandro's Job Posting Google Group
  - ask him for an invite
- Bay Area Data Science Facebook Group
  - courtesy of David Yerrington
  - https://www.facebook.com/groups/bayareadatascience/
- Study groups / Meetups!

Study Groups / Meetups

- My personal favorite
- Open invite, co-working space for data scientists
- Can discuss what we're working on, resources we can share with each other, hold workshops, and just hang out with other like-minded people
- Each week, one person can be the lead for a subject of his/her choice and teach the rest of us about that domain.

Study Groups / Meetups

- Biggest need will be coordination (when and where we'll meet). I'd be happy to be one of the co-leaders for this.
- We might be able to have a regular spot at GA on weekends (from Vanessa)
- Message me if interested! If Hipchat is deprecated after the class, we'll move to Slack :)

Thanks for a great class!