

Mathematicians, tech needs you!



Ryan Walker, UK PhD 2013



Outline

- **What** is the major challenge facing tech companies today?
- **Why** are mathematicians the right people to solve these problems?
- **How** can I get involved?

We'll also have 3 mathematical interludes during the talk!

About Me

- PhD from UK in 2013 under Prof. Peter Perry
 - Studied inverse problems
 - Started grad school convinced I wanted to study pure math only but became increasingly drawn to applications
- Interned in a medical imaging lab and at a health insurer
- Worked at several startups in machine learning, software engineering, and data science roles
- Currently Chief Technology Officer of Casetext. We use machine learning to help lawyers research and draft legal documents

Industry Problem: ML revolution

- We're generating vast amounts of messy, unstructured data every day (think [about 2MB](#) per person on earth per second). Most of this data is unanalyzed.
- Groundbreaking advances in machine learning (especially in image and language processing) have been made in the past 3-5 years
- There is huge opportunity to **apply** these new techniques and methods to all that messy data and solve *BIG* problems

Interlude I: Parallel Search

 Add Standards  Add Cases

Parallel Search
Enter a [full sentence](#) that describes how the legal standard you've chosen applies to the facts of your case. [Learn more](#)

Luxtottica Should Not Be Permitted to Force Chinese Employees to Fly to Chicago for Depositions.

[Search Cases](#)

500 Recommended cases Show other jx

Gagasoules v. MBF Leasing LLC [Add](#)
08-CV-2409 (ADS)(ARL) (E.D.N.Y. Sep. 29, 2012)

...2. As to Counsel's Conduct During Discovery Between July of 2009 and February of 2010, the parties engaged in a contentious and protracted discovery dispute over the appropriate location for the Plaintiffs' depositions. **Niblett and Garner, both California residents, maintained that they should not be required to travel to New York to be deposed.** MBF took the position that since the Plaintiffs chose to litigate in this forum, they should be expected to travel to New York for depositions. After various discussions by counsel, letter motions, court appearances and an appeal to...

 Add Standards  Add Cases

Parallel Search
Enter a [full sentence](#) that describes how the legal standard you've chosen applies to the facts of your case. [Learn more](#)

the scent of weed is sufficient justification to search a car.

[Search Cases](#)

500 Recommended cases Show other jx

U.S. v. Ibarra-Sanchez [Add](#)
199 F.3d 753 (5th Cir. 1999)

...van, and Mattas testified that he could smell it from five to ten feet away. Regardless of precisely how near or far from the van the officers were when they detected the odor, once they did so they possessed probable cause to search the van. **This Court has consistently held that the smell of marijuana alone may constitute probable cause to search a vehicle.** See, e.g., *McSween*, 53 F.3d at 686-87; *United States v. Reed*, 882 F.2d 147, 149 (5th Cir. 1989) (observing that the smell of marijuana "in itself would have justified the subsequent search of Reed's vehicle"); *United States v. Henke*, 775...

Interlude I: Parallel Search

- *Offline:*
 - Train a deep neural net to map sentences into ***fixed length*** vector representations (that encode the semantics of the sentence)
 - build a corpus of all sentences from case law mapped into this vector representation
- *Online:*
 - User enters query sentence
 - Map query into the fixed length vector representation
 - Find the n nearest neighbors to the query vector in the corpus according to some similarity measure (e.g. cosine similarity)

Why *YOU*?

Hard skills

- **Math**

- Linear algebra!!!
- Vector calculus / optimization
- Probability
- Numerical methods

- **Programming**

- Scientific computing
- Data wrangling
- Data architecture / model serving

Soft skills

- **Research**

- Pose business problems in a way that is answerable with data
- Read papers, find the interesting parts, apply the techniques – fast

- **Teaching/Leading**

- Help smart, non-technical people understand technical approaches, risks, and trade offs
- Work with teams to solve complex problems

Machine Learning Researcher



Research

Data Engineer



Coding

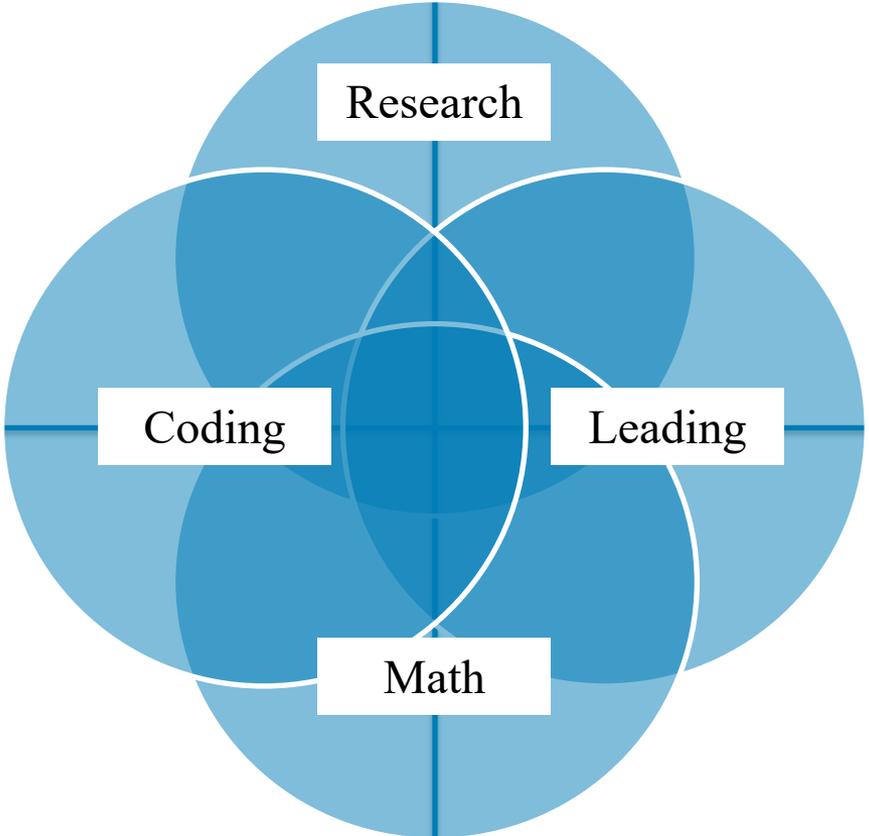
Leading

Data Scientist



Math

Machine Learning Engineer

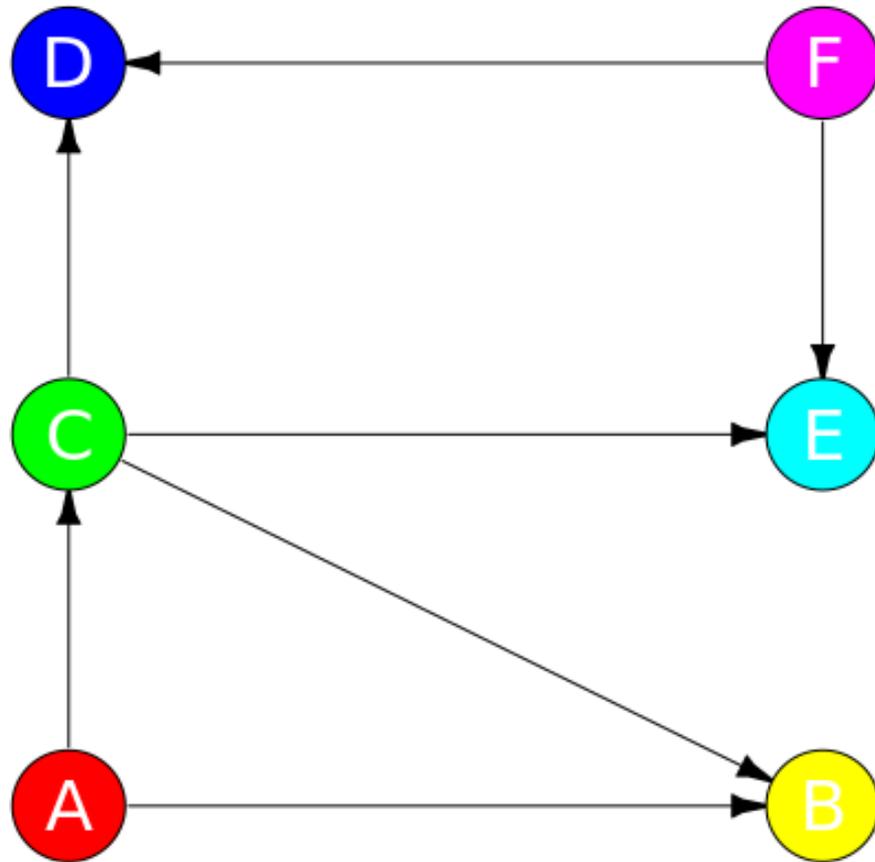


Where is ML happening?

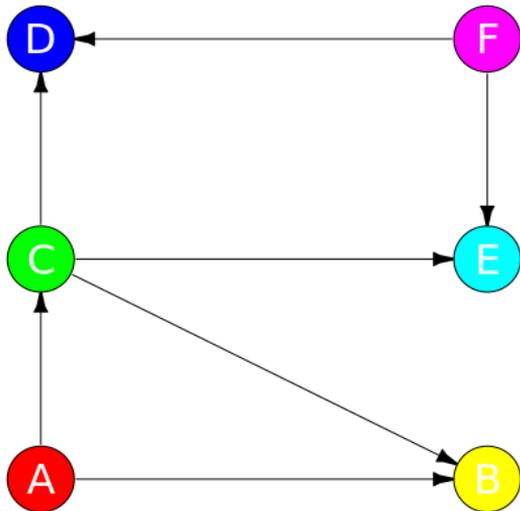
Everywhere

- Big tech (“FAANG”):
 - Pros: great places to start career; lots of structure; top colleagues and cutting edge tech/research
 - Cons: Highly competitive (need to do a *lot* of preparation); work on a small piece of a big problem
- Startups:
 - Pros: Tons of responsibility/ownership and rapid career progression
 - Cons: Less/no structure, little training, very tight timelines
- Traditional companies:
 - Pros: Lots of opportunity as these companies transform to apply ML to their (mostly unanalyzed) big data; Lots of structure – great places to learn how businesses really work
 - Cons: Slow moving; don’t really understand ML tech (yet)

Interlude II: Caselaw Recommendation



Interlude II: Caselaw Recommendation



transpose(G)

x

G

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

x

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

=

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Hmm, what is $(\text{transpose}(G) \times G)^2$?

How to prepare

- Take a broad range of math classes. Make sure to take a few courses that focus on scientific computing
- Develop coding skills. Learn Python and its major scientific computing / ML frameworks
- Try before you buy: **be an intern!**
 - *Applications begin as early as the fall but there are opportunities throughout the year*
- Build a **portfolio** of numerical / machine learning work on GitHub that you can show in interviews

Interviewing

- Preparation:
 - Engineering roles: expect algorithm whiteboard questions
 - Content covers what you'd find in a standard Undergraduate "Algorithms" course from CS Department
 - There are (almost) no new questions – you pass the interviews by **doing a lot of practice problems**
 - Data science / ML roles:
 - Much more variety of questions depending on the roles
 - Expect to be able to describe or analyze an ML technique in detail
 - Probability and stats questions like you'd see in an undergraduate stats course
 - Project based interviews:
 - Show the team a project you are currently working on or work with a team on a current project
 - Take home assignments

Interlude III: Obamacare

Scene: It's early 2013 and you are a data scientist at a large health insurer. Everyone is freaking out because Obamacare is going live next year.

Can you use data to make educated predictions about how marketshare will redistribute among competitors?

Let x be a vector where each x_i represents the number of insured people covered by a particular carrier in a particular market segment. We only know our own book of business and not our competitor's, so many of the x_i are **unknown**.

But we know *a lot of facts* about the market, e.g. California has roughly 8M uninsured. Represent these facts as linear combinations of the x_i :

$$Ax = b$$

Company analysts can make solid guesses about the world. These guesses look like:

$$Qx \approx g$$

Interlude III: Obamacare

So one way to solve this problem is to setup a constrained optimization:

$$\begin{cases} \min_x (Qx - g)^T (Qx - g) \\ \text{subject to: } Ax = b \end{cases}$$

This is a quadratic program, a well studied flavor of optimization problem

Parting thoughts

- Chase the things you find most interesting...but round out your skill set with some numerical computation and programming experience
- Build an industry portfolio alongside your academic portfolio
- Take opportunities to work with people in industry – this is a very interesting (and profitable!) time for applied mathematics

Thank You!

Questions?

Feel free to email me or connect on LinkedIn
ryan@ryanwalker.us

Recommended Resources

- News:
 - [Hacker News](#)
 - [NA Digest mailing list](#)
- Recruiting:
 - [Angel List](#)
- Interview Prep:
 - [leetcode](#)
 - [kaggle](#)
 - [Cracking the Coding Interview](#)
- Programs/workshops
 - [SIAM Summer School](#)
 - [Insight Data Science](#)