Mathematicians, tech needs you!

Ryan Walker, UK PhD 2013

University of Kentucky

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!

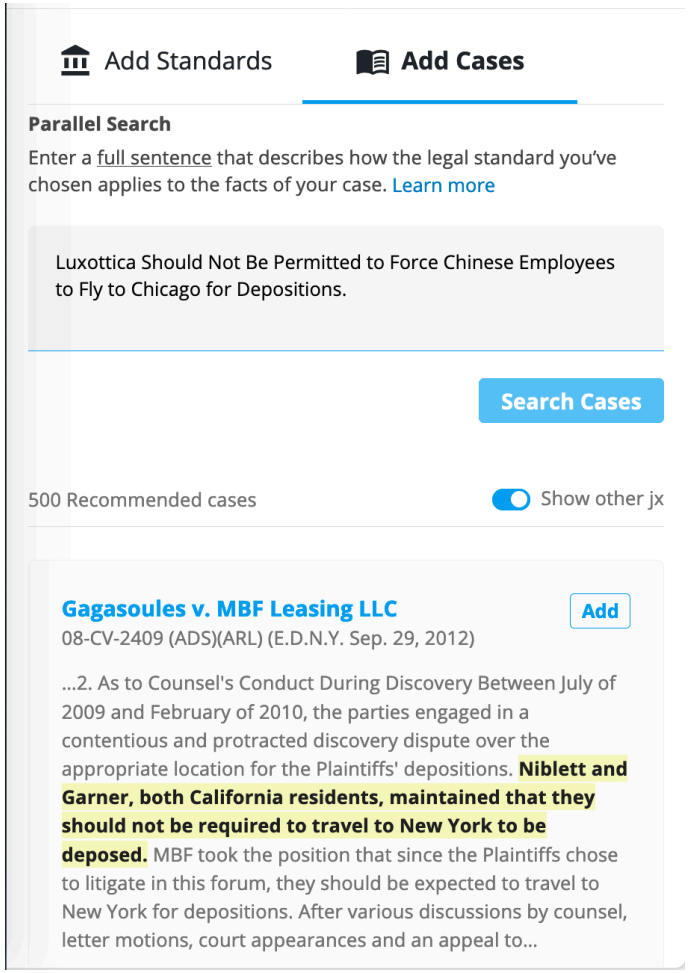
About Me

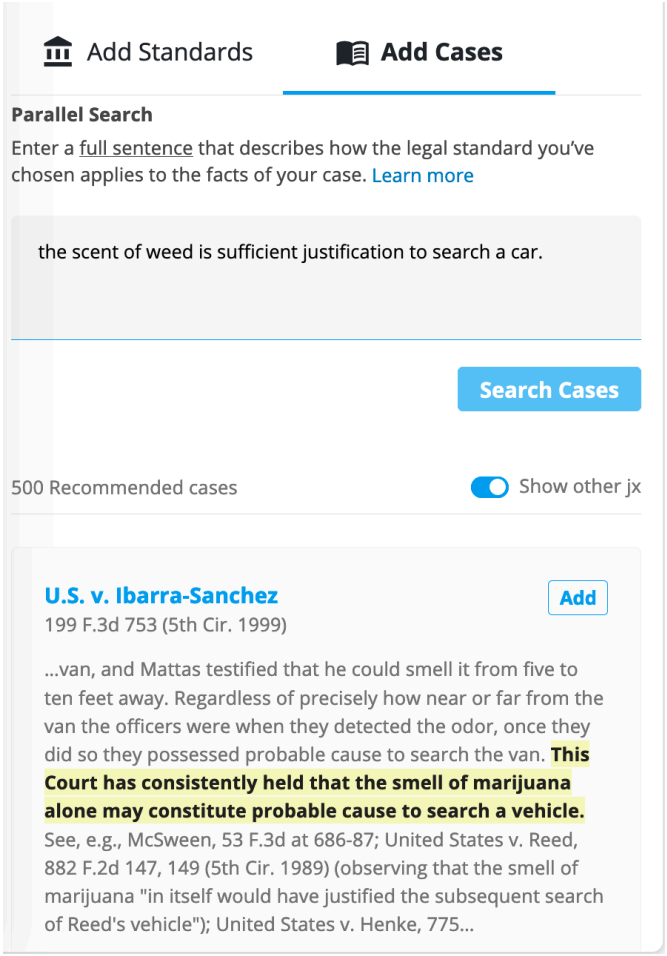
- PhD from UK in 2013 under Prof. Peter Perry
  - Studied inverse problems
  - Started grad school convinced I wanted to study pure math only but became increasingly drawn to applications
- Interned in a medical imaging lab and at a health insurer
- Worked at several startups in machine learning, software 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





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)

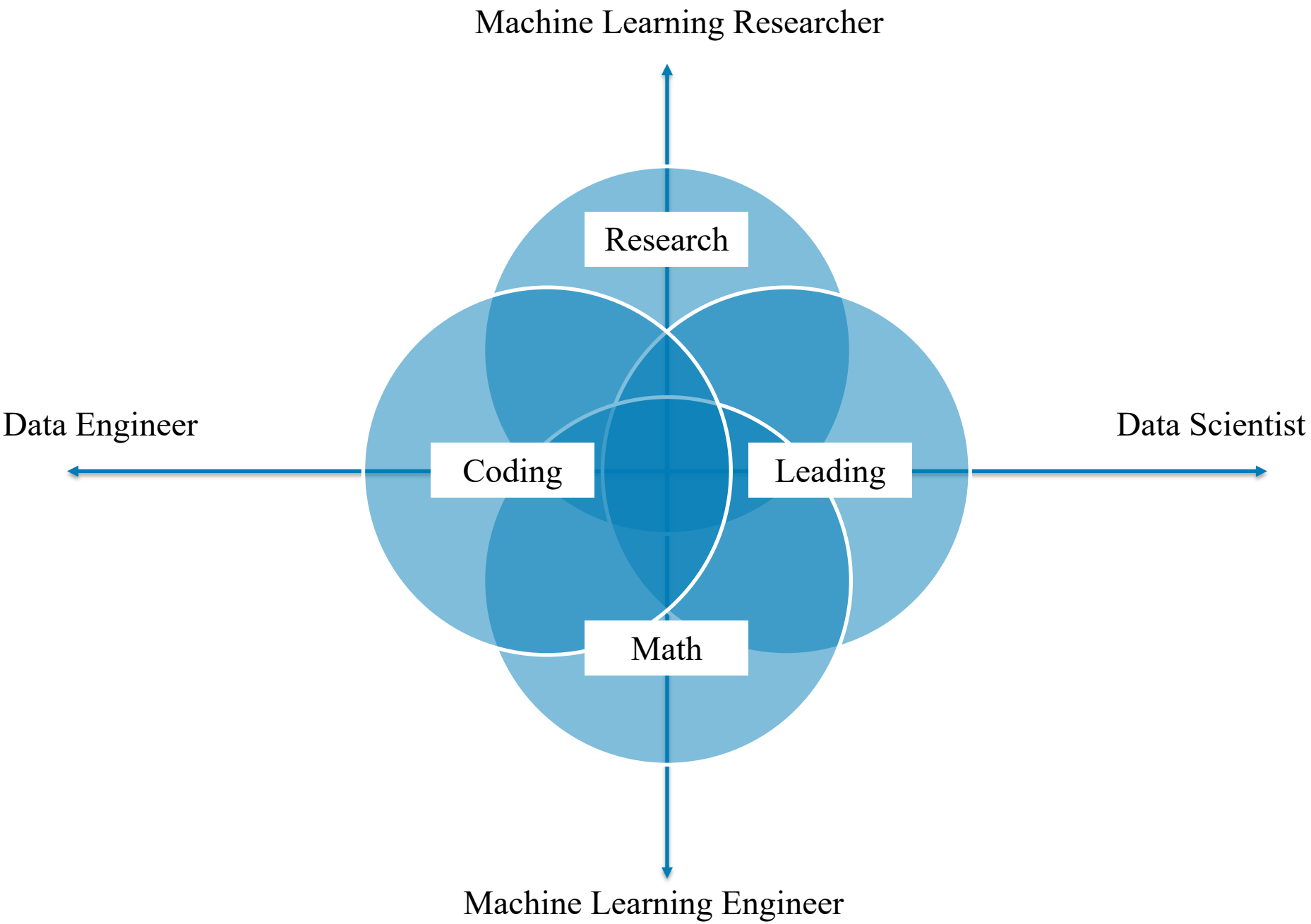
Why *YOU*?

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

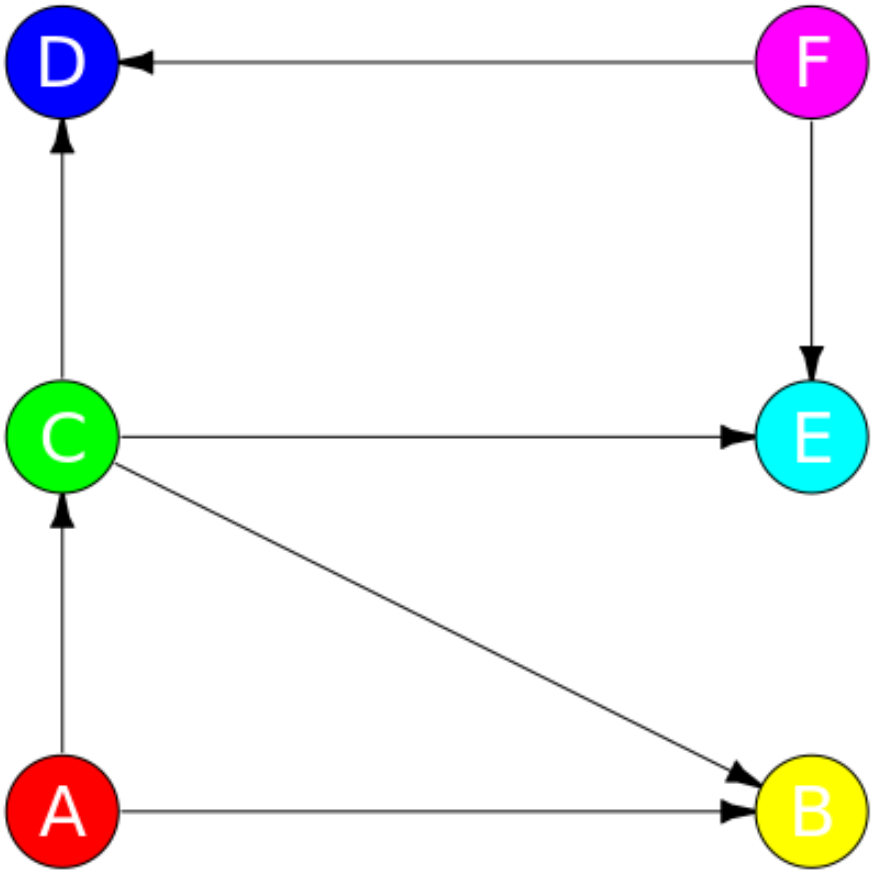

Machine Learning Researcher
Research
Data Engineer
Coding
Leading
Data Scientist
Math
Machine Learning Engineer

Where is ML happening?

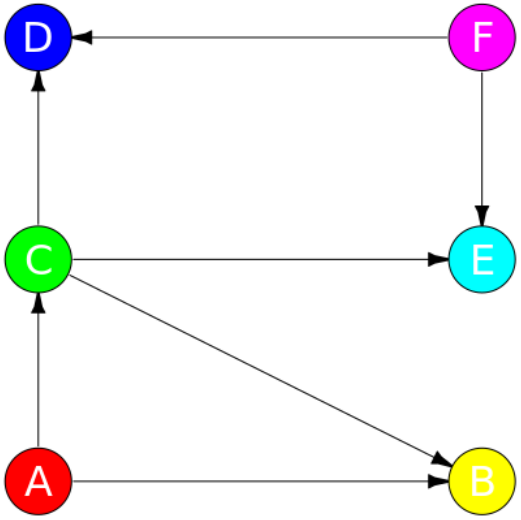
Everywhere

- 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

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

Hmm, what is (transpose(G)xG)^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