Monday April 13th, 2015

KNEWTON

Ruben Naeff
Data Scientist

Data Science at Knewton

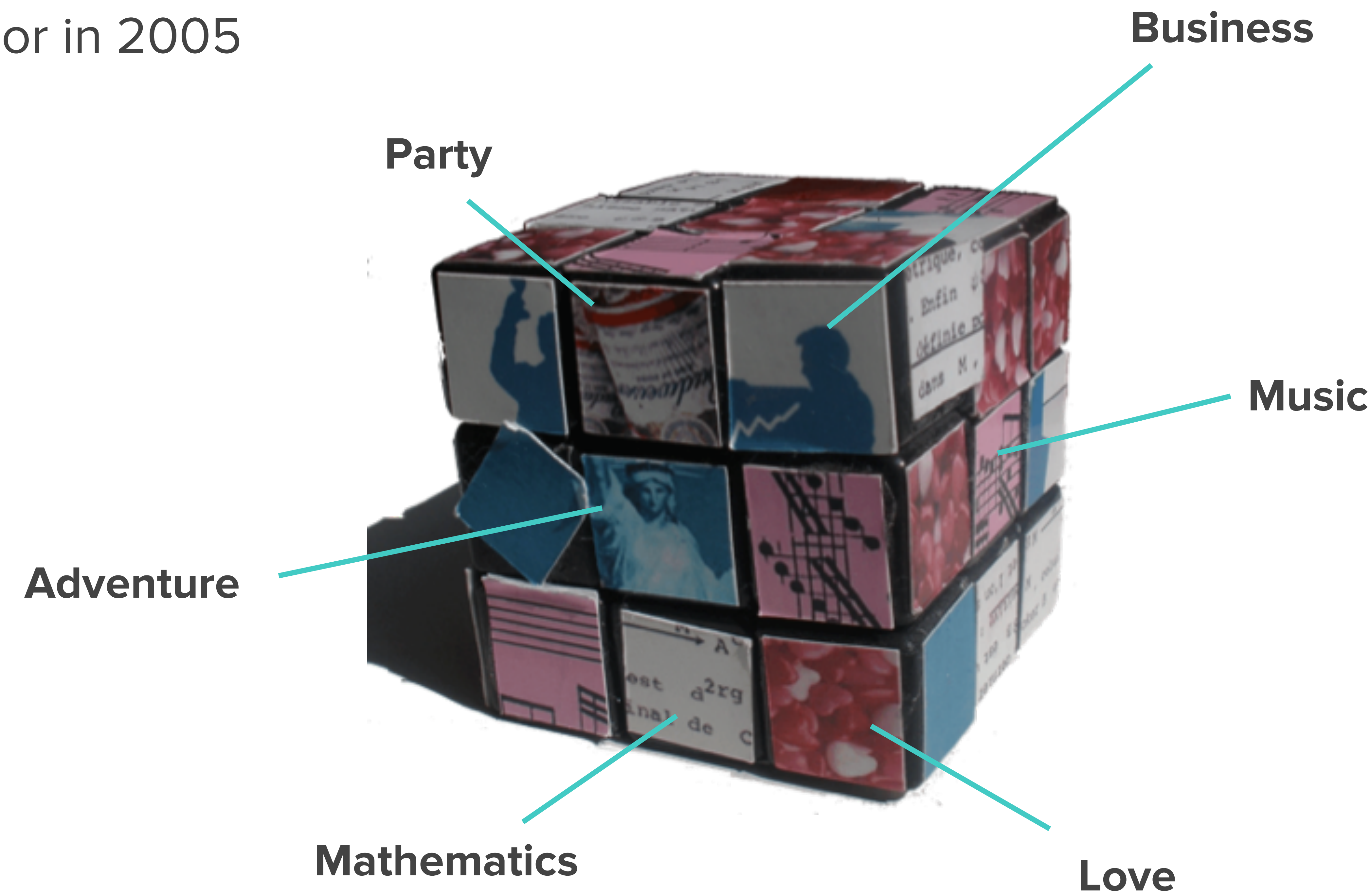
Agenda

- Who am I
- What is Knewton
- Predicting Student Responses
- Q&A

Who am I

Who am I?

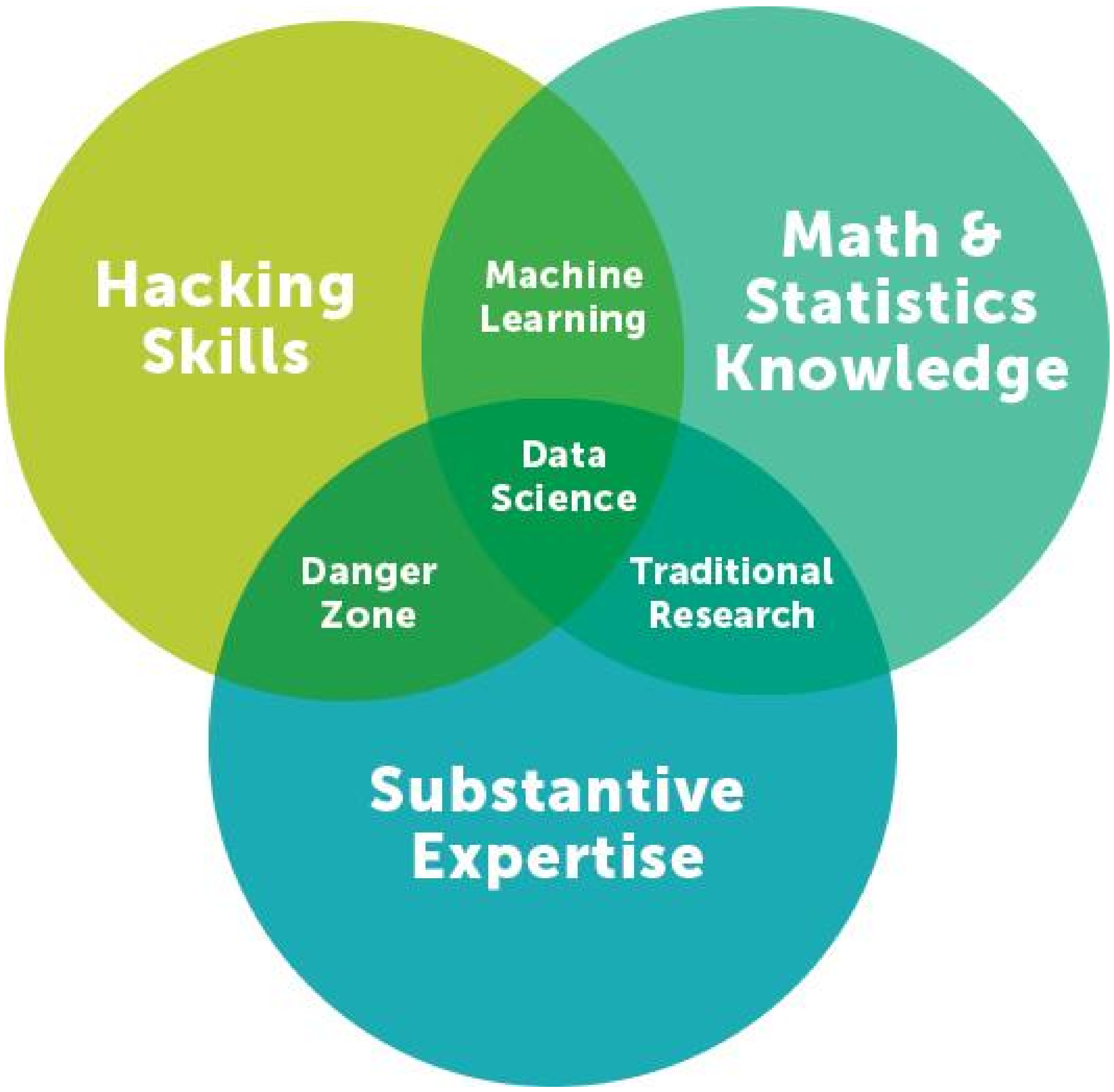
Ruben’s Cube, gift from graduation supervisor in 2005



My resume looked like a hodgepodge...

12-14	HS Math teacher, Dutch teacher, Music teacher
10-12	New York University, M.M. Music Theory and Composition
08-10	Netherlands Competition Authority, Economic Researcher
06-07	OC&C Strategy Consultants, Associate Consultant
99-05	University of Amsterdam, M.S. Mathematics
always	Freelance composer of concert music
..and..	Leisurely coder: as teenager and later (youopera.org)

… until I discovered data science



How does data science compare to the other jobs I had?

	Intellectual work	Actual products	Work conditions
Data Science	Crunch your brain Lots of code, higher math	Fancy visualizations vs. *too* scientific talks	Substantial hours but nothing crazy (yet)
Strategy Consulting	Think on your feet Excel analyses	Powerful, intimidating pitches & presentations	Deadline-driven Crazy hours
Economic research	Thorough research Conceptual analyses	Tough long reads Safe, could be boring	4 days x 9 hours (governmental)
Teaching (HS)	Clever, creative Multiple perspectives	Engaging and lively performances	For early birds Lots of vacation
Composing Music	Creative, trial & error Strive for perfection	Engaging concerts Recordings, oeuvre	Always working, daytime jobs, shows at night

What is Knewton

Knewton helps education publishers making their content adaptive

Empower Every Student
Knewton's mission is to bring personalized learning to the world. We help prepare all students to succeed.

Improve Any Learning Product
Publishers can integrate Knewton adaptive learning into new or existing products across grade levels and subject areas.

KNEWTON

Knewton has three core products

RECOMMENDATIONS

Knewton figures out what each student knows, then recommends the exact activities she should focus on next to meet learning goals.

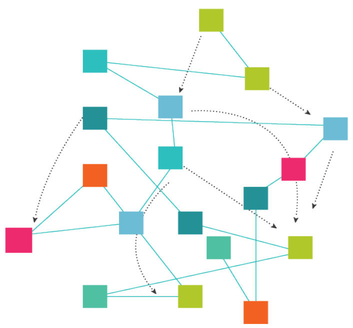
ANALYTICS

Knewton shows educators at a glance what a student needs help with, so they can intervene before students fall behind.

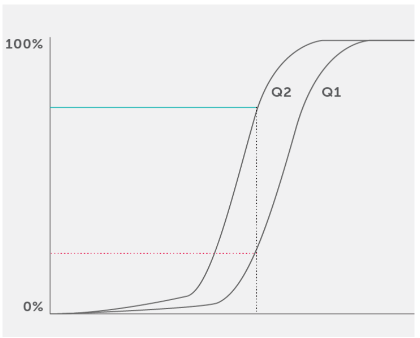
CONTENT INSIGHTS

Content creators can see exactly how well different lessons work for different students, and improve poor performing or insufficient content.

Knewton data scientists are spread out throughout the entire company



Knowledge representation



Core Models e.g., proficiency estimation

Recommendations

Student Analytics

Content Insights

Content Insights

currently being built by yours truly

Duration	How long do students work on this?
Difficulty	How hard is this content?
Assessment	How well do questions assess underlying topics?
Exhaustion	Do I have enough content?
Instruction	How does content contribute to learning?
Engagement	Does some content make students lose interest?

investigated

KNEWTON

Predicting Student Responses

What would Ann, Ben and Charles respond to the last question?

	Q1	Q2	Q3	Q4	Q5
Ann	✓		✓	✓	
Ben		✗	✓	✗	
Charles	✗	✗		✓	
Dan					
Emma					
Frank					

What would Ann, Ben and Charles respond to the last question?

	Q1	Q2	Q3	Q4	Q5	
Ann	✓		✓	✓	Y	100% – *excellent*
Ben		✗	✓	✗	N	33%
Charles	✗	✗		✓	N	33%
Dan						
Emma						
Frank						

Method #1. Look at mean student scores
High-performing students will likely answer correctly, low-performing students will answer incorrectly

What would Ann, Ben and Charles respond to the last question?

	Q1	Q2	Q3	Q4	Q5	
Ann	✓		✓	✓		100% – *excellent*
Ben		✗	✓	✗		33%
Charles	✗	✗		✓		33%
Dan					✓	
Emma					✓	
Frank					✗	
					33%	

Method #2. Classical Test Theory
Also look at question scores. A student is likely to respond correctly if the student's score is higher.

What would Ann, Ben and Charles respond to the last question?

	Q1	Q2	Q3	Q4	Q5	
Ann	✓		✓	✓	Y	100% – *excellent*
Ben		✗	✓	✗	Y	33%
Charles	✗	✗		✓	Y	33%
Dan					✓	
Emma					✓	
Frank					✗	
					33%	

Method #2. Classical Test Theory
Also look at question scores. A student is likely to respond correctly if the student's score is higher.

What would Ann, Ben and Charles respond to the last question?

	Q1	Q2	Q3	Q4	Q5	
Ann	✓		✓	✓		100% – *excellent*
Ben		✗	✓	✗		33%
Charles	✗	✗		✓		33%
Dan		✗	✓	✓	✓	
Emma	✓	✗	✓	✓	✓	
Frank		✗	✓	✗	✗	
		hard	***easy***			

Method #3. Item Response Theory (IRT)
Give more ‘points’ for hard questions

What would Ann, Ben and Charles respond to the last question?

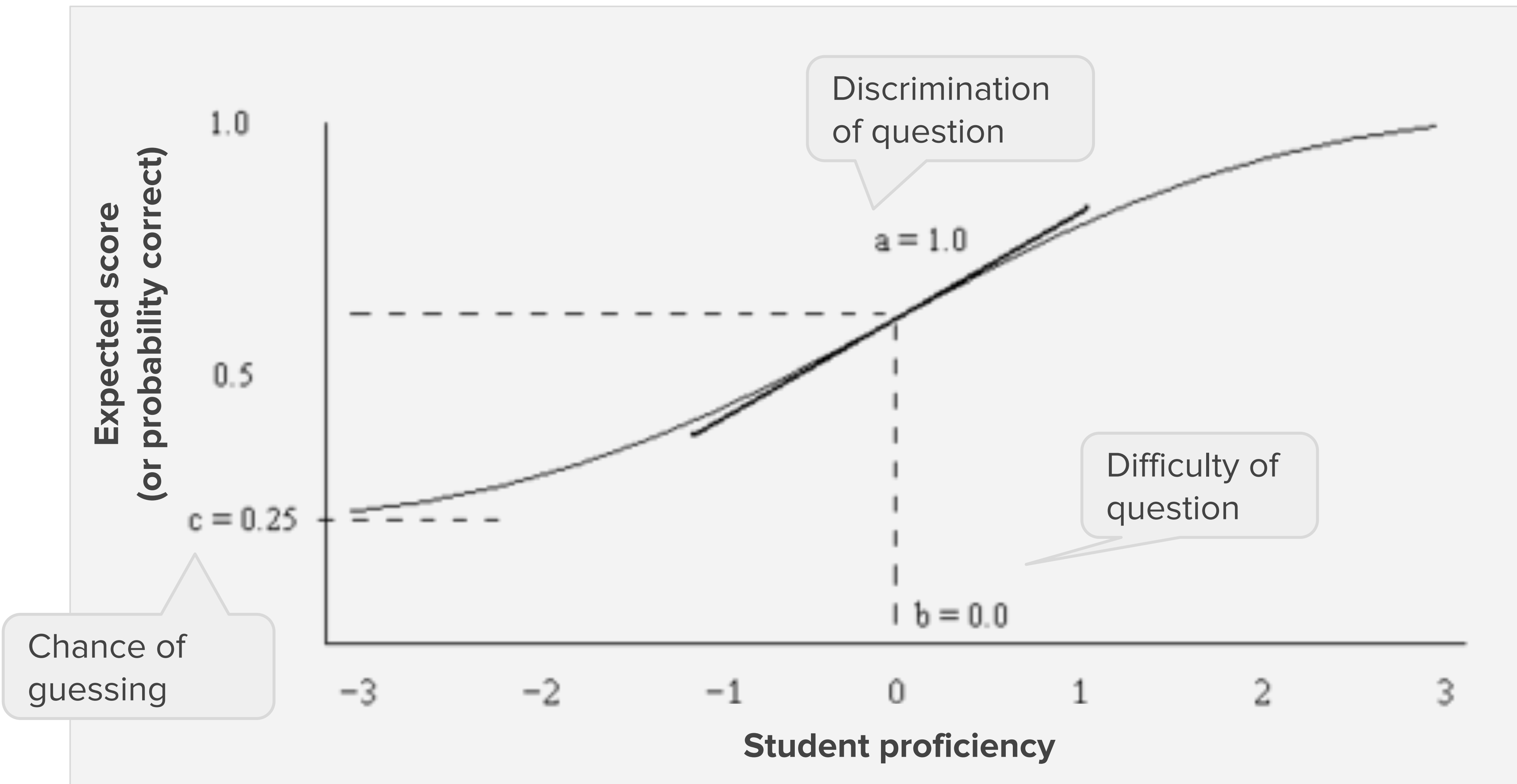
	Q1	Q2	Q3	Q4	Q5	
Ann	✓		✓	✓	Y	100% – *excellent*
Ben		✗	✓	✗	N	33%
Charles	✗	✗		✓	Y	33%
Dan		✗	✓	✓	✓	
Emma	✓	✗	✓	✓	✓	
Frank		✗	✓	✗	✗	
		hard	***easy***	***similar***		

Method #3. Item Response Theory (IRT)
Give more ‘points’ for hard questions

Item Response Theory (IRT)

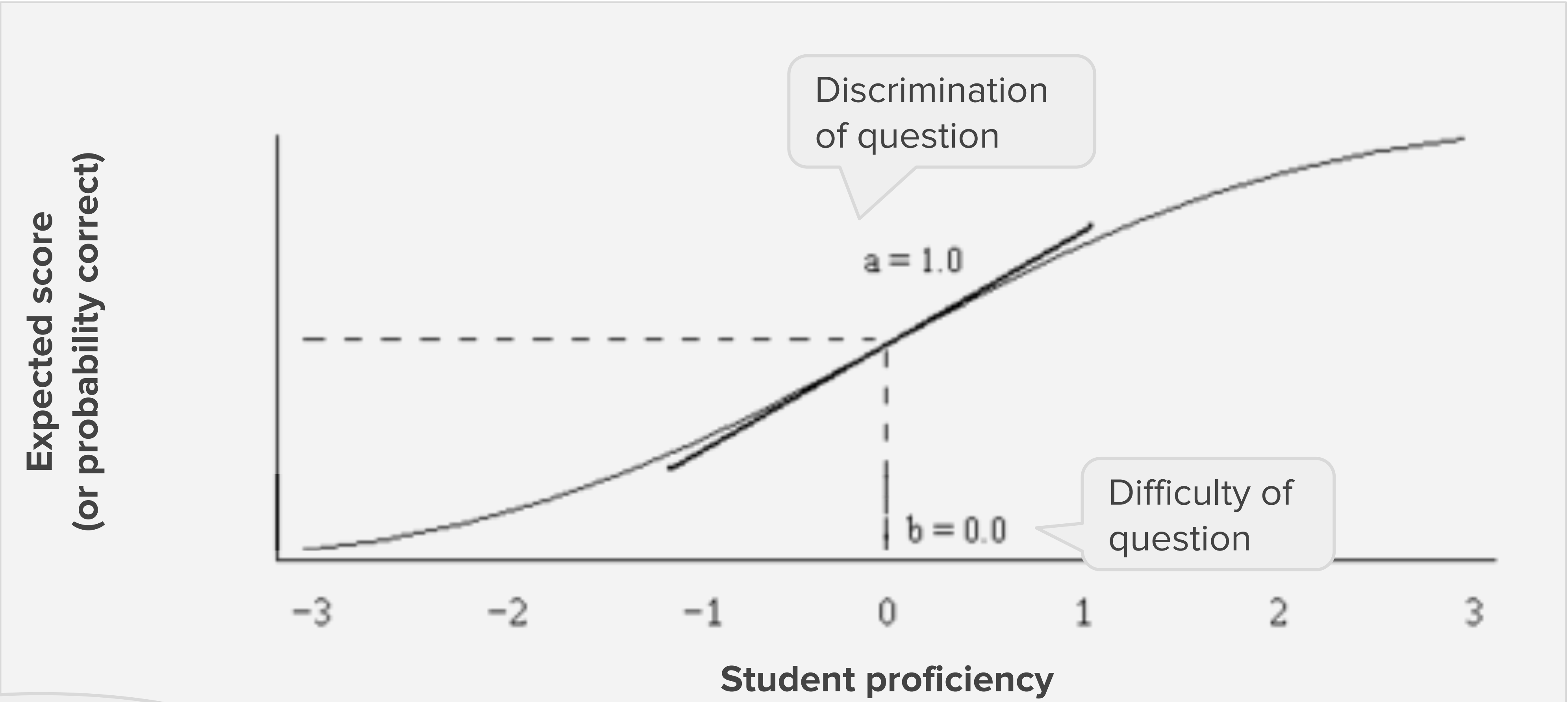
IRT defines student and item parameters

predicted response of student θ on item i $= c_i + \dfrac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$



IRT defines student and item parameters

predicted response of student θ on item i $= \dfrac{1}{1 + e^{-a_i(\theta - b_i)}}$



Assume no correct guessing ($c = 0$)

How do we find

the students' proficiencies

and questions'

difficulty and discriminations?

Bayesian approach for 1PL

(a = 0, c = 0)

$$P(\text{ correct} \mid \theta, \beta\,) = \frac{1}{1 + e^{-(\theta - \beta\,)}}$$

Bayes' rule

$$P(\theta, \beta \mid responses) = \frac{P(responses \mid \theta, \beta) \times P(\theta, \beta)}{P(responses)}$$

Bayes' rule

Likelihood

Prior

$$P(\theta, \beta \mid responses) = \frac{P(responses \mid \theta, \beta) \times P(\theta, \beta)}{P(responses)}$$

Posterior

Bayes' rule

Likelihood

Prior

$$P(\theta, \beta \mid r) = \frac{P(r \mid \theta, \beta) \times P(\theta, \beta)}{P(r)}$$

Posterior

Bayes' rule

$$\max_{\theta,\, \beta} \log P(\theta, \beta \mid r) = \max_{\theta,\, \beta} \log \frac{P(r \mid \theta, \beta) \times P(\theta, \beta)}{P(r)}$$

Bayes' rule

$$\max_{\theta,\beta} \log P(\theta, \beta \mid r) = \max_{\theta,\beta} \log \frac{P(r \mid \theta, \beta) \times P(\theta, \beta)}{P(r)}$$

$$= \max_{\theta,\beta} \log \; P(r \mid \theta, \beta) \times P(\theta, \beta)$$

Bayes' rule

$$\max_{\theta,\,\beta} \log P(\theta, \beta \mid r) = \max_{\theta,\,\beta} \log \frac{P(r \mid \theta, \beta) \times P(\theta, \beta)}{P(r)}$$

$$= \max_{\theta,\,\beta} \log \quad P(r \mid \theta, \beta) \times P(\theta, \beta)$$

$$= \max_{\theta,\,\beta} \Big(\log P(r \mid \theta, \beta) + \log P(\theta, \beta) \Big)$$

Bayesian approach for 1PL

(a = 0, c = 0)

Likelihood

$$P(r \mid \theta, \beta) \quad = \quad \Pi_i \, P(r_i \mid \theta, \beta)$$

Bayesian approach for 1PL

(a = 0, c = 0)

Likelihood

$$P(r \mid \theta, \beta) = \prod_i P(r_i \mid \theta, \beta)$$

$$= \prod_{\text{correct resp.}} P(r = 1 \mid \theta, \beta) \times \prod_{\text{incorrect resp.}} P(r = 0 \mid \theta, \beta)$$

Bayesian approach for 1PL

(a = 0, c = 0)

Likelihood

$$P(r \mid \theta, \beta) = \prod_i P(r_i \mid \theta, \beta)$$

$$= \prod_{\text{correct resp.}} P(r = 1 \mid \theta, \beta) \times \prod_{\text{incorrect resp.}} P(r = 0 \mid \theta, \beta)$$

$$= \prod_{\text{correct resp.}} \frac{1}{1 + e^{-(\theta - \beta)}} \times \prod_{\text{incorrect resp}} \left(1 - \frac{1}{1 + e^{-(\theta - \beta)}} \right)$$

Bayesian approach for 1PL

(a = 0, c = 0)

Likelihood

$$P(r \mid \theta, \beta) = \Pi_i \, P(r_i \mid \theta, \beta)$$

$$= \Pi_{\text{correct resp.}} \, P(\, r = 1 \mid \theta, \beta\,) \times \Pi_{\text{incorrect resp.}} \, P(\, r = 0 \mid \theta, \beta\,)$$

$$= \Pi_{\text{correct resp.}} \, \frac{1}{1 + e^{-(\theta - \beta)}} \times \Pi_{\text{incorrect resp.}} \left(1 - \frac{1}{1 + e^{-(\theta - \beta)}} \right)$$

$$\log P(r \mid \theta, \beta) = \Sigma_{\text{correct resp.}} - \log\left(1 + e^{-(\theta - \beta)}\right) + \Sigma_{\text{incorrect resp.}} \text{ etcetera...}$$

Bayesian approach for 1PL

(a = 0, c = 0)

Gaussian prior

$$P(\theta, \beta) = \prod_{x \in \{\theta, \beta\}} P(x)$$

Bayesian approach for 1PL

(a = 0, c = 0)

Gaussian prior

$$P(\theta, \beta) = \prod_{x \in \{\theta, \beta\}} P(x)$$

$$= \prod_{x \in \{\theta, \beta\}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Bayesian approach for 1PL

(a = 0, c = 0)

Gaussian prior

$$P(\theta, \beta) = \prod_{x \in \{\theta, \beta\}} P(x)$$

$$= \prod_{x \in \{\theta, \beta\}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\log P(\theta, \beta) = \sum_{x \in \{\theta, \beta\}} -\tfrac{1}{2} x^2 \quad + \quad \text{constant}$$

(for standard normal $\mu = 0$, $\sigma = 1$)

Bayesian approach for 1PL – putting it all together

Posterior

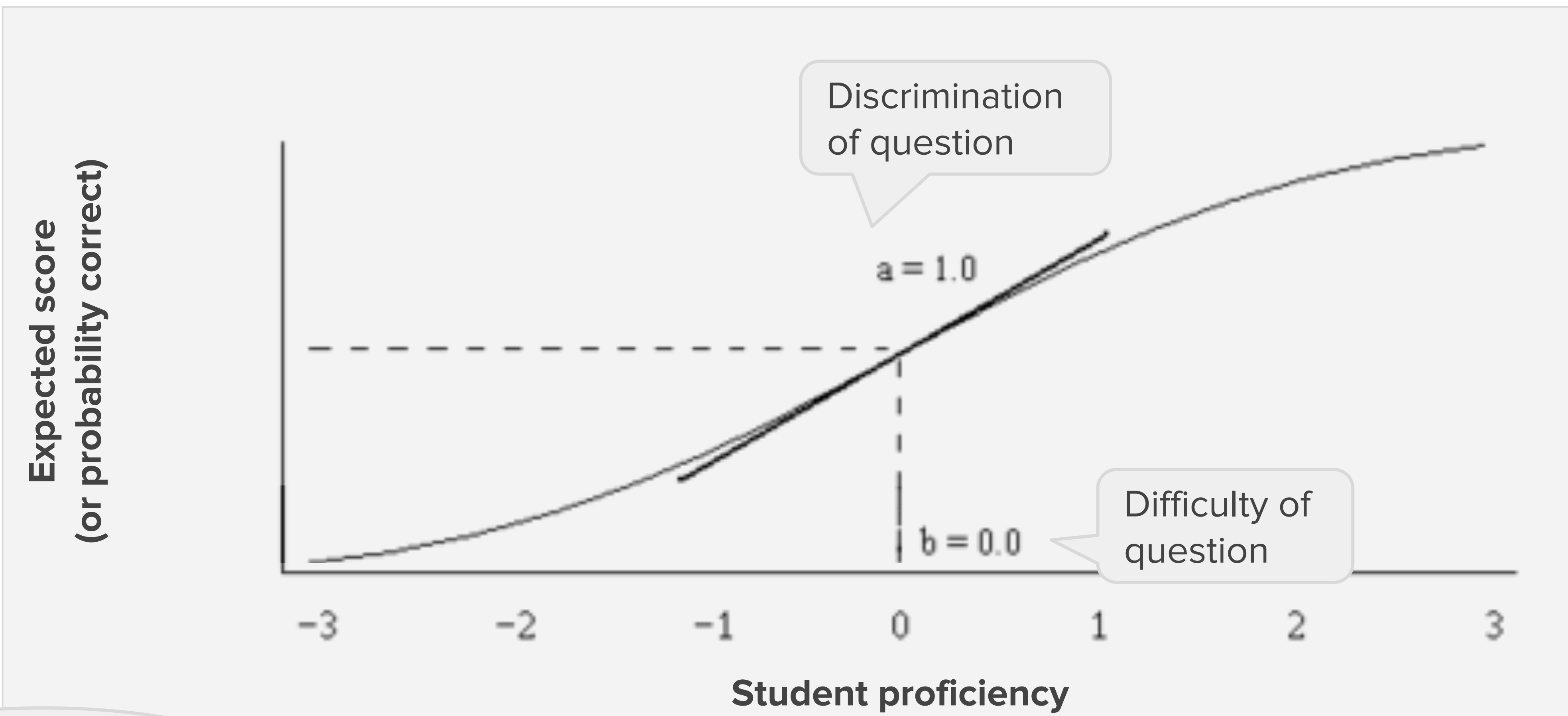
$$\max_{\theta,\,\beta} \log P(\theta, \beta \mid r) \;=\; \max_{\theta,\,\beta} \quad \Sigma_{s,i \in R} \left(\log\left(1 + e^{-(\theta s - \beta i)}\right) + (1 - r_{s,i})(\theta_s - \beta_i) \right)$$

$$- \quad \Sigma_s \tfrac{1}{2}\, \theta_s^{\,2} - \Sigma_i \tfrac{1}{2}\, \beta_i^{\,2}$$

Tournament Ranking

IRT defines student and item parameters

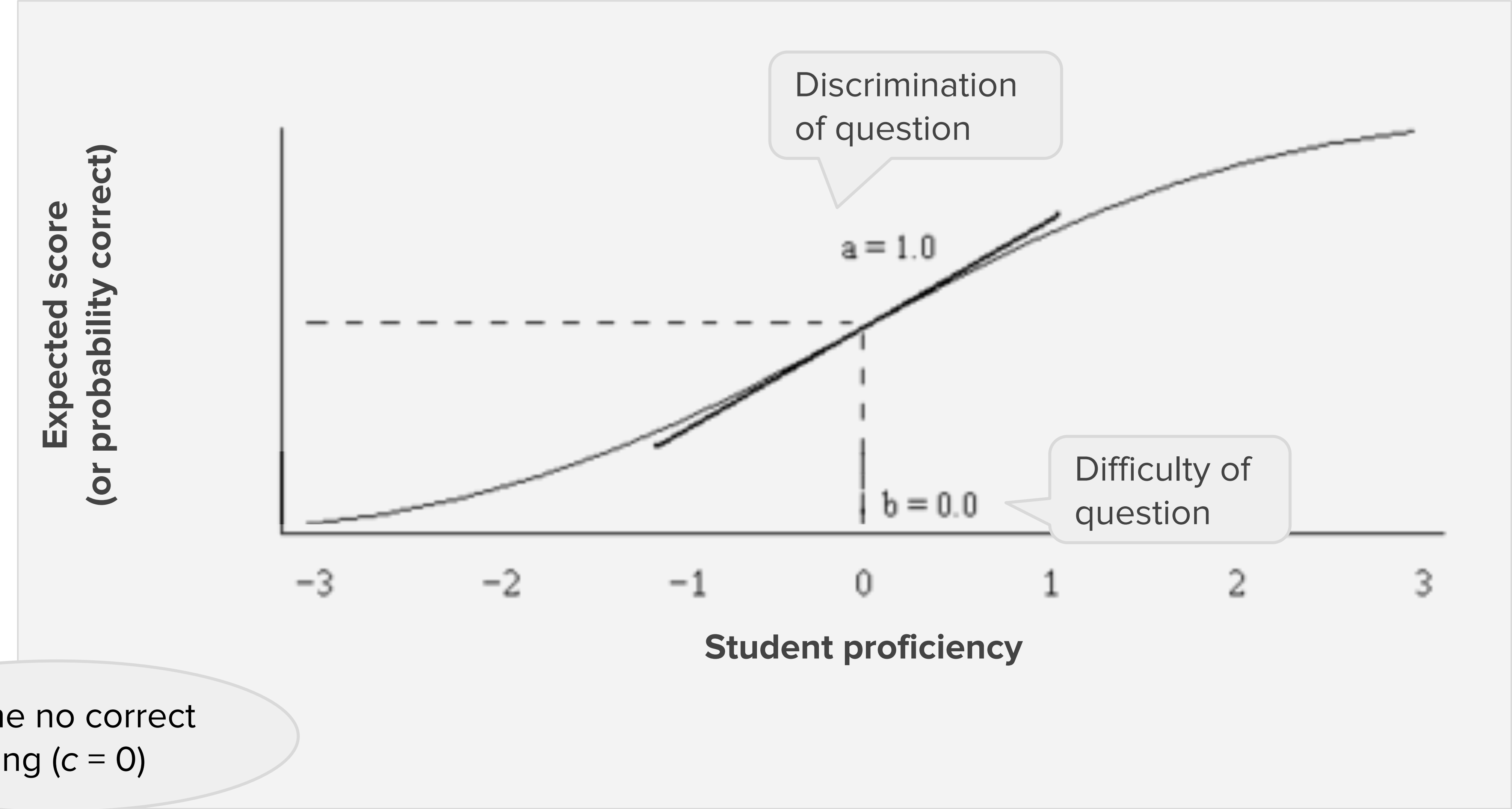
predicted response of student θ on item i $= \dfrac{1}{1 + e^{-a_i(\theta - b_i)}}$



Assume no correct guessing ($c = 0$)

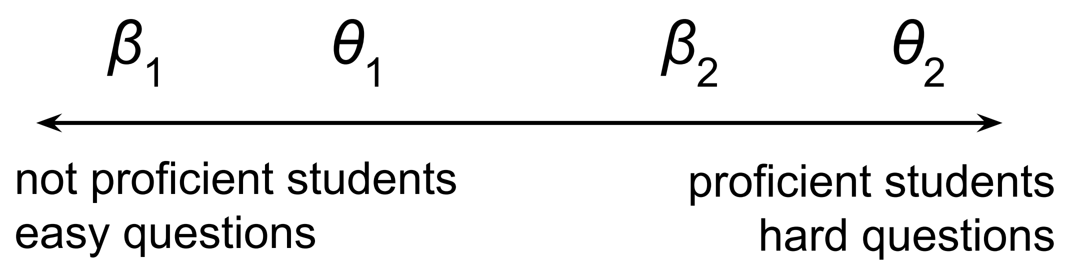
… but when predicting responses, we only look at which one is bigger: *θ* or *β*

predicted response of student θ on item i $= \dfrac{1}{1+e^{-a_i(\theta-b_i)}} = \begin{cases} 1 \text{ if } \theta > \beta \\ 0 \text{ if } \theta < \beta \end{cases}$



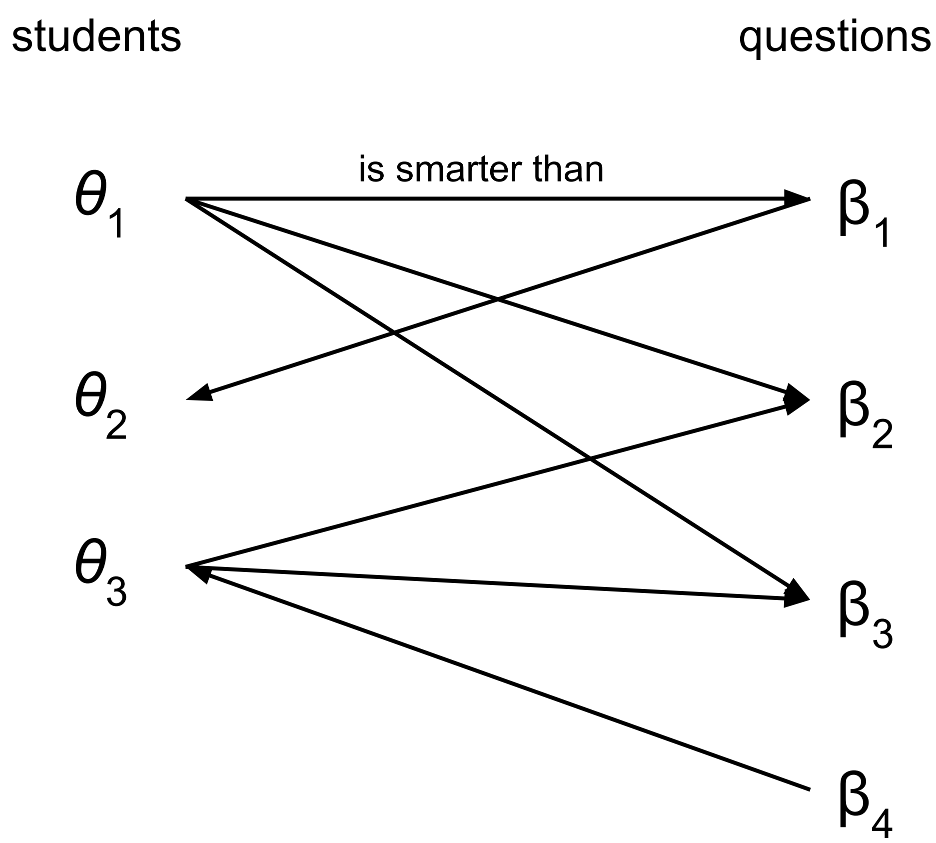
Assume no correct guessing ($c = 0$)

Hence, the highest accuracy is obtained by the optimal ordering of θs and βs



What ordering has the highest accuracy?
How many proficient students answered easy questions incorrectly?
How many non proficient students answered easy questions incorrectly?

…which is equivalent of finding the topological sorting of a bipartite digraph



Graph Theoretic Background

Responses form a directed graph G = (V,E).

Vertices: students and items

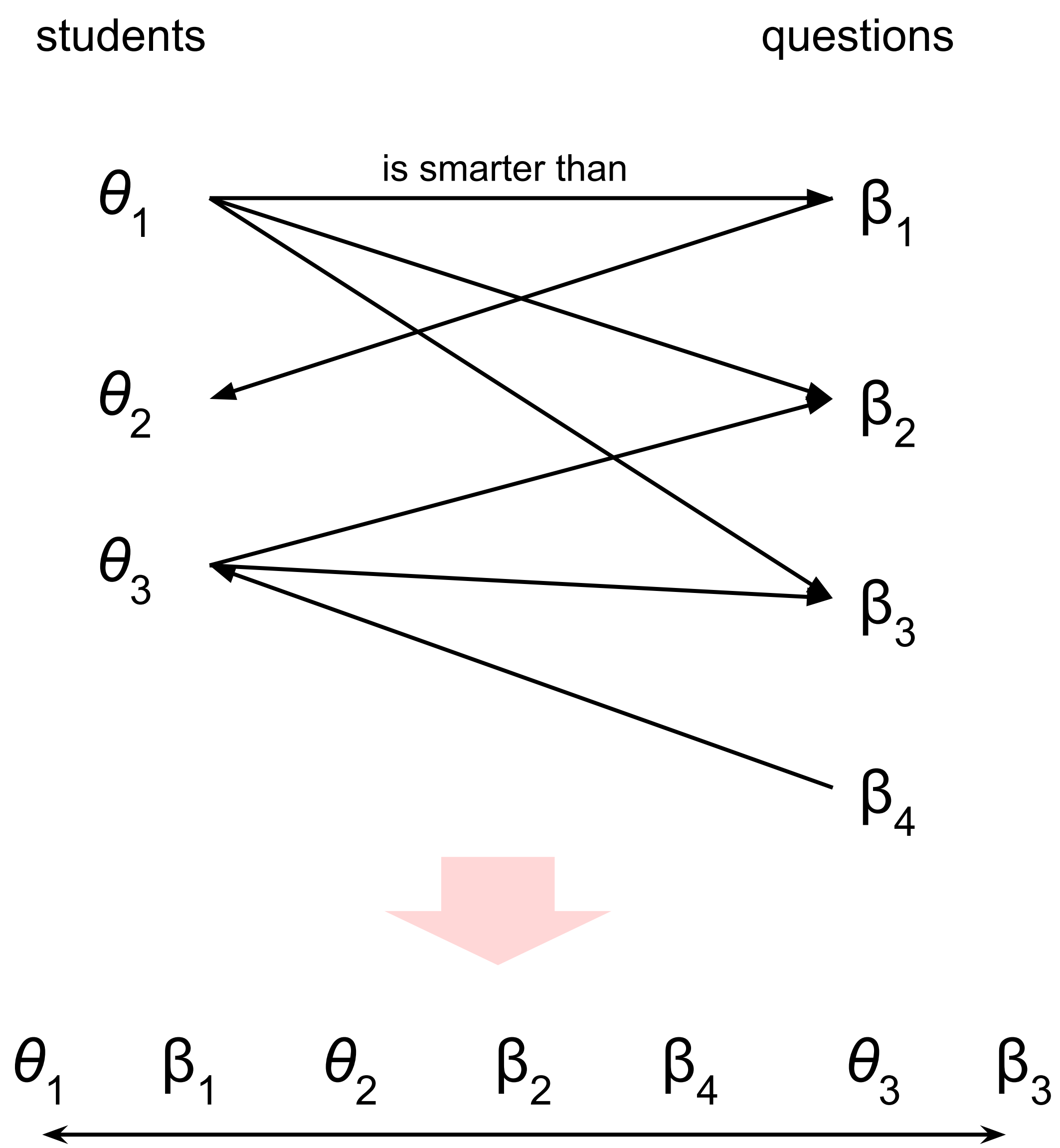
Edges: s → i if correct response,
i → s if incorrect response.

Note that G is

- bipartite (only even cycles), and
- simple (no loops; 1 edge per vertex pair).

Goal: find topological ordering of G
(which only exists if G is acyclic)

...which is equivalent of finding the topological sorting of a bipartite digraph



Graph Theoretic Background

Responses form a directed graph G = (V,E).

Vertices: students and items

Edges: s ➔ i if correct response,
i ➔ s if incorrect response.

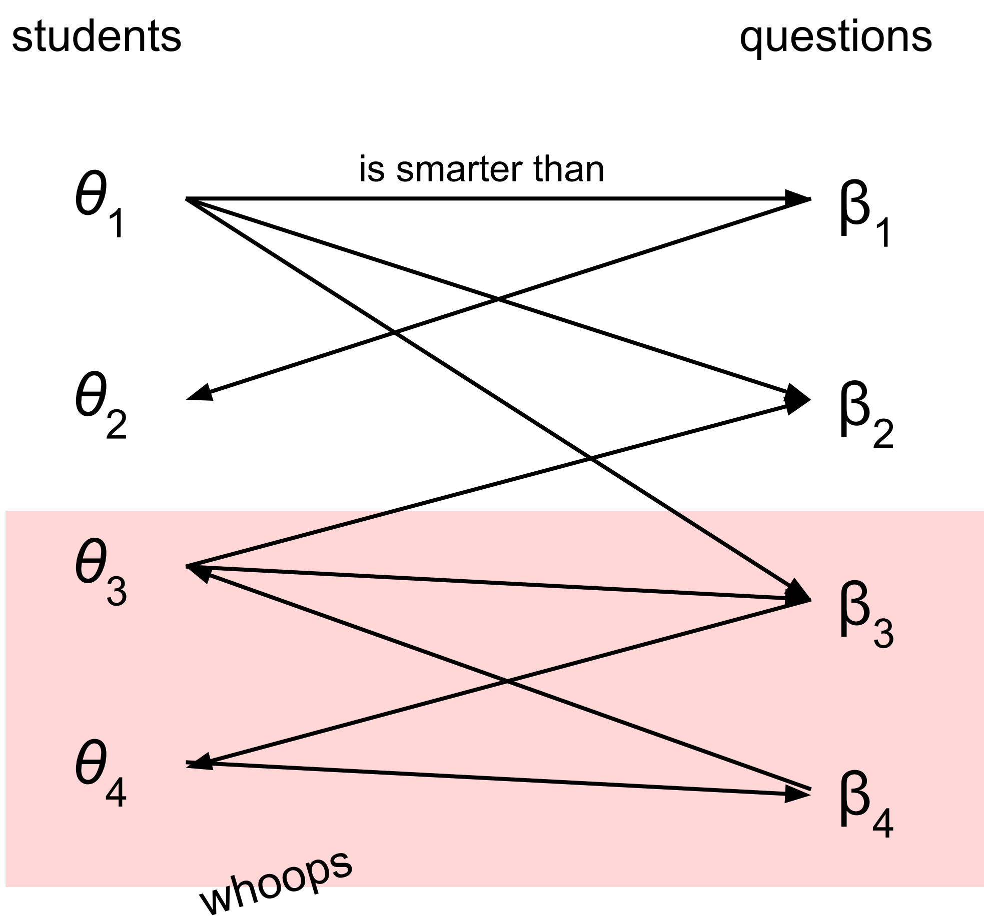
Note that G is

- bipartite (only even cycles), and
- simple (no loops; 1 edge per vertex pair).

Goal: find topological ordering of G
(which only exists if G is acyclic)

Sorting generally not unique

But actual student data often don't fit this model, and cycles occur



Graph Theoretic Background

Responses form a directed graph G = (V,E).

Vertices: students and items

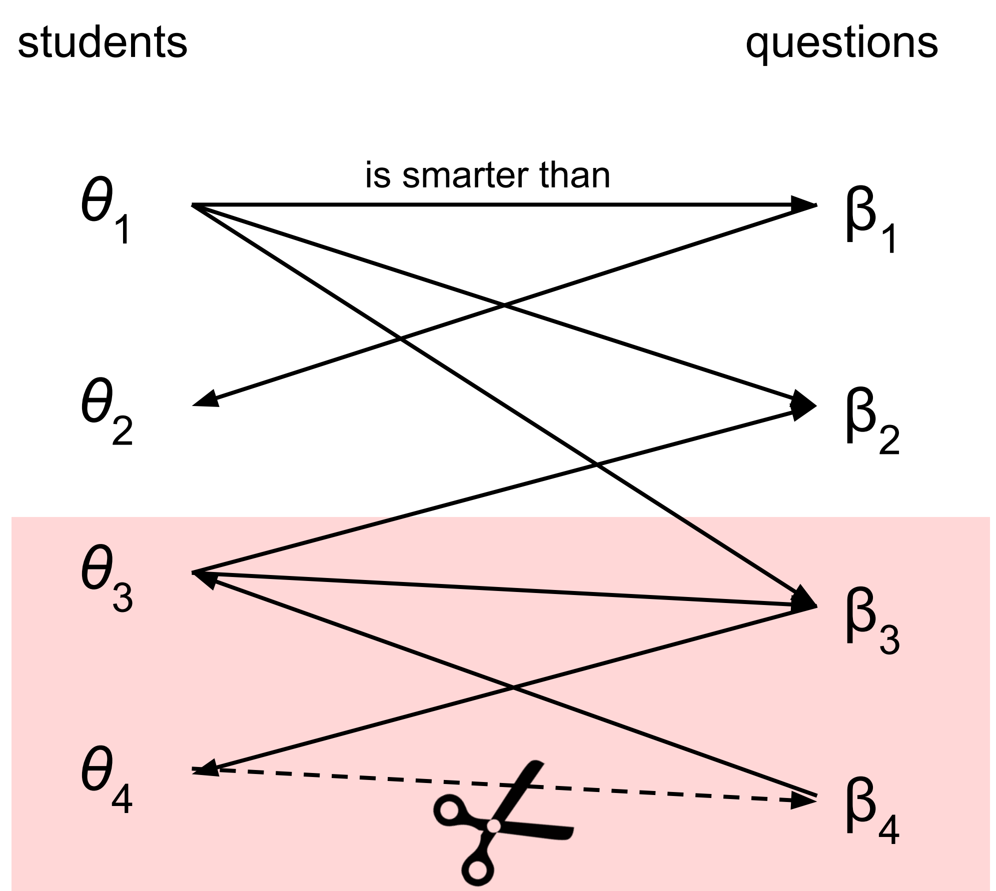
Edges: s → i if correct response,
i → s if incorrect response.

Note that G is

- bipartite (only even cycles), and
- simple (no loops; 1 edge per vertex pair).

Goal: find topological ordering of G
(which only exists if G is acyclic)

Cutting cycles gives us the max IRT (i.e., cutting the min. feedback arc set)



Graph Theoretic Background

Responses form a directed graph G = (V,E).

Vertices: students and items

Edges: s → i if correct response,
i → s if incorrect response.

Note that G is

- bipartite (only even cycles), and
- simple (no loops; 1 edge per vertex pair).

Goal: find topological ordering of G
(which only exists if G is acyclic)

NP-complete :(

Feedback arc set is the smallest set of edges that needs to be removed to make G acyclic

Experiments and Results

Experiments and results: further reading

Popular blog post
published on Knewton's tech blog

(^n_k) N choose K

THE KNEWTON DEVELOPER BLOG

Search

How Well Can You Predict Student Responses?

Posted on January 20, 2015 by Ruben Naeff . View the latest posts.
Like what you see? We're hiring. Check out the career opportunities at Knewton.

NIPS 2014 and Human Propelled Machine Learning

Last month Knewton gathered along with the global data science community at NIPS, which has become one of the top conferences for cutting-edge machine learning. Topics included game theory, climate informatics, energy infrastructure, neural nets and deep learning, and a wealth of different optimization algorithms.

The workshop *Human Propelled Machine Learning* was of particular interest, as much of the research presented is directly relevant to online learning. Jacob Whitehill (HarvardX) presented a model predicting student dropout in online courses, largely based on log-in times. Jonathan Huang (Google) discussed how to automatically grade programming homework on Coursera by clustering different solutions into equivalence classes. Richard

Follow @knewton
Visit knewton.com
Open Source contributions
Jobs at Knewton
Subscribe to RSS feed

CATEGORIES

- data visualization
- deployment
- distributed systems
- graphs
- machine learning
- modeling
- programming languages
- quality assurance
- Uncategorized

Please read on http://www.knewton.com/tech/blog/2015/01/

Academic paper
presented at NIPS 2014 conference

On the Limits of Psychometric Testing in Online Education

Ruben Naeff, Zachary Nichols
Knewton, Inc.
100 5th Ave, 8th Floor, New York NY
{ruben, zack, data-science}@knewton.com

Abstract

The rapid growth of web-based educational products has garnered renewed interest into psychometric assessment methods. Here we compare two traditional psychometric approaches – classical test theory (CTT) and item response theory (IRT) – that model student proficiency and item difficulty, and introduce a graph-theoretical algorithm based on tournament ranking that approximates an upper bound to both traditional models' single interaction prediction accuracy. We apply each model to two large dataset collections, each containing tens of millions of student interactions, and find via cross-validation that CTT is more susceptible to over-fitting than IRT. We also find via the graph theoretical method that for this data, the prediction accuracy of models with fixed student and item parameters has an upper bound of 93% on average (and as low as 82% for some datasets). This is significant, given the reported prediction accuracy for IRT on web-based data [1]. We also find that the upper bound on prediction accuracy can be well-fit by a simple model from basic statistics of the sample (such as mean correctness, number of students, or number of questions answered per student), suggesting that the maximum accuracy is a feature of student behavior that itself can be described by or utilized in future assessment methods.

Please download from http://dsp.rice.edu/HumanPropelledML_NIPS2014

Thank you

Ruben Naeff

Adaptive Products, Data Science

Knewton, Inc.

ruben@knewton.com

@knewton