



The Online Revolution: Education at Scale

Daphne Koller & Andrew Ng
Stanford University & Coursera

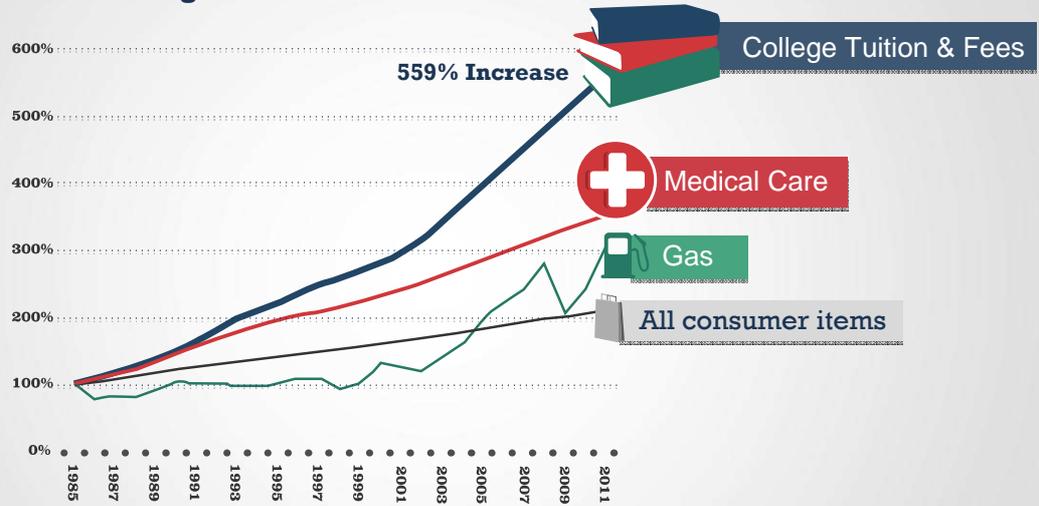


South Africa
Univ. of Johannesburg



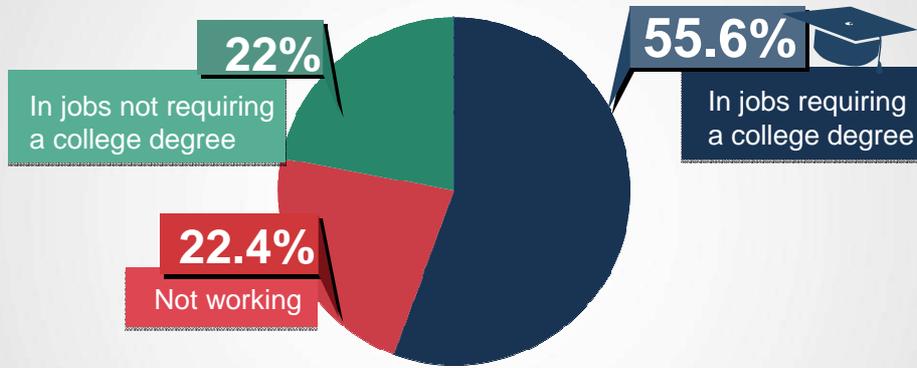
Availability

Price Changes Since 1985



Source: Bureau of Labor Statistics

Affordability



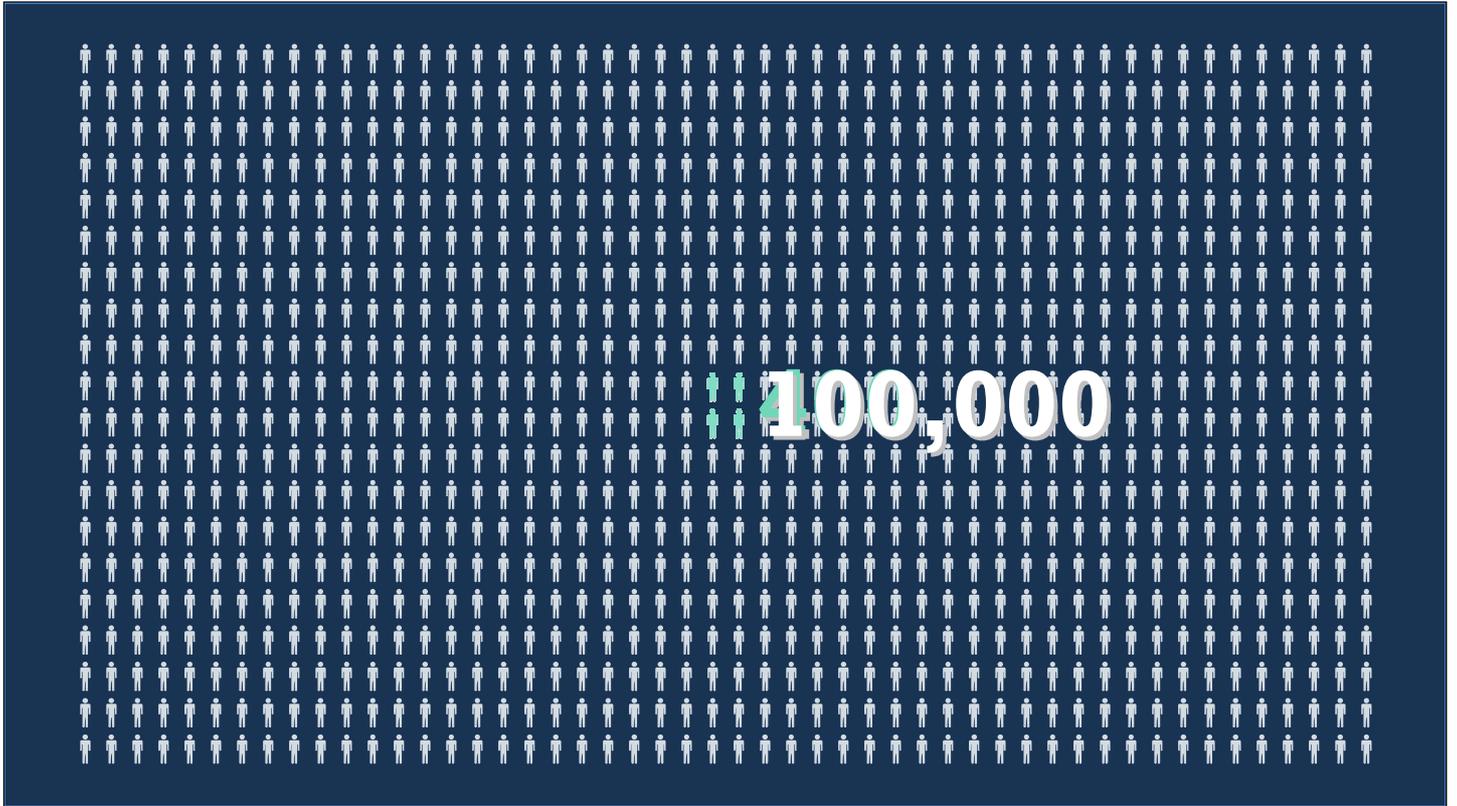
Source: "A College Degree, but Not a College Job." (A. Sum, *New York Times*, 19 May 2011)

Opportunity

“Big breakthroughs happen when what is suddenly possible meets what is desperately necessary.”

—Thomas Friedman
May 15, 2012 · New York Times

Tom Friedman summarized better than anyone what I'm about to tell you: Big breakthroughs are what happens when what is suddenly possible meets with is desperately necessary. We've talked about the desperately necessary. What about the suddenly possible?



What is suddenly possible is to use technology to offer education at scale. This started with 3 courses that Stanford opened to the world in the fall. For example, the machine learning class was taught by my colleague Andrew Ng. Andrew's ML class is one of the larger at Stanford, with 400 students. Andrew's public machine learning class had an enrollment of over 100K students. So for Andrew to reach the same audience by teaching his Stanford class, he would have to teach it for ... 250 years.

“

Coming from a middle class family from a small town in India. Never had the luck and guidance to reach Stanford for education. Guess what? God has sent the opportunity right across my door step! Heartfelt thanks to the great team and teachers who made this happen!

(Akash Goswami)



”

“

I'm a single mom 39 years old with two college boys... I've been trying to get back to college for my masters and want to learn more about computers. Looking forward to this class!

(Jenny Ramirez)



”



High academic standards

900K students

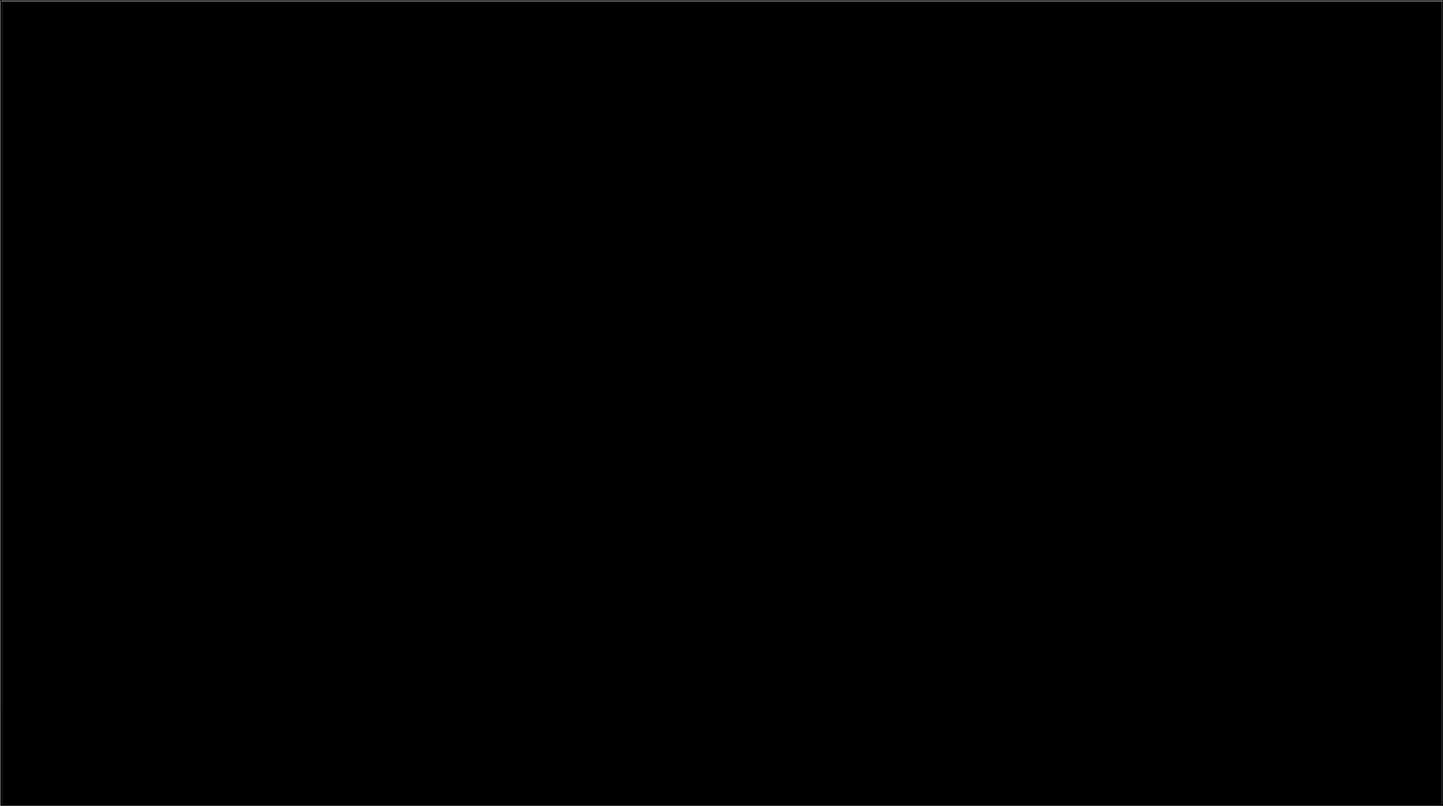
190 countries

2.4 million enrollments

16 Universities

111 courses

It turns out that people like to get great content from the best instructors ... for free. Since mid-February, when our website opened, we have accumulated more than 600K students from 190 countries. We have close to 1.5 million enrollments in 41 courses, across a range of disciplines. In the 15 courses that have already launched, we have 14 million video views and 6 million quiz submissions.

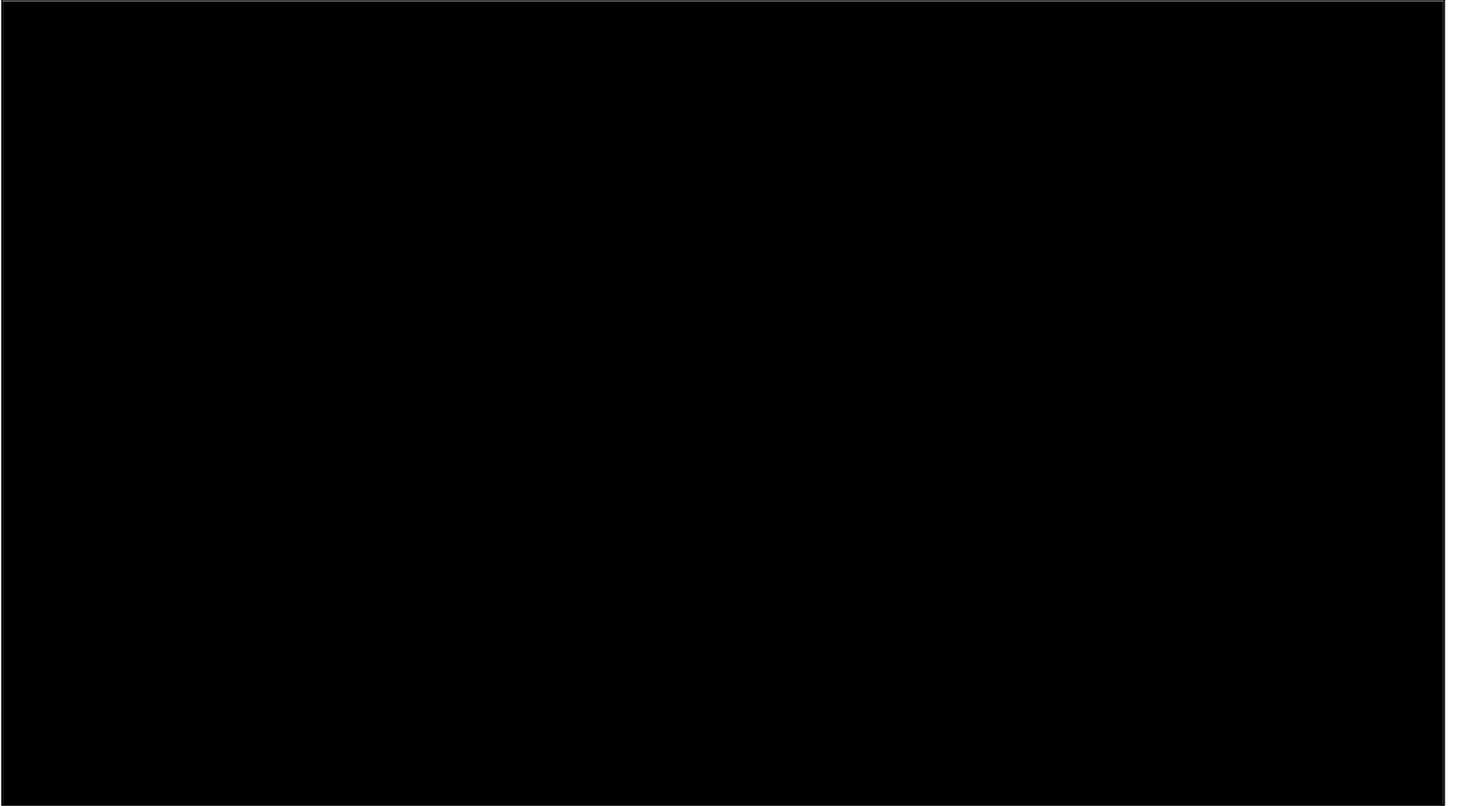




Real Course

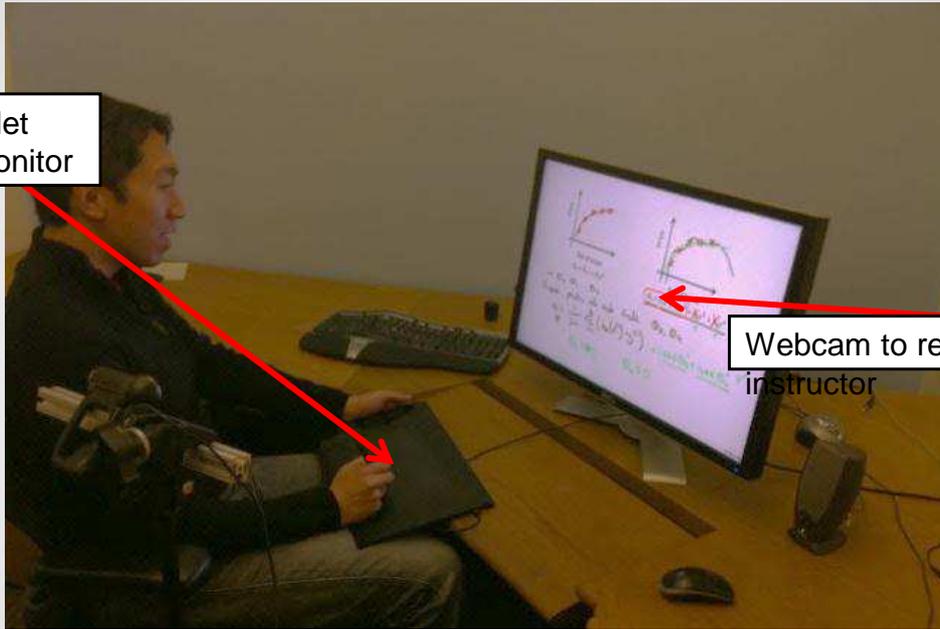
VIDEO-BASED INSTRUCTION

Coursera



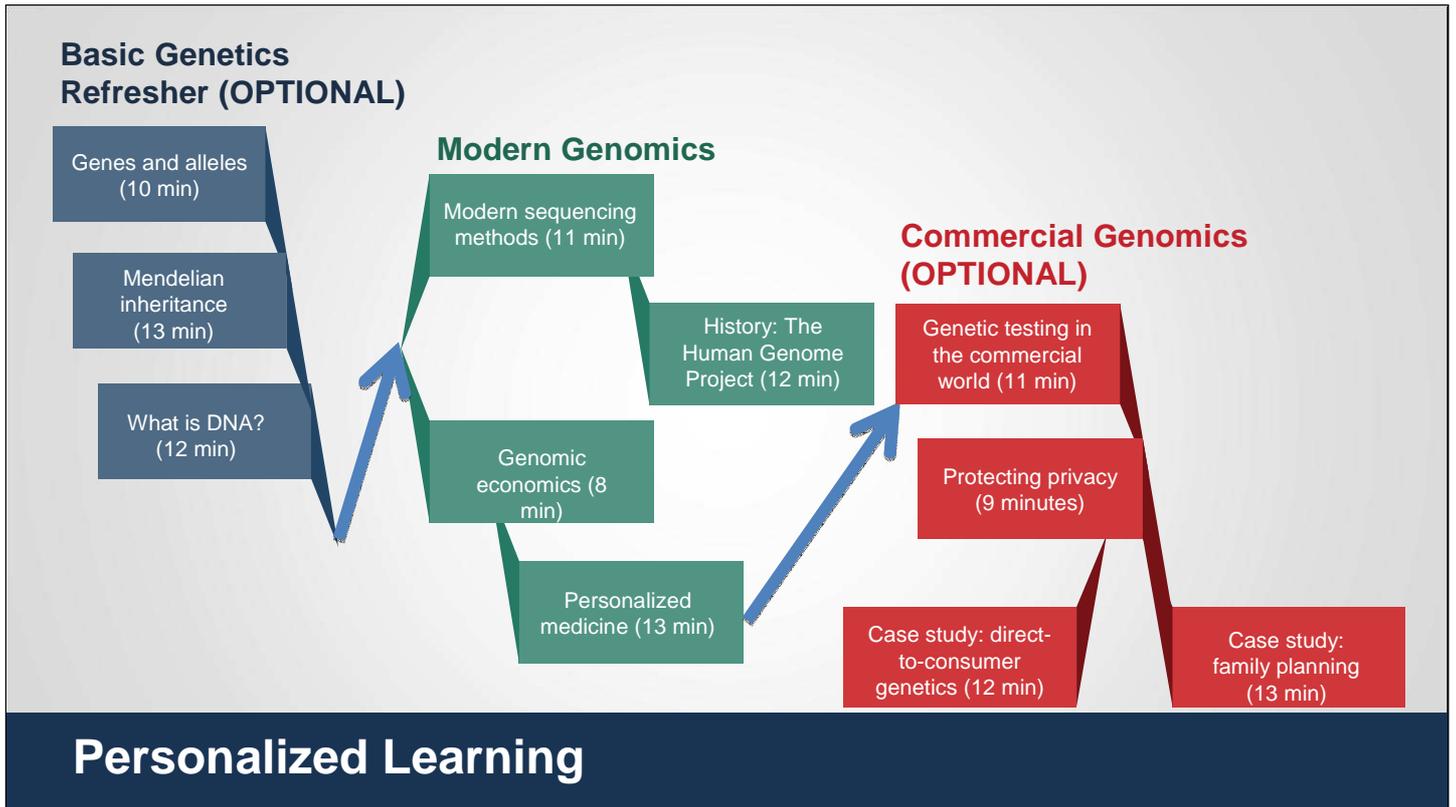
ADD GRAPH?

Writing on tablet
appears on monitor



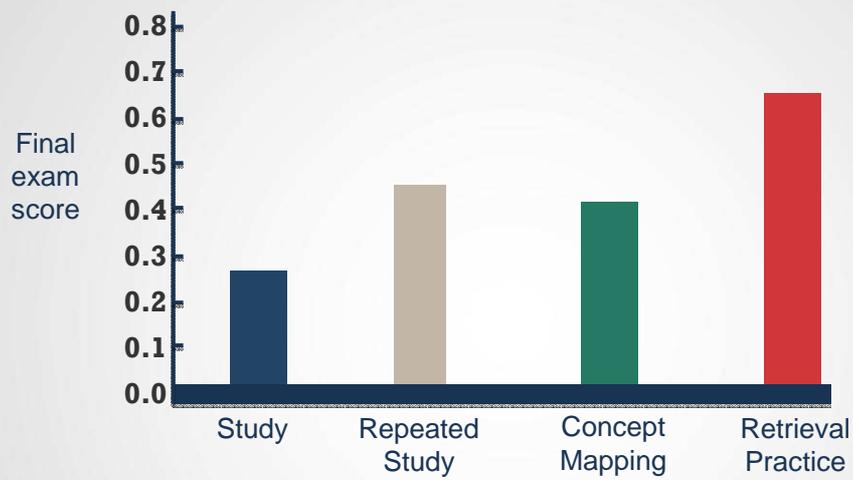
Webcam to record
instructor

Lecture Recording



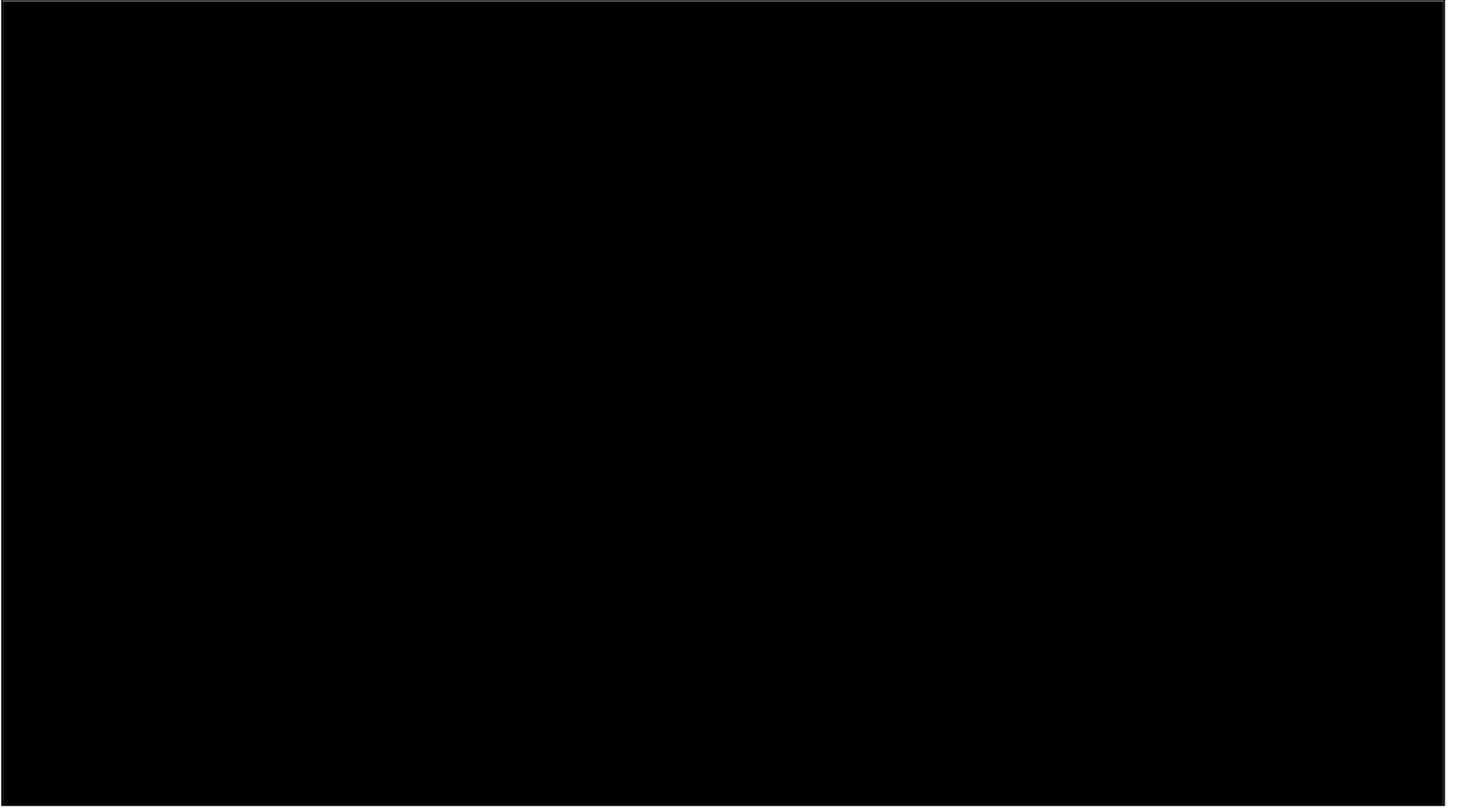
Let's talk about some of the key elements that went into these courses. First, when we design content for the online environment, we can remove the constraints of classroom scheduling, and abandon the monolithic 1hr lecture. The content can be divided into short, coherent modules of 8-12 minutes each, which can be traversed in different ways by different students. Some students might benefit from some additional preparation. Others might want to go into depth in certain topics. So we can finally move away from the one-size-fits-all model of education into a much more personalized experience.

ASSESSMENTS



"Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping."
J. Karpicke, J. Blunt. *Science* (2011).

Testing Improves Learning



- Multiple-choice (radio button, checkbox)
- Numerical answers
- Short answer
- Math
- Structured outputs: Programming and modeling assignments
- Peer assessment

Multiple forms of assessment

To show: $e^{e^{e^x+x}}$ Description: Let $a = e^{e^x}$. $a' = ?$

Required Rules: Assumptions: $\frac{d}{dx}(e^{e^x})$

Identities	Properties of 0 and 1
Power Rule	$\frac{d}{dx}(x^n) = n \cdot x^{n-1}$
Derivative	$\frac{d}{dx}(A+B) = \frac{d}{dx}(A) + \frac{d}{dx}(B)$, $\frac{d}{dx}(c \cdot A) = c \cdot \frac{d}{dx}(A)$
Chain Rule	$\frac{d}{dx}(e^{f(x)}) = e^{f(x)} \cdot \frac{d}{dx}(f(x))$
Exponents	$A^X \cdot A^Y = A^{X+Y}$

Free Rules:

Commutative	$A+B = B+A$, $A \cdot B = B \cdot A$
Associative	$(A+B)+C = A+(B+C)$, $(A \cdot B) \cdot C = A \cdot (B \cdot C)$
Negative and Divide	

Proof:

$e^{e^x} \cdot \frac{d}{dx}(e^x)$	Chain Rule	$\frac{d}{dx}(e^{e^x})$	✕
$e^{e^x} \cdot (e^x)^n \cdot \frac{d}{dx}(e^x)$	Chain Rule	$e^{e^x} \cdot \frac{d}{dx}(e^x)$	✕
$e^{e^x} \cdot (e^x)^1$	Power Rule	$e^{e^x} \cdot \frac{d}{dx}(e^x)$	✕
$e^{e^x} \cdot (e^x)$	Identities	$e^{e^x} \cdot (e^x)$	✕
e^{e^x+x}	Exponents	$e^{e^x} \cdot (e^x)$	✕

Update Proof Clear Proof

✔ You've finished this assignment.

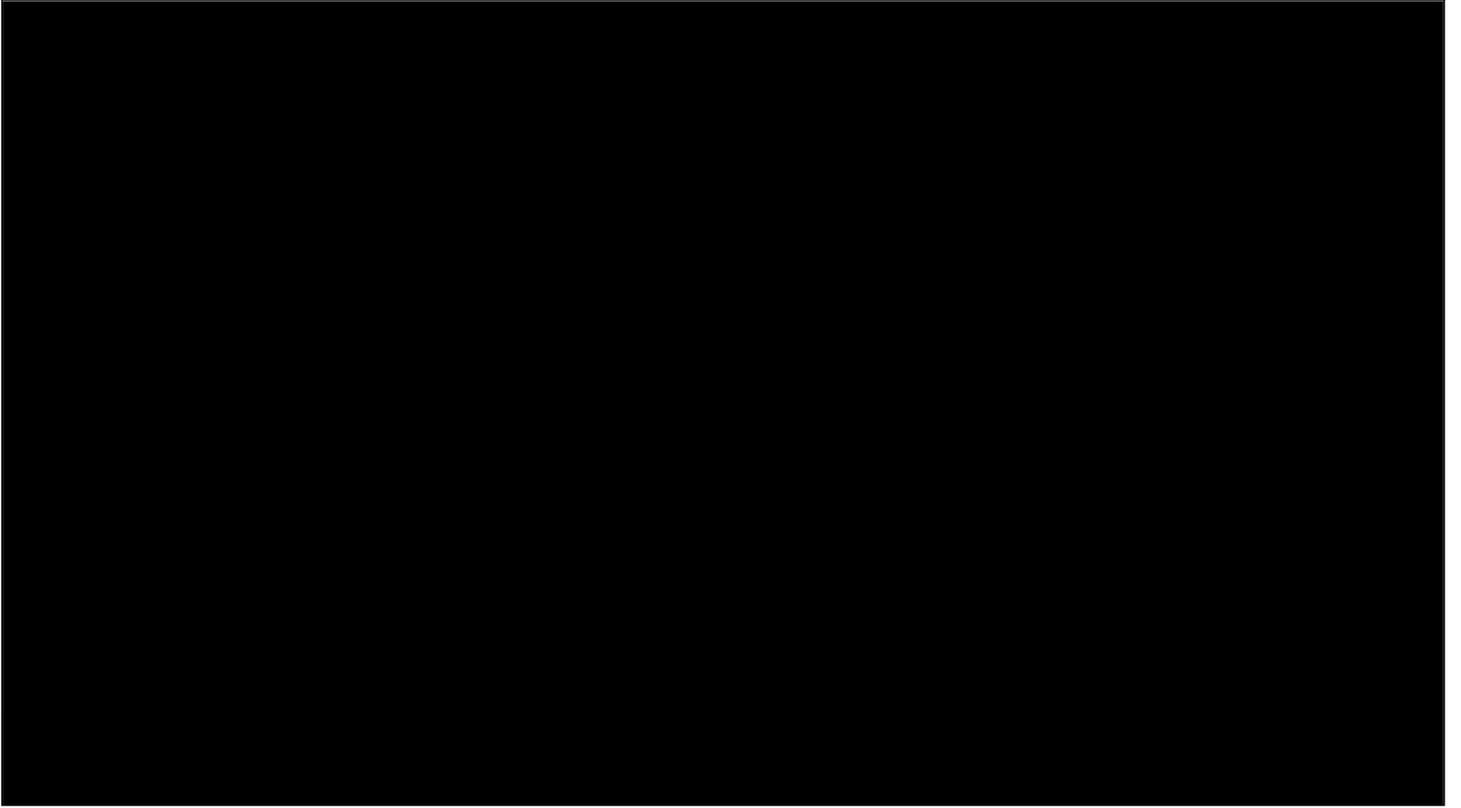
Colleen Lee, Ethan Fast, Alex Aiken

Math assessments

Expression equivalence (e.g., $a^2 - b^2 = (a+b)(a-b)$..)
Theorem proving

- Programming assignments
 - Support for any language: R, Python, Matlab, Java, ...
 - Advanced assessments: timing, memory use, test set performance, ...
 - Spell-check, Kinect™ action recognition
- Modeling assignments (e.g., electrical circuits)
- Excel spreadsheets (data analysis, financial models, ...)

Grading structured outputs



Question 8

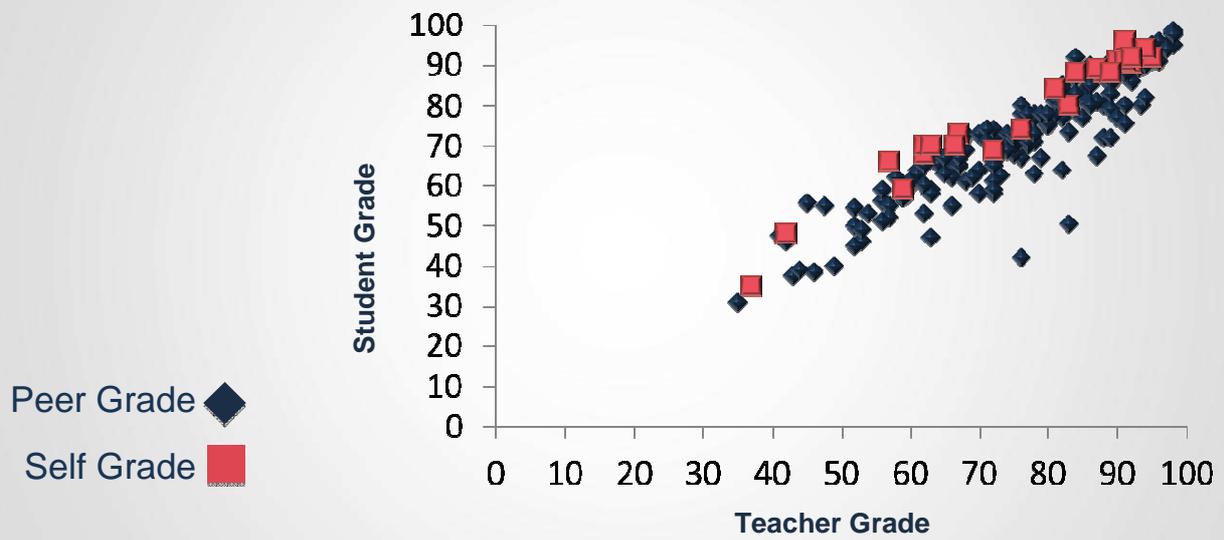
I-maps. I-maps can also be defined directly on graphs as follows. Let $I(G)$ be the set of independencies encoded by a graph G . Then G_1 is an I-map for G_2 if $I(G_1) \subseteq I(G_2)$.

Which of the following statements about I-maps are true? You may select 1 or more options (or none of them, if you think none apply).

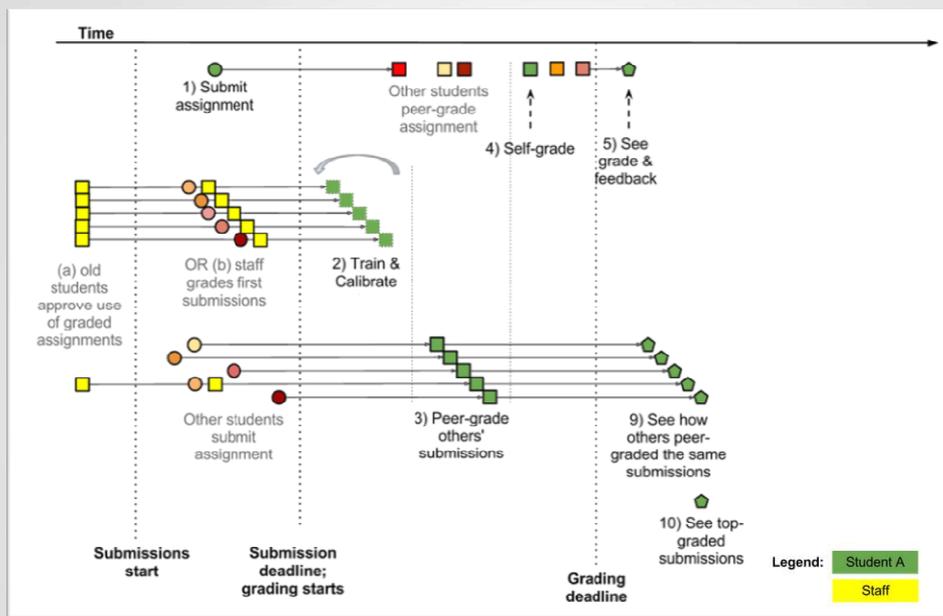
- A graph K is an I-map for a graph G if and only if K encodes exactly the same independencies as G .
- I-maps are Apple's answer to Google Maps.
- An I-map is a function f that maps a graph G to itself, i.e., $f(G) = G$.
- A graph K is an I-map for a graph G if and only if K and G are identical, i.e., they have exactly the same nodes and edges.
- A graph K is an I-map for a graph G if and only if all of the independencies encoded by K are also encoded by G .

Personalization and Mastery

“The Impact of Self-and Peer-Grading on Student Learning”.
P. Sadler, E. Good. *Educational Assessment* (2006).



Open-Ended Work



Cf. Calibrated Peer Review™
(Chapman, 2001)

Peer Grading Workflow

Evaluation criteria & Grading rubric

Grade value 40 points

Guiding questions

0-15 points

16-20 points

Did you make informal prototypes of two ideas? Points off if the prototype is too formal. (As a rough rule of thumb, a detail-oriented computer mock-up is too formal.) (max 20)

Fewer than 2 prototypes; ineffectual prototypes; unnecessary formality.

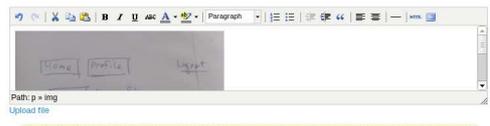
Two prototypes, created rapidly.

Did you test your prototype with at least 5 (3 if the activity is long) users waiting in a real line? (max 20)

0-7: Not really.
8-15: The testing was hasty, and done with your friends or family for the sake of convenience.

Yes. With real users who were waiting in a real line.

Photos of your prototypes



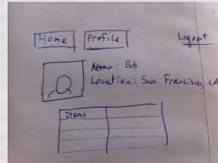
Evaluation

Did the student make informal prototypes of two ideas? Points off if the prototype is too formal. (As a rough rule of thumb, a detail-oriented computer mock-up is too formal.) (max 20)

- 0-15 points: Fewer than 2 prototypes; ineffectual prototypes; unnecessary formality.
- 16-20 points: Two prototypes, created rapidly.

Comments:

Photos of your prototypes



Evaluation

Did the student make informal prototypes of two ideas? Points off if the prototype is too formal. (As a rough rule of thumb, a detail-oriented computer mock-up is too formal.) (max 20)

Aggregate score: 17.5

Comments:

student1: Your prototypes were at the right level of formality.

student2: I'm glad you chose to highlight the navigation buttons and de-emphasized the less important actions.

student3: You clearly put a lot of effort, but the assignment asked a high-level prototype, and your submission had too much detail.

student4: pretty good

student5: I was a bit confused about which parts of your 2nd prototype to focus on. The professor said a good informal prototype doesn't show details for views that don't impact the flow of the UI.

Peer Grading



Sociology 101 Student Map

Global community of learners

^
16
vote(s)
v

In one of the questions, it is suggested that we "Use the unix command line utilities". How do I solve this on a non-unix OS? I have never worked with this, so I am at a complete loss what to do. Thanks in advance

Posted by (Student)
on Wed 7 Mar 2012 6:41:09 PM PST

Add New Comment

Time (Oldest to Newest) Time (Newest to Oldest) Votes (Most to Least)

^
11
vote(s)
v

If you want the true 'Unix experience', running a virtual machine is better than using cygwin / gnuwin, imo.

[Oracle VirtualBox](#) is great, and you can find many preconfigured virtual machine image on [Virtualboxes.org](#). Choose Ubuntu, if you are new to the linux world. That one also has Python 2.7 included, while many other distributions like Debian still have Python 2.6

Open a terminal window by entering 'terminal' in the dashboard. You can also install 'Guake Terminal', then you can always open and close a terminal by pressing F12.

Posted by (Student)
on Thu 8 Mar 2012 4:59:24 AM PST

Add New Comment

^
31
vote(s)
v

Assuming you're using windows, you can

- install cygwin, or
- install gnuwin32, or
- run live linux in virtualbox (e.g. ubuntu live cd)

Posted by (Student)
on Wed 7 Mar 2012 7:40:21 PM PST

Students

Discussions

76
vote(s)

Ordering for assigning factors to cliques in `ComputeInitialPotentials`. The order of assignment of factors to cliques should happen in the order cliques are given to you at the end of the `CreateCliqueTree` function. Each factor should be assigned to the first clique that contains the variables in the factor, where ordering of the cliques is given in `C.nodes` (`C` is the argument for `ComputeInitialPotential` function).

For example: in function `ComputeInitialPotentials`, the argument `C` has a field `nodes`. Now let's say the contents of `C.nodes` are:

```
C.nodes[1] = [1 2]
```

```
C.nodes[2] = [2 3]
```

And your factors are [1], [2], [3]. So [1] and [2] should be assigned to the 1st clique. Even though [2] can be assigned to the second clique, for the purpose of this assignment we are going to assign [2] to the first clique that contains it.

Order of Variables in Cliques. You should use `CliqueTree.nodes()` to get the ordering for your variables and those nodes are in numerical order.

Empty cliques. It is possible that you may end up with cliques with no factors assigned to them. If that is the case, set the initial potential to 1 for all variable assignments for that clique.

`CliqueTreeCalibrate` (for max-sum) If you are having problems with this part, but your code is otherwise correct for sum-product message passing, make sure that your `FactorMaxMarginalization` works properly with `logspace-potentials`.

Clique Potentials If you have a clique over variables [1 2 3] with only one factor assigned to it, say [1], then you should assume that there's an initial potential over [1 2 3] with all 1s and multiply it.

This is the implementation we have and it doesn't really affect the answer because you will end up multiplying stuff about all your variables in the clique. If a variable does not appear in any factor, then it shouldn't be in any clique. However, if a variable is in a clique and just that the factors assigned to the clique don't contain the variable then there must be some other clique that contains that variable, and to which the factor is assigned. By running intersection property that variable has to be in the sepset so you will get messages for it.

Community TA

Posted by  (Community TA)
on Mon 9 Apr 2012 2:36:28 PM PDT

Comments

- ^ Thanks a lot! It clarified several things for me. However, when you mention empty cliques, apart from the initial potential set to 1: what variable (I mean, field `.var`) should we put? Zero? An empty vector []? And the cardinality? Because this has an effect on the amount of values
- ^ (all ones, as you said).

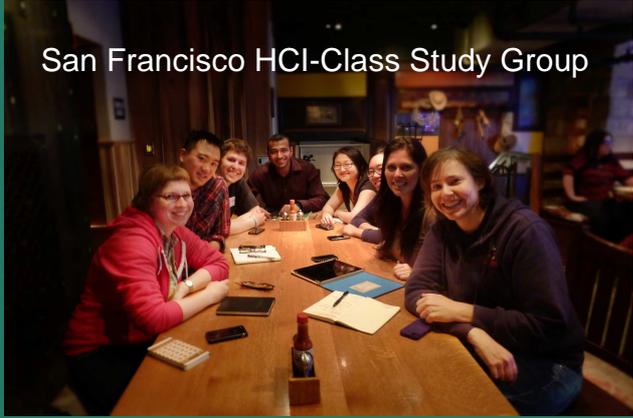
[Delete] Posted by  (Student)

Community TAs



“Fellow students on these forums really gave the sense that I wasn't just sitting in my office working on it by myself. The spirit of cooperation and information sharing has been far more than any “non-virtual” course I ever took. (Sanjaya Kumar)”

San Francisco HCI-Class Study Group



Vietnam

Czechoslovakia

Nigeria Miami

Russian Austin, Texas
speaking students

Minnesota

Guatemala

Arabic speaking students

**A multilingual universal
study group**

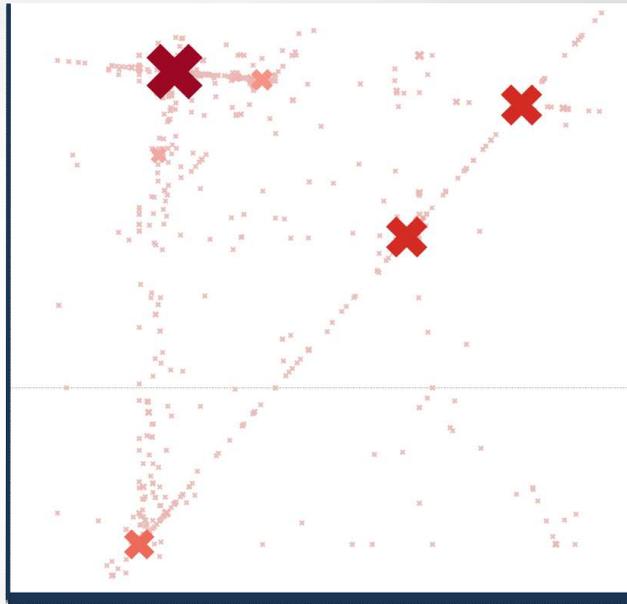
Athens Nepal **Kenya**

Student Study Groups



STATISTICS & ANALYTICS

Wrong answers submitted
for machine learning class
programming assignment



New Window into Human Learning

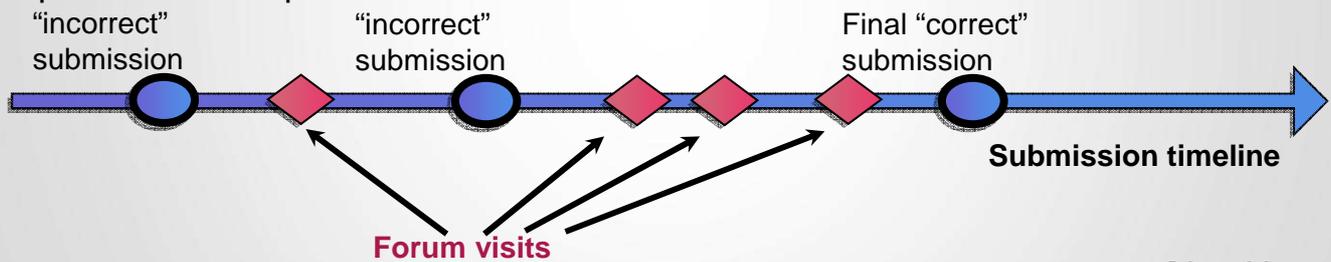
ON HOLD

For each forum question q , estimate:

Increase in probability of transitioning to correct response given that a user viewed question q

(focusing only one "cluster" of incorrect responses at a time)

Experimental setup:



[Jon Huang]

New Window into Human Learning

Question 830: 98 views and 1949 not viewed.

- Fraction transitioned to correct if viewed: 0.643
- Fraction transitioned to correct if not viewed: 0.34
- Fraction improvement by viewing: 1.893

computeCost

12
▲▼

I have obtained ans = 32.073 for the computeCost function in ex1 but when I submit 2) Computing Cost (for one variable) [computeCost.m] I am told == Sorry, your answer was incorrect.

How can this be?

Tags: [exercise 1](#)

Posted by: [redacted]
6 months 2 days ago

Sort By [Relevance](#) | [Date \(Newest First\)](#) | [Date \(Oldest First\)](#) Page 1 of 1

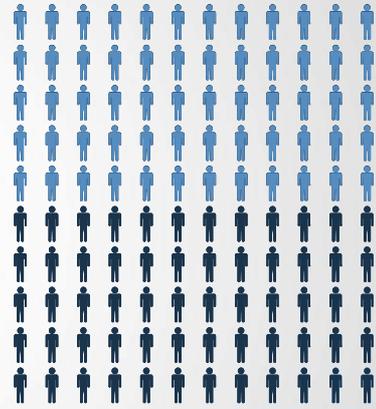
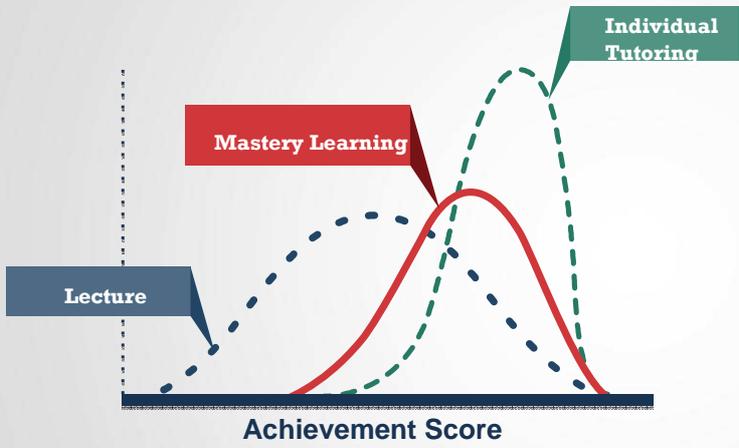
14
▲▼

Something to consider is the matrix X. Remember that a column of 1s are added so this will change the way you are indexing values in X. I had the same answer and was also failing until I considered this.

Posted by: [redacted]
6 months 2 days ago

[Jon Huang]

What helped the most?



"The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring." *B. Bloom, Educational Researcher (1984).*

The 2 Sigma Problem

“

College is a place where a professor's lecture notes go straight to the students' lecture notes, without passing through the brains of either.

”

—Mark Twain

“

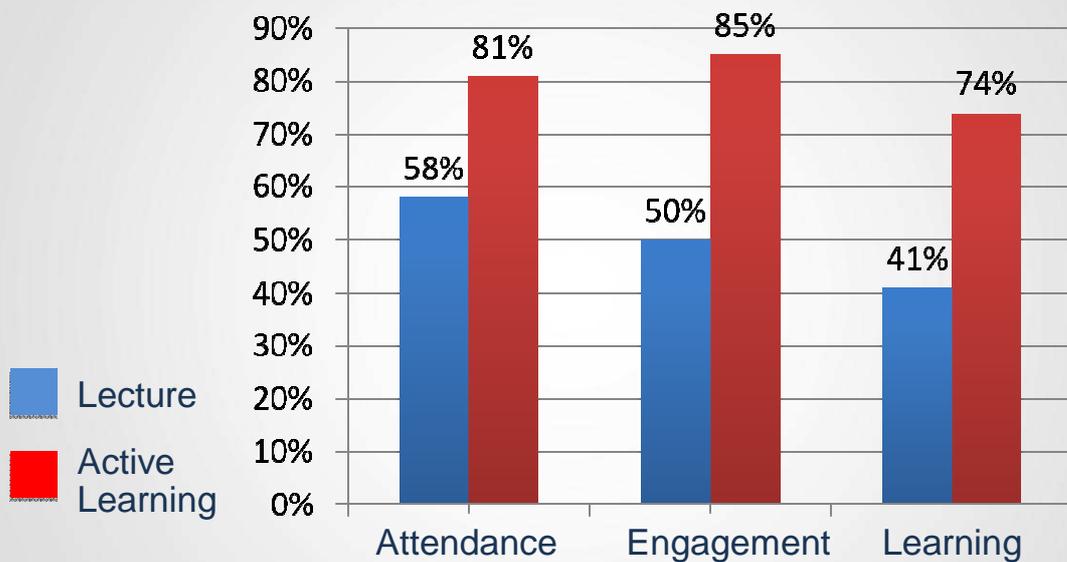
The mind is not a vessel that needs filling, but wood that needs igniting.

”

—Plutarch

from Ian Kidd's translation of Essays

"Improved Learning in a Large-Enrollment Physics Class."
L. Deslauriers, E. Schelew, and C. Wieman. *Science* (2011).



Igniting Minds

- Just-in-time teaching

- Small group problem solving

- Higher-level discussion of material

- Presentation of real-world applications

- Optional interactive sessions better attended than lectures in standard televised courses

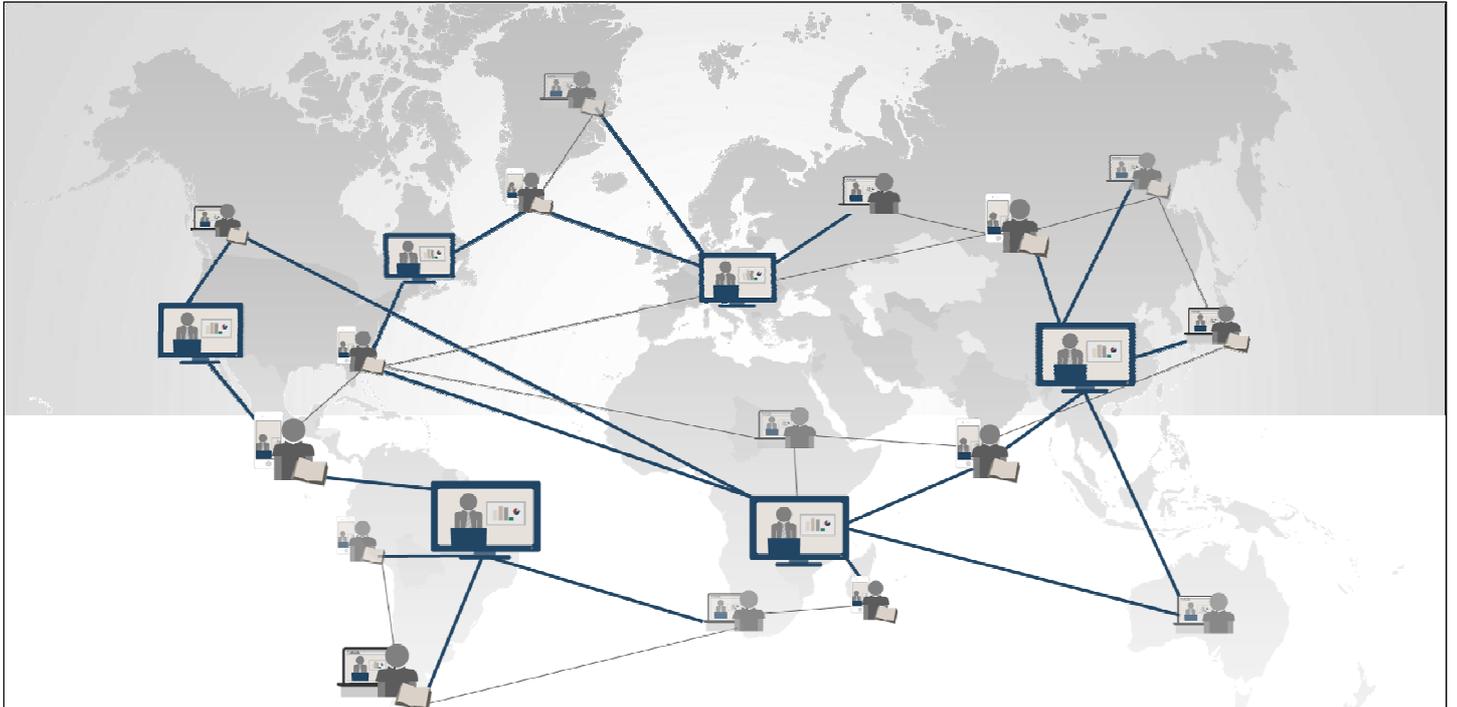
Question 4
 Place terms in sets. Let A and B be random variables inside a compartment indexed by i . Which of the following statements must be true? You may select more than one option.

Option	Submissions
For each i , $A(i)$ and $B(i)$ have CPDs that begin with the same values, regardless of the plate.	3 / 149
For each i , $A(i)$ and $B(i)$ have different CPDs.	10 / 143
If there is an instance of A for some i , then there is no instance of B for that i .	0 / 158
For each i , $A(i)$ and $B(i)$ have the same CPDs.	28 / 151
There is an instance of A and an instance of B for every i .	291 / 301
For each i , $A(i)$ and $B(i)$ are not independent.	5 / 162
For each i , $A(i)$ and $B(i)$ are independent.	12 / 149

In-class activities

- Enhanced interaction, with immediate feedback
- Individual tailoring of flow and pace through content
- Less threatening environment for students
- Detailed analytics to improve courses
- Time for meaningful engagement between students and faculty, students and peers
- **Interaction and creative problem solving are the real value of top universities**

Benefits to on-campus teaching



Universal Access to Education

What would we get if we could offer a free high-quality education to everyone? First, it establishes education as a basic human right, so that anyone with the motivation and the ability would have the opportunity to get the skills that they need to make a better life for themselves, their families and their communities. Second, it enables lifelong learning. It's a shame that for most of us, learning stops when we finish our formal education. With these amazing courses, we would always have the opportunity to explore new directions, whether to expand our minds or to make a change in our lives. Finally, it opens the door to a wave of innovation. Because talent can be found everywhere. Maybe the next Albert Einstein or the next Steve Jobs is living in some remote village in Africa. With access to education, he or she can come up with the next big idea, and help make the world a better place for all of us.