



Gradescope AI for Grading

NIPS 2016 Education Workshop

Pieter Abbeel
Professor UC Berkeley
Co-founder Gradescope

http://dsp.rice.edu/ml4ed_nips2016

- How we got started
- How grading works
- ML for analytics: projects we'd love to collaborate on
- AI-assisted grading: what we are working on



Arjun Singh
PhD Robotics



Sergey Karayev
PhD Computer Vision



Ibrahim Awwal
MS AI



Pieter Abbeel
Professor Robotics

- With TAs, built a tool to grade paper assignments online: faster, more consistent, and prevented cheating.
- Founded company at 50 courses. 2000+ courses today (200+ schools, 15M answers graded).
- Mission is to be the central place for all assessment.

Current Submission

Exam 1

QUESTION 1

Implement the Fibonacci sequence using recursion.

```
def fib(n):  
    if n < 2:  
        return n  
    else:  
        return fib(n-1) + fib(n-2)
```

Submission: 11

Rubric

Question 1

10/50 GRADED

TOTAL POINTS
10.0/10.0 pts

- | Rubric Item | Point Value | Description |
|-------------|-------------|--------------------------------|
| 1 | - 0.0 | Correct |
| 2 | - 4.0 | Incorrect base case |
| 3 | - 8.0 | Iterative instead of recursive |
| 4 | - 6.0 | Incorrect recursive case |

+ Add Rubric Item

Next >

Next Submission

<https://www.youtube.com/watch?v=UullQiXIjeQ>

The image shows a screenshot of the Gradescope web interface. At the top, there's a header bar with the Gradescope logo and the word "Grading". Below this is a browser-like window with the URL "www.gradescope.com". The main content area displays a question titled "Question 1". The question text is "What is the integral of x?". A handwritten submission is shown: $\frac{1}{2}x^2 + C$. To the right of the question, there's a summary: "0/3 GRADED", "TOTAL POINTS 3/3.0 pts", and a "Correct" feedback message with a score of "-0.0". There's also a "Add Rubric Item" button. At the bottom, a teal bar shows "Submission: 1" and a "Next >" button.

QUESTION 1

What is the integral of x?

$\frac{1}{2}x^2 + C$

Submission: 1

Next >

Question 1

0/3 GRADED

TOTAL POINTS
3/3.0 pts

1 - 0.0
Correct

+ Add Rubric Item

The screenshot shows the Gradescope interface for a question titled "Question 1". The question asks, "What is the integral of x?". A student submission of $\frac{1}{2}x^2$ has been graded. The grade breakdown is as follows:

- Submission 1: 0.0 points (Correct)
- Submission 2: -2.0 points (Missing constant (+C))

The total points available are 1/3.0 pts. There is a button to "Add Rubric Item".

QUESTION 1
What is the integral of x?

1/2 x^2

Question 1
1/3 GRADED

TOTAL POINTS
1/3.0 pts

1 - 0.0
Correct

2 - 2.0
Missing constant (+C)

+ Add Rubric Item

Submission: 2

Next >

The screenshot shows the Gradescope interface for a question titled "What is the integral of x?". The question has been graded with 0/3.0 pts. Three rubric items are listed:

- 1 - 0.0: Correct
- 2 - 2.0: Missing constant (+C)
- 3 - 1.0: Missing factor (1/2)

A button to "Add Rubric Item" is visible. At the bottom, it says "Submission: 3" and "Next >".

QUESTION 1
What is the integral of x?

Submission: 3

Next >

www.gradescope.com

Question 1

2/3 GRADED

TOTAL POINTS
0/3.0 pts

1 - 0.0
Correct

2 - 2.0
Missing constant (+C)

3 - 1.0
Missing factor (1/2)

+ Add Rubric Item

The screenshot shows the Gradescope grading interface. At the top, it displays the URL www.gradescope.com. Below this, there are two student submissions for "Question 1".

Student 1 (Ann): Name: Ann
QUESTION 1: What is the integral of x ?
Answer: $\frac{1}{2}x^2$

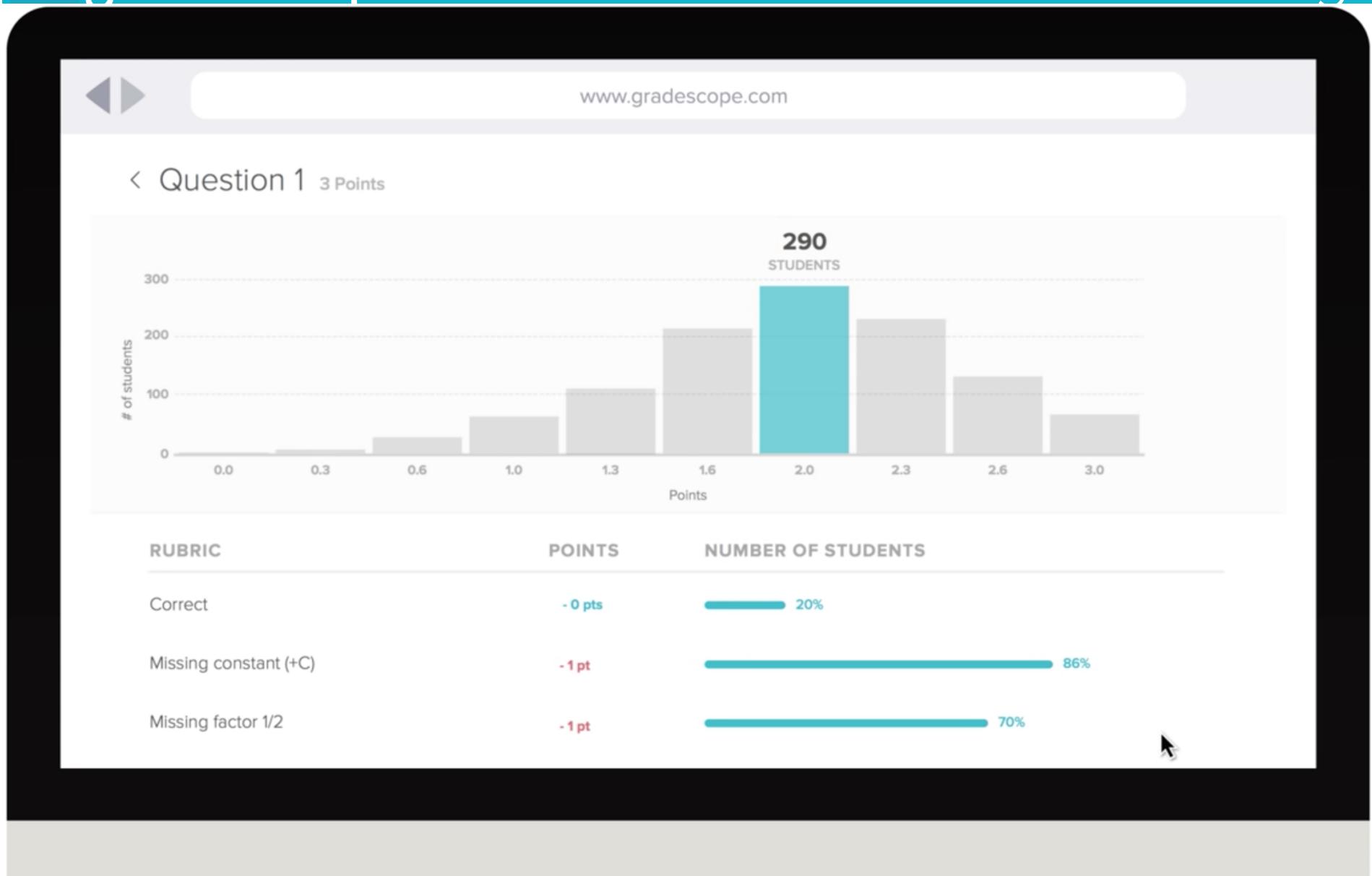
Student 2 (Minh): Name: Minh
QUESTION 1: What is the integral of x ?
Answer: x^2

Question 1 Rubric:

Category	Score	Description
1	- 0.0	Correct
2	- 1.0	Missing constant (+C)
3	- 1.0	Missing factor (1/2)

Submission Summary: Submission: 3

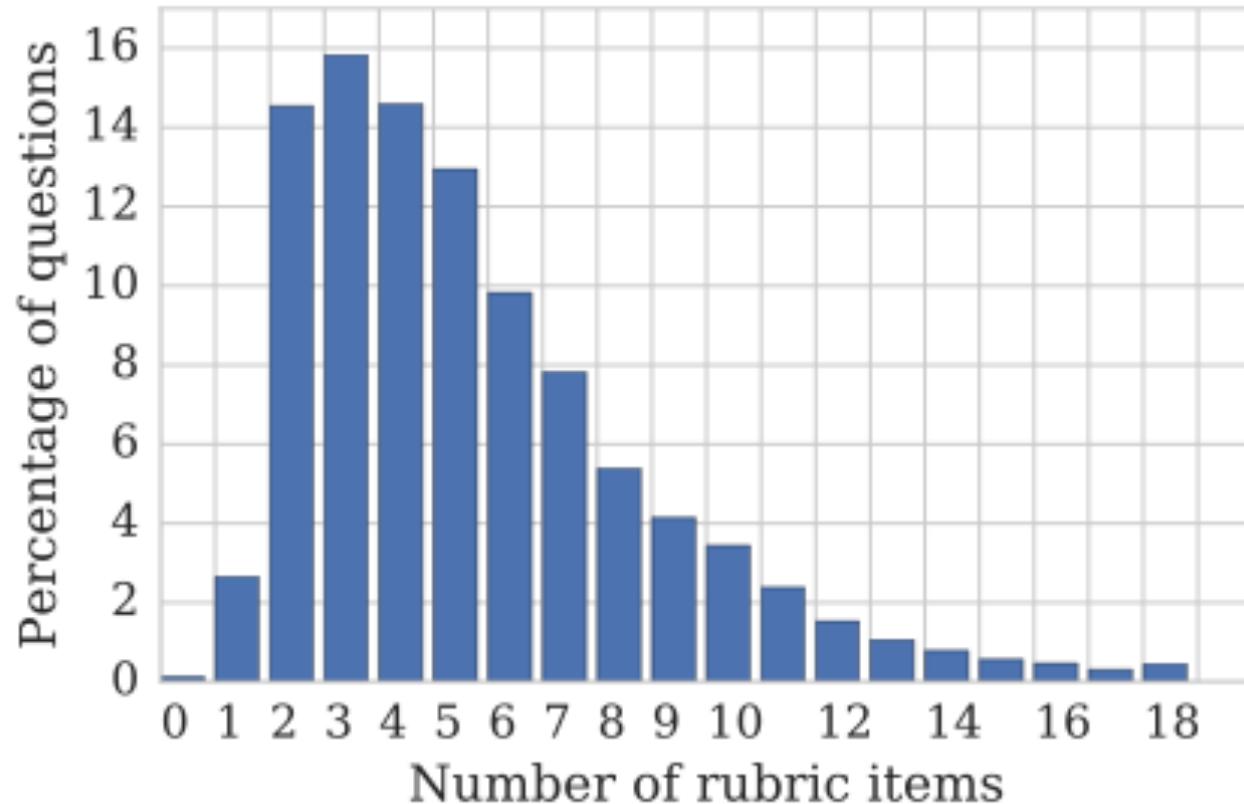
Next >



We would love to collaborate on these!

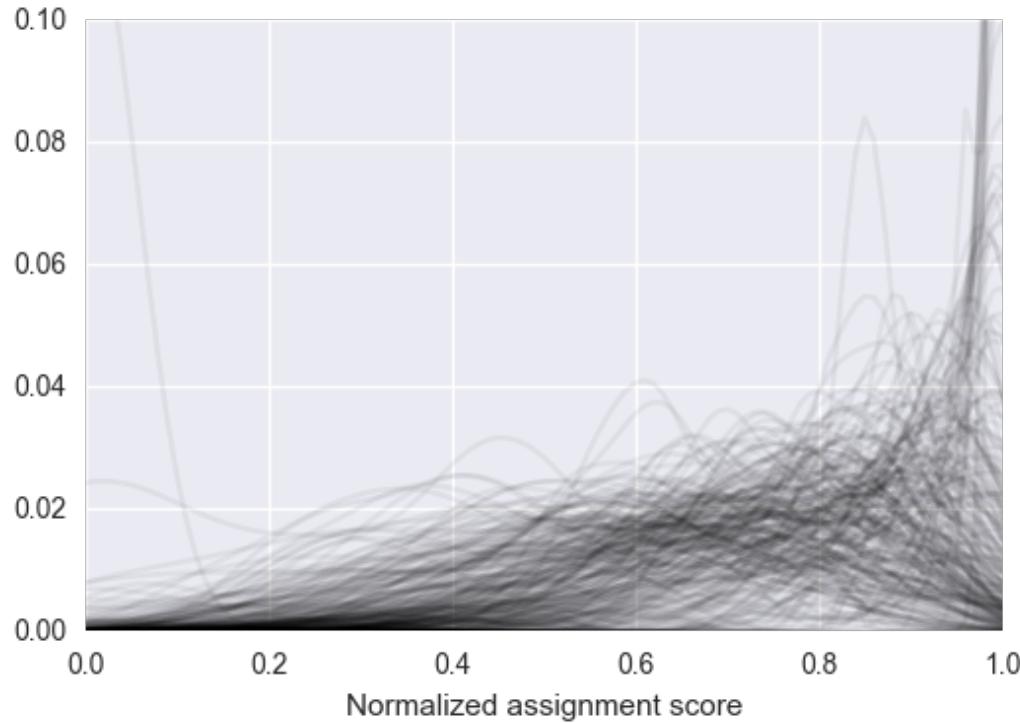
- Item response theory for rubric-graded questions
- Inferring concepts
- Predicting student outcomes

- On Gradescope, mean number of applied rubric items is 5.6.
More questions have 8 or more rubric items than 2 or less.
- Approximately equal number of additive and subtractive rubrics.

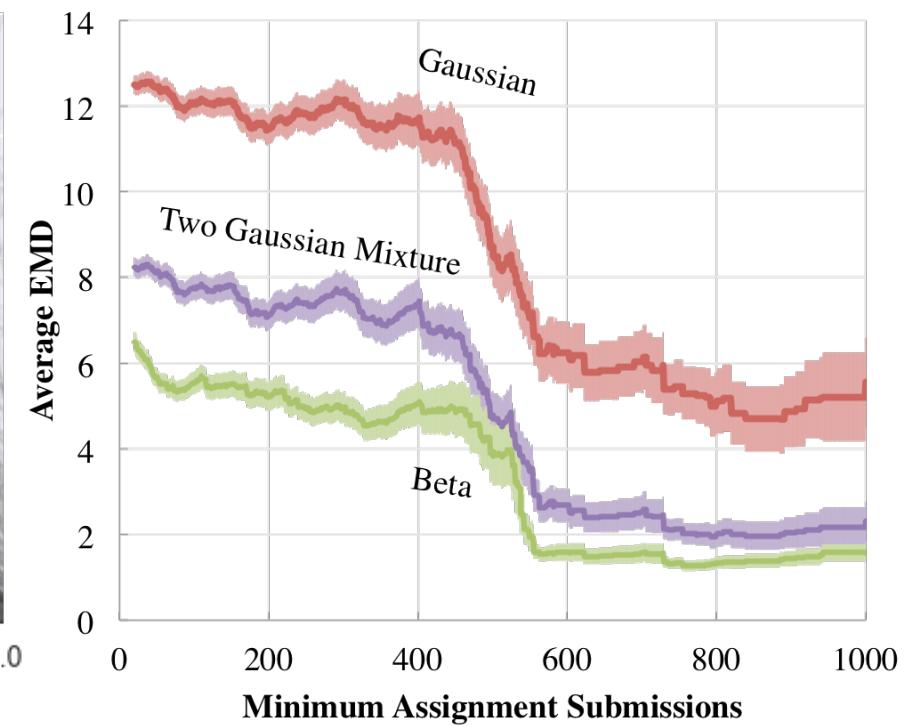


- Assignment and question difficulty do not appear to be normally distributed.

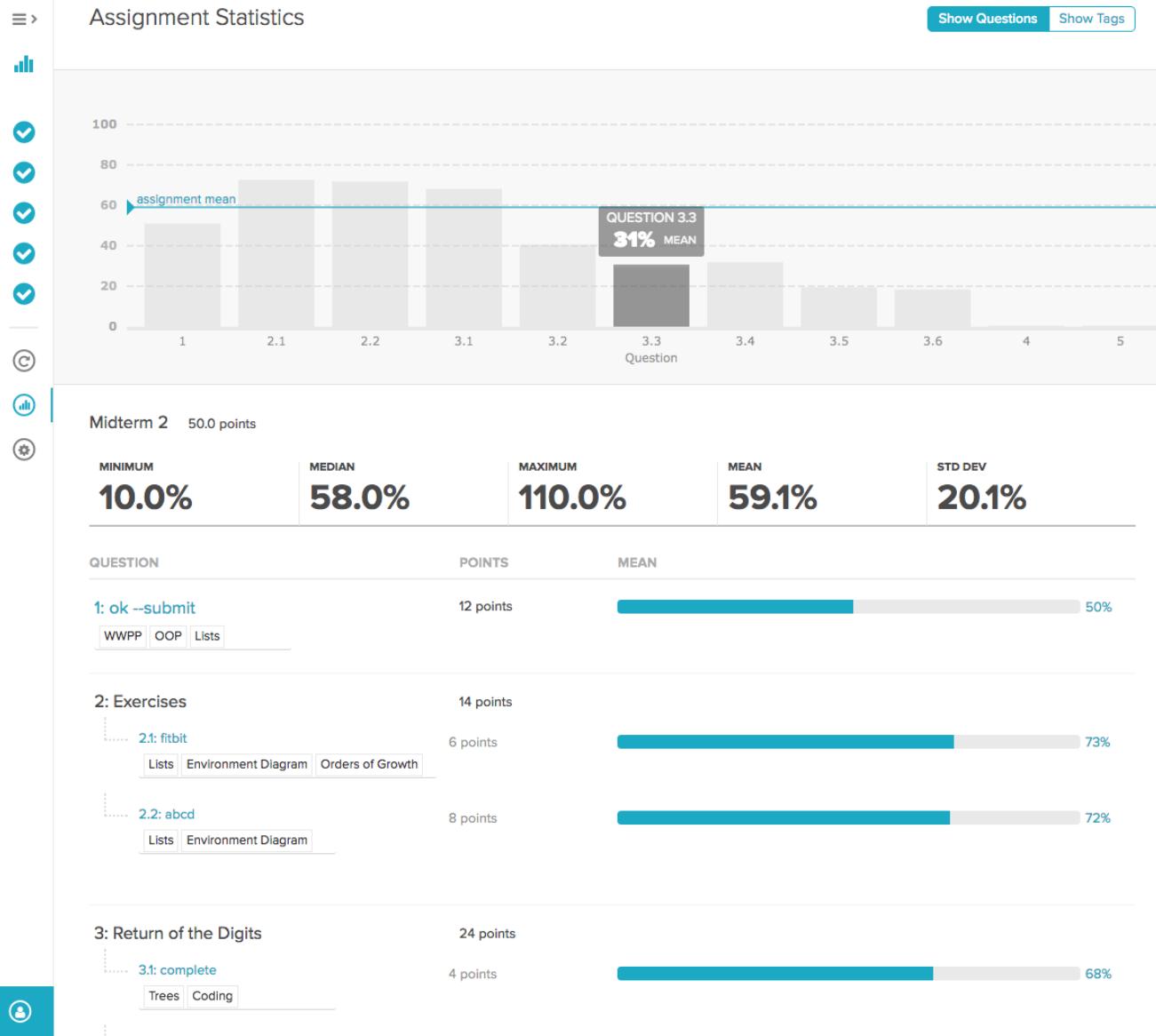
Gradescope assignment score distributions



Gaussian distribution does not fit distributions well



- Would love to collaborate on evaluating polytomous IRT models on our additive and subtractive rubric-graded questions.
- Will probably need to develop new models.



Instructors can tag questions with concepts to see how their assignment breaks down by concept.



Assignment Statistics

[Show Questions](#)[Show Tags](#)100
80
60
40
20
0

WWPP OOP Lists Environment Diagram Orders of Growth Trees Tag Coding Nonlocal Linked Lists Recursion Higher Order Functions



Midterm 2 50.0 points

MINIMUM

10.0%

MEDIAN

58.0%

MAXIMUM

110.0%

MEAN

59.1%

STD DEV

20.1%

TAG

POINTS

MEAN

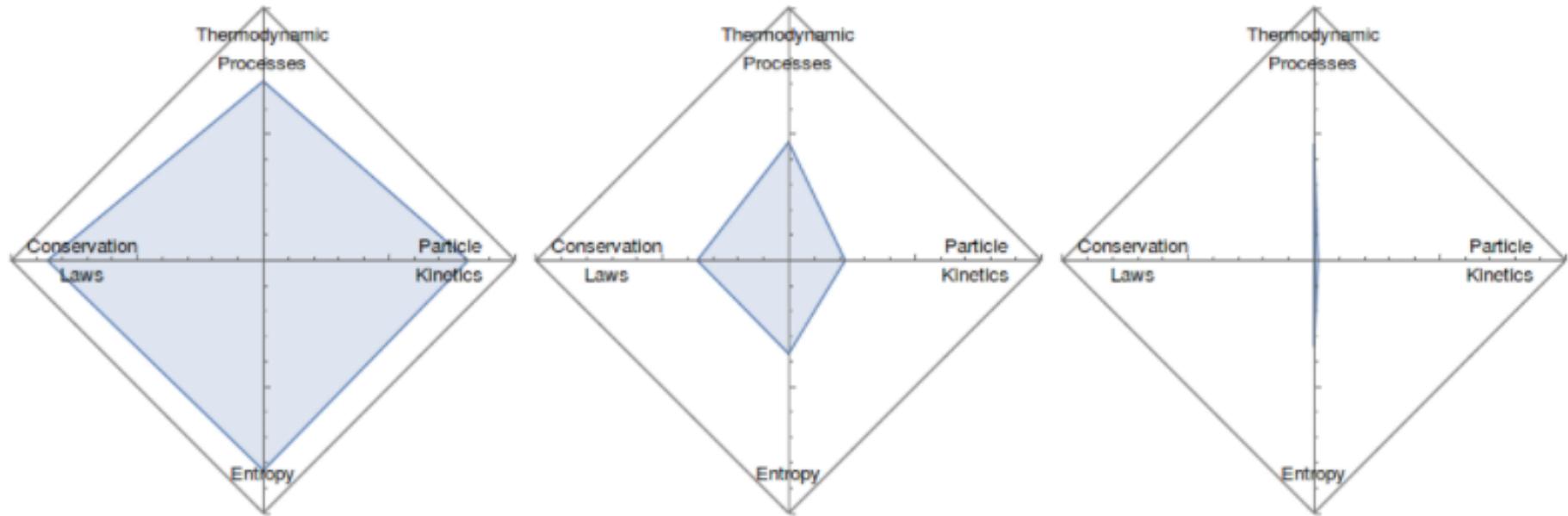
WWPP

24 points



50%

- We would like to infer concepts automatically.
- One way forward: Gradescope user Lenny Evans used topic modeling on rubric item applications.



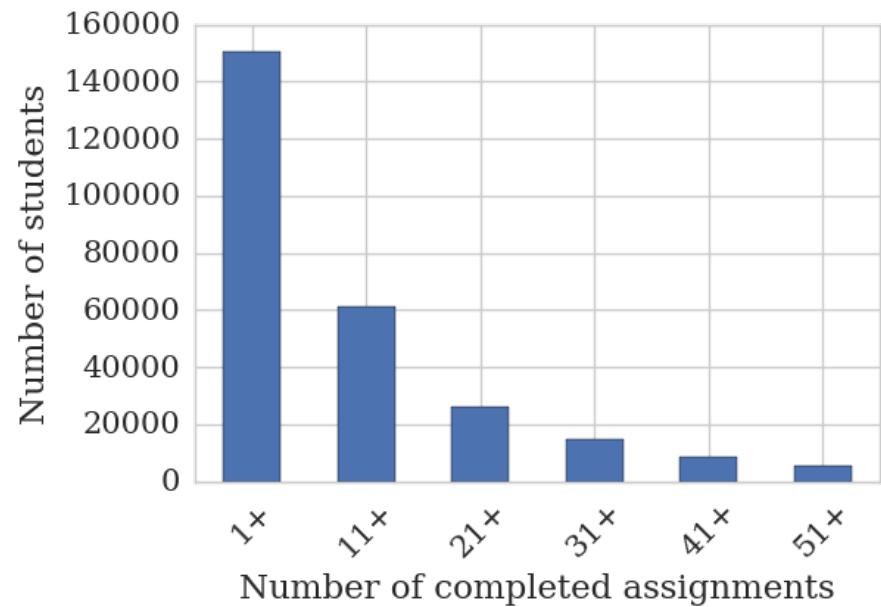
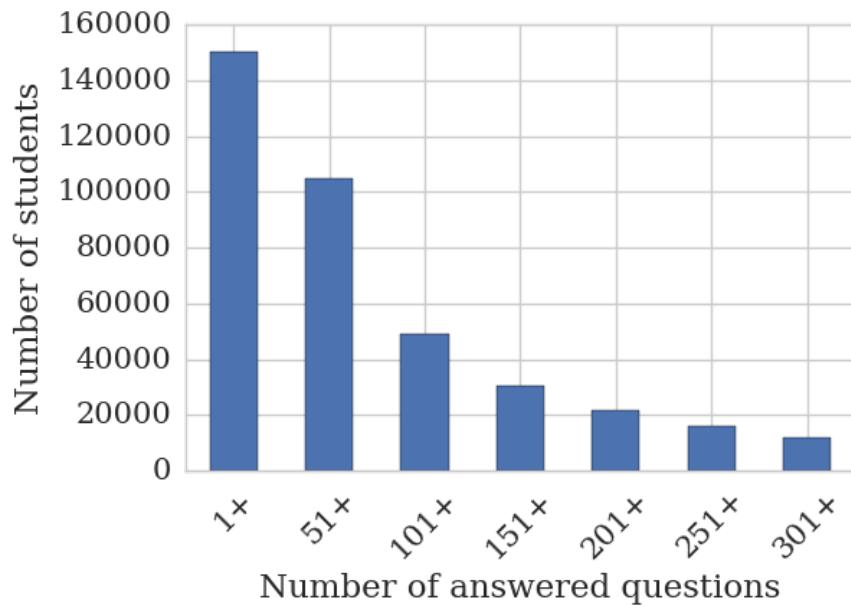
Best-scoring
students

Medium-scoring
students

Low-scoring
students

<http://blog.ultramarineneutrinos.com/topic-modeling-and-gradescope/>

- Thousands of students have a significant amount of real work on Gradescope.
- We would like to help institutions with pre-requisite analysis and other tasks.



- Our path to auto-grading: grouping answers for the instructors to grade
- Spectrum of question complexity
- Group review interface

- We value rubric-based grading: partial credit and useful feedback to the student.
Goal is to scale, not replace, instructor grading.



Scantron-like

- $f(s) + g(s) + h(s)$
 $f(s)/6 + g(s)/3 + h(s)/2$
 $\min(f(s), g(s), h(s))$
 $\max(f(s), g(s), h(s))$

- $f(s) + g(s) + h(s)$
 $f(s)/6 + g(s)/3 + h(s)/2$
 $\min(f(s), g(s), h(s))$
 $\max(f(s), g(s), h(s))$

Freeform M.C.

Circle *True* or *False*: meow is cat[0]

Circle *True* or *False*: meow[0][0] is cat[0][0]

Circle *True* or *False*: meow is cat[0]

Circle *True* or *False*: meow[0][0] is cat[0][0]

Constrained
text

- (a) ISN
(b) DHCP

- (a) ACK
(b) DHCP

Math

 Q1.1 [3pt] What is the integral of x ?

$$\frac{1}{2}x^2 + C$$

 Q1.1 [3pt] What is the integral of x ?

$$\frac{1}{2}x^2 + C$$

Code

(a) (2 pt) Select the titles of all movies that have a rating greater than 7 in alphabetical order.

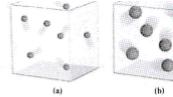
```
select title from ratings
where rating > 7
order by title;
```

(a) (2 pt) Select the titles of all movies that have a rating greater

```
select title from ratings
where rating > 7
order by title.
```

Short answer

2. (5 points) The figure to the right represents three ideal gas samples (a, b, c). Assume that the mass of each particle is proportional to its size, and that all the gas samples are at the same temperature, in which sample is the pressure highest? Explain your reasoning.

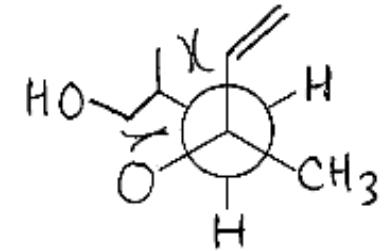
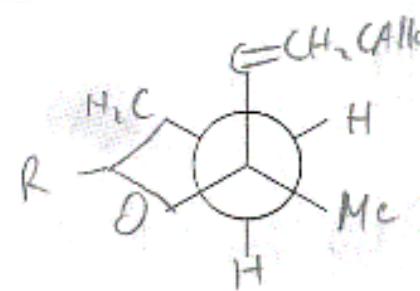
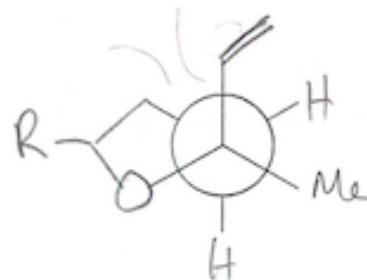


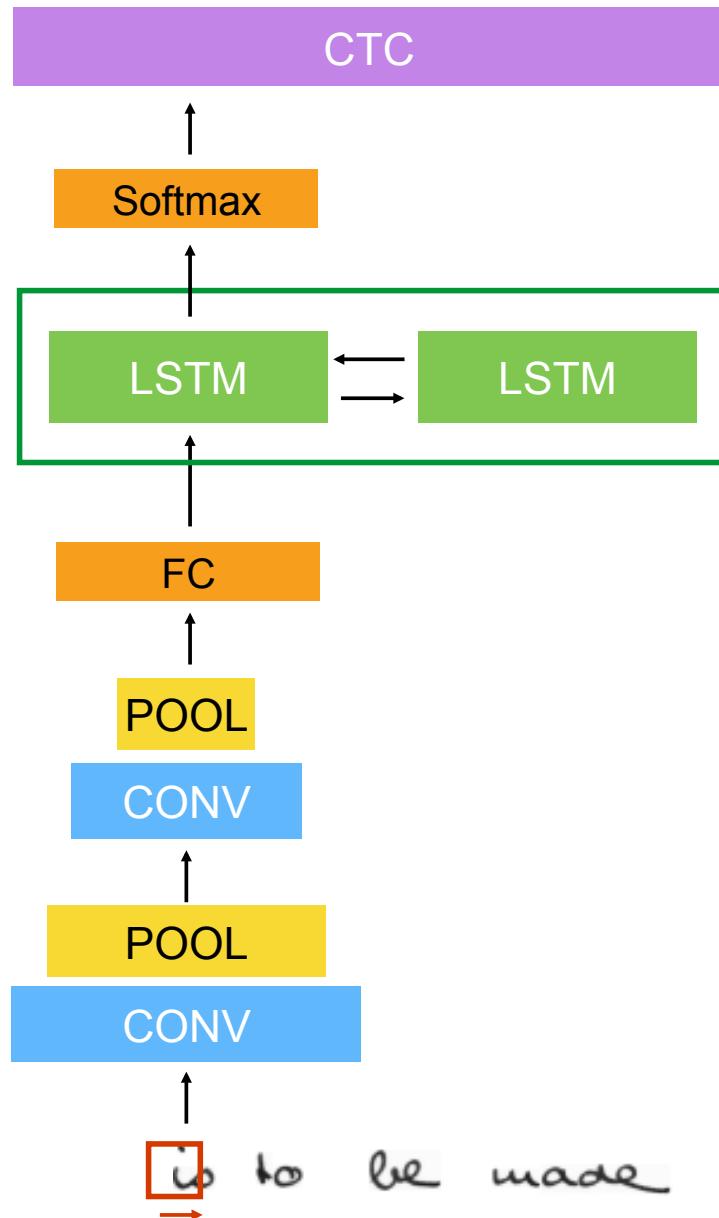
Sample A, because more open space
allows to move move freely &
b/c particles are moving faster!

2. (5 points) The figure to the right represents three ideal gas samples (a, b, c). Assume that the mass of each particle is proportional to its size, and that all the gas samples are at the same temperature, in which sample is the pressure highest? Explain your reasoning.

C has the highest pressure. For ideal gases,
 $P = \frac{nRT}{V}$, since R, T, and V are the same, the gas
with larger n has higher pressure. By counting, c has
11 molecules while a and b only have 7.

Diagrams





→ “is to be made at a meeting at Labour”

Model architecture based on a combination of LeNet and Alex Graves PAMI 2009.

IAM Offline Handwriting Recognition DB (2002)

	Trainval	Test
# lines	7081	2781
# writers	339	161

is to be made at a meeting at Labour

That's how he got the votes, that's how

Thal's how he got the votes, that's how

with black, watted hair and striking

with bleek, watted hair and sfriking

But all three were to the casual

But all three were to the casual

He love frou his breakfast-nook bech

He love frou his breakfast -nook bech

wars impatient far it. Teis afternoon, ther, we climbed

wars impatient far it. Teis affroon, then, wechinbed

83.5% character accuracy
(no language model)

$$acc = 100 * \left(1 - \frac{\text{insertions} + \text{substitutions} + \text{deletions}}{\text{total length of test set transcriptions}} \right)$$

 gradescope[◀ Back to Course](#)**Demo Quiz** Edit Outline Manage Scans Manage Submissions **Grade Submissions** Review Grades Regrade Requests Statistics Settings**Q 2 Answer Groups**

0/4 Groups Confirmed

0/18 Confirmed Answers

2 Ungrouped Answers >

[? Explanation](#)  [Settings](#)We found **4 groups** for this **Fill-in-the-blank Math Question** ([Edit type](#)).

Review each Group for correctness, and then process Ungrouped Answers.

X =

20

Group 1 · Unconfirmed - 11 answers

X =

80

Group 2 · Unconfirmed - 4 answers

X =

Blank

Group 3 · Unconfirmed - 2 answers

 Admin User

Review unconfirmed groups.

Review Groups >

gradescope 

[Back to Course](#)

Demo Quiz

- Edit Outline
- Manage Scans
- Manage Submissions
- Grade Submissions
- Review Grades

- Regrade Requests
-  Statistics
-  Settings

Q 2 Answer Groups > Group 1 / 4

0/4 Groups Confirmed      

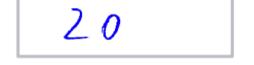
0/18 Confirmed Answers [2 Ungrouped Answers >](#)

$x =$ 

20
Group 1 · Unconfirmed - 11 answers
[Rename](#) [Delete \(and Ungroup Answers\)](#)

Make sure all answers here actually belong in the group.
Answers we were not quite sure about are presented first, in red.

ANSWERS TO CONFIRM

$x =$ 	$x =$ 	$x =$ 	$x =$ 
$x =$ 	$x =$ 	$x =$ 	$x =$ 

 Admin User 

 View Groups [Previous Group](#) [Next Group](#) [Confirm & Review Next Group >](#)

gradescope ≡

[Back to Course](#)

Demo Quiz

- Edit Outline
- Manage Scans
- Manage Submissions
- Grade Submissions
- Review Grades

- Regrade Requests
- Statistics
- Settings

Q 2 Answer Groups > Group 2 / 4

1/4 Groups Confirmed Progress bar

11/19 Confirmed Answers 1 Ungrouped Answers >

Settings

$x = \boxed{80}$

80
Group 2 · Unconfirmed - 5 answers
✎ Rename ✖ Delete (and Ungroup Answers)

Make sure all answers here actually belong in the group.
 Answers we were not quite sure about are presented first, in red.

ANSWERS TO CONFIRM

$x = \boxed{80}$

$x = \boxed{80}$

$x = \boxed{80}$

$x = \boxed{20}$
REMOVE space
Q

$x = \boxed{80}$

View Groups ◀ Previous Group Next Group ▶ Confirm & Review Next Group ▶

[← Back to Course](#)

Demo Quiz

- Edit Outline
- Manage Scans
- Manage Submissions
- Grade Submissions
- Review Grades

- Regrade Requests
- Statistics
- Settings

Admin User

2 answers selected

qs.dev:3000/courses/1/questions/18/answer_groups/39

Q 2 Answer Groups > Ungrouped Answers

4/4 Groups Confirmed

18/18 Confirmed Answers 2 Ungrouped Answers >

Here are answers that are about to be graded individually.
You can continue on to Grade Answers, or you can move them into groups here.

x = 20 x = 20

DESELECT space

2 answers selected.

+ Create a Group

1 x = 20 (11)

2 x = 80 (4)

3 x = Blank (2)

4 x = 2 (1)

Name:

Student ID:

Introduction to Gradescope

Spring 2015 Quiz 1

Q1. [6 pt] Math
1: Integral (1 points)

Q1.1 [3pt] What is the integral of x ?

Q1.2 [3pt] What is x , if $50 - x = 30$?
Write your final answer in the box.

Q2. [1pt] Geography
3: Geography (1 points)
Shade the South Dakota in this map.



Q3. [2pt] Chemistry
4: Chemistry (1 points)
Which of the following are noble gases?
 Argon
 Helium
 Oxygen
 Hydrogen

169.6 ms

Outline for Demo Quiz

Select name region

Create questions and subquestions via the + buttons below, or by dragging boxes on the document. Reorder and indent questions by dragging them in the outline.

#	TITLE	POINTS
1	Integral	1
2	Algebra	1
3	Geography	1
4	Chemistry	1

+ new question

Save Outline **Cancel**



Thank you!

NIPS 2016 Education Workshop

Pieter Abbeel
pabbeel@cs.berkeley.edu

Professor UC Berkeley
Co-founder Gradescope

http://dsp.rice.edu/ml4ed_nips2016