

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

Rubric

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)
```

Question 1

10/50 GRADED

TOTAL POINTS
10.0/10.0 pts

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

Submission: 11
Next >

Rubric Item
Point Value

Rubric Item
Description

Next Submission

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

The screenshot shows the Gradescope interface for a question. The question asks for the integral of x. A handwritten answer, $\frac{1}{2}x^2 + C$, is shown in the main area. The right sidebar displays the question details, including the total points (3/3.0 pts) and a grading entry for 1 point, marked as correct. A 'Next' button is visible at the bottom right.

QUESTION 1
What is the integral of x?

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

Question 1
0/3 GRADED
TOTAL POINTS 3/3.0 pts
1 - 0.0 Correct
+ Add Rubric Item

Submission: 1 Next >

The screenshot shows the Gradescope interface for a question. The question text is "QUESTION 1 What is the integral of x?". The student's answer is $\frac{1}{2}x^2$. The grading sidebar on the right shows "Question 1" with a progress bar for "1/3 GRADED". It displays "TOTAL POINTS 1/3.0 pts" and a gear icon. There are two items in the rubric: item 1 is "Correct" with a score of -0.0, and item 2 is "Missing constant (+C)" with a score of -2.0. A button "+ Add Rubric Item" is visible below the items. At the bottom of the interface, it says "Submission: 2" and a "Next >" button.

The screenshot shows the Gradescope interface for a question. The question text is "QUESTION 1 What is the integral of x?". The student's answer is x^2 . The grading panel on the right shows a progress bar for "Question 1" at "2/3 GRADED". The total points are "0/3.0 pts". The rubric items are:

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

At the bottom, it says "Submission: 3" and a "Next >" button.

◀ ▶
www.gradescope.com

-1.0

Missing constant (+C)

Exam Name: Ann

QUESTION 1
What is the integral of x?

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

-1.0

Missing constant (+C)

-1.0

Missing factor 1/2

Exam Name: Minh

QUESTION 1
What is the integral of x?

$$x^2$$

Question 1

2/3 GRADED

TOTAL POINTS ⚙

1/3.0 pts

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

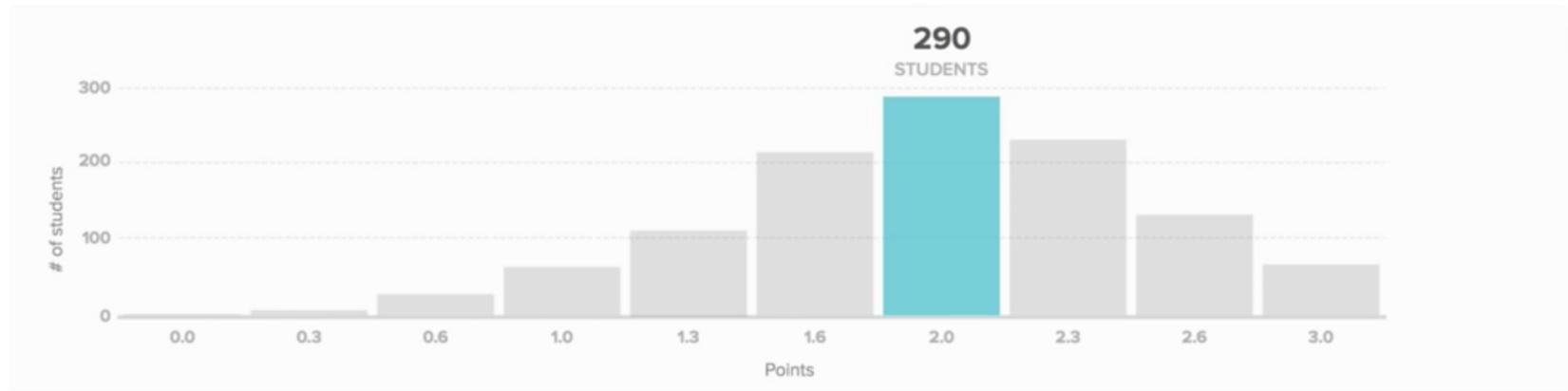
+ Add Rubric Item

Submission: **3**
Next >



www.gradescope.com

< Question 1 3 Points

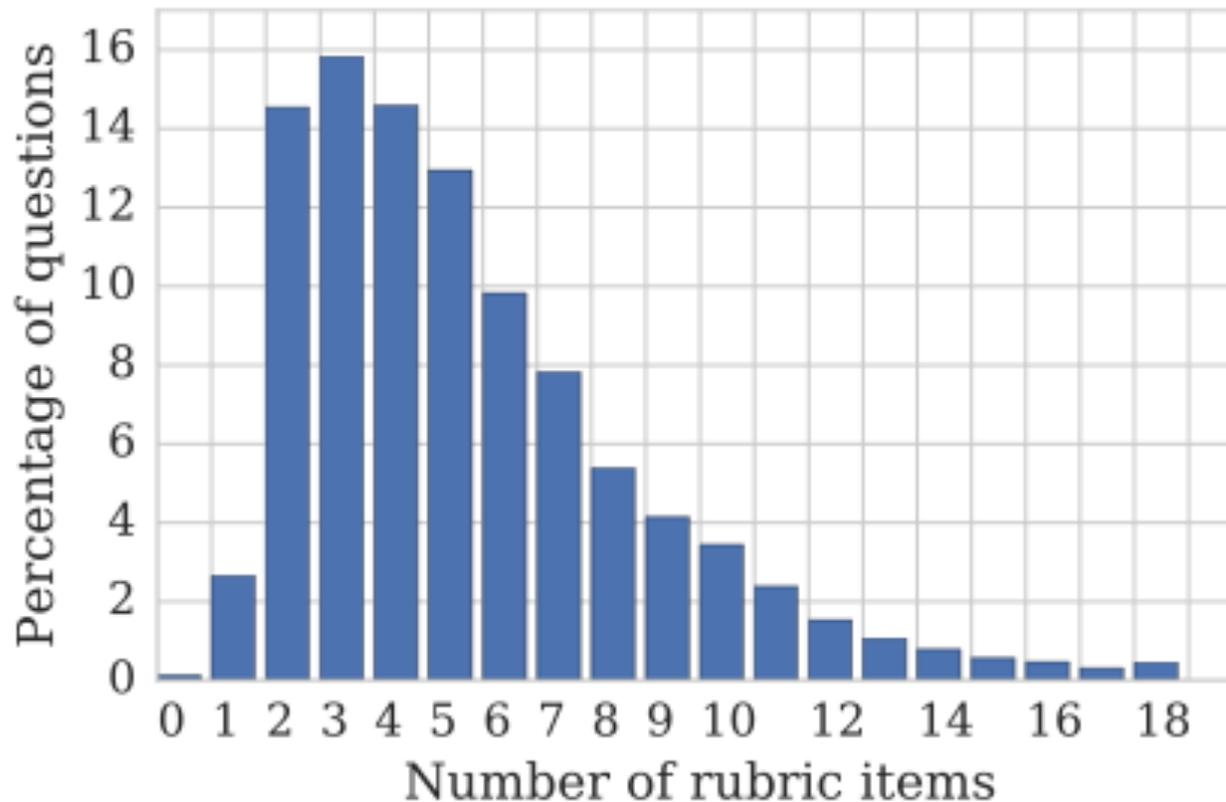


RUBRIC	POINTS	NUMBER OF STUDENTS
Correct	- 0 pts	20%
Missing constant (+C)	- 1 pt	86%
Missing factor 1/2	- 1 pt	70%

We would love to collaborate on these!

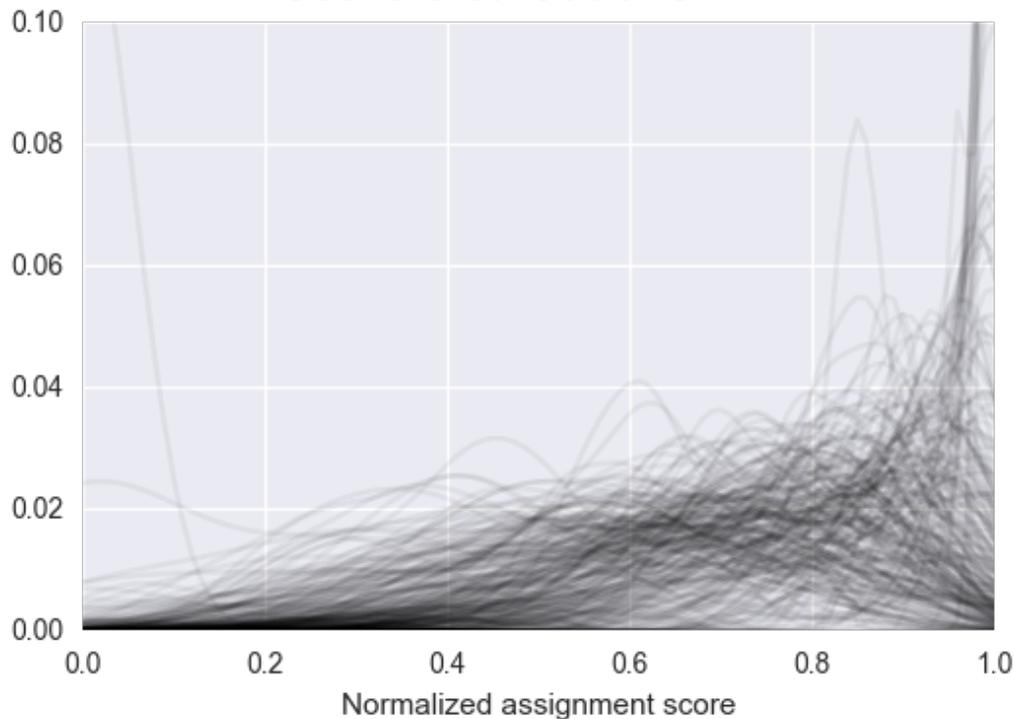
- Item response theory for rubric-graded questions
- Inferring concepts
- Predicting student outcomes

- 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.

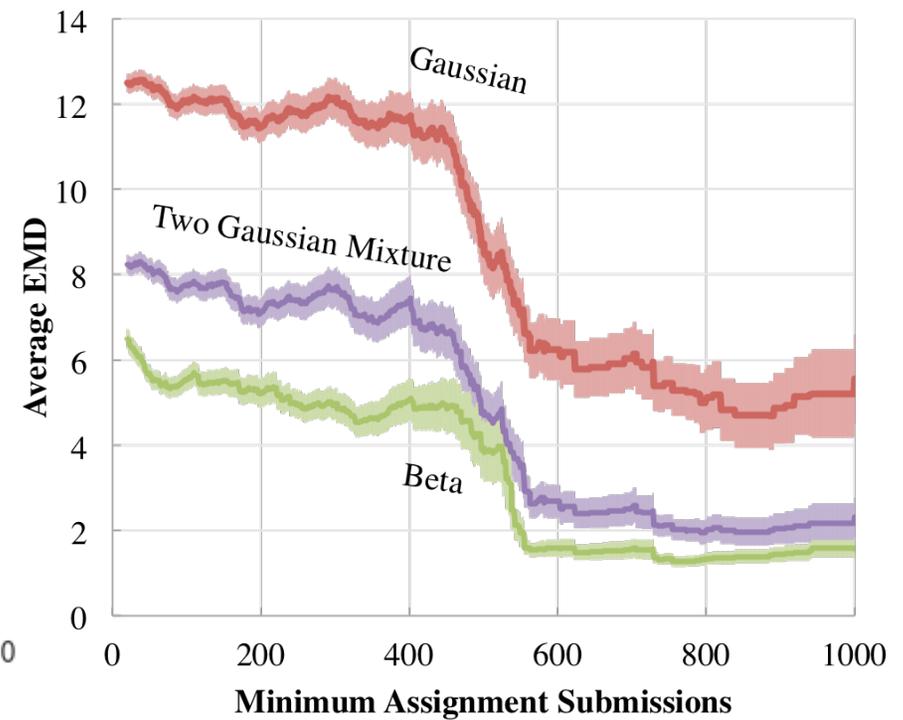


- Assignment and question difficulty do not appear to be normally distributed.

Gradescope assignment score distributions

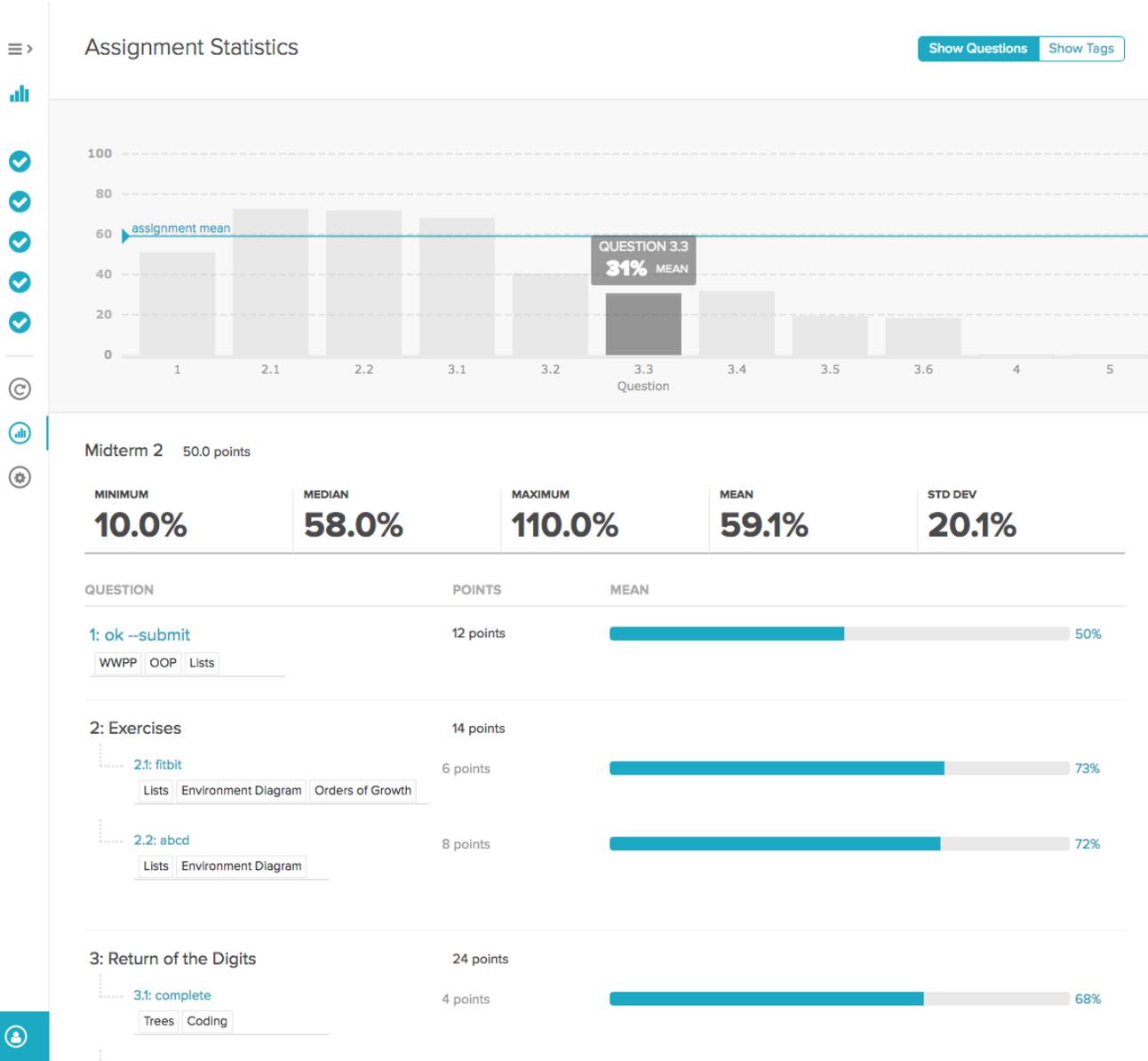


Gaussian distribution does not fit distributions well



(figure courtesy Chris Piech)

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



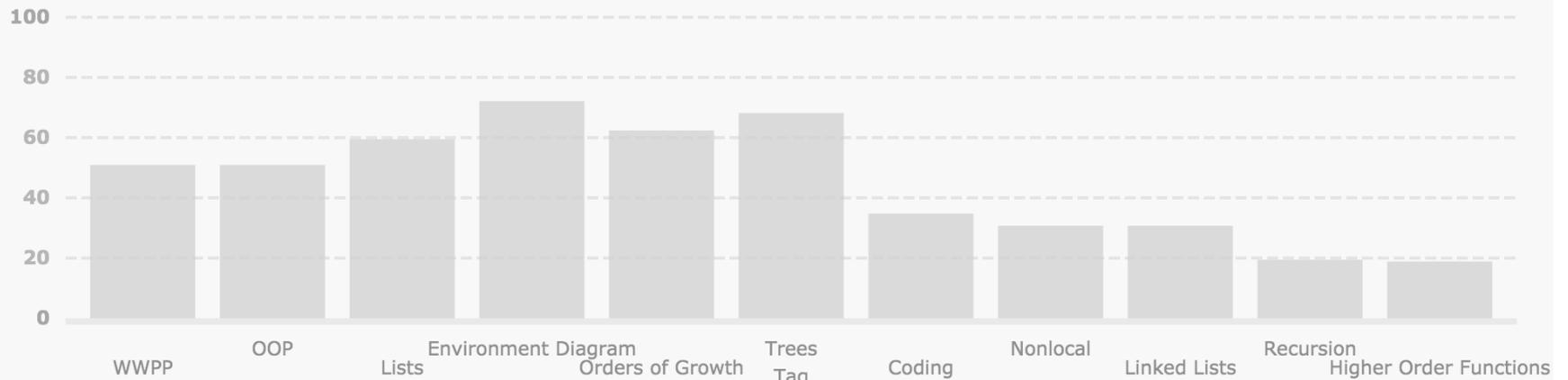
Instructors can tag questions with concepts to see how their assignment breaks down by concept.



Assignment Statistics

Show Questions

Show Tags



Midterm 2 50.0 points

MINIMUM

10.0%

MEDIAN

58.0%

MAXIMUM

110.0%

MEAN

59.1%

STD DEV

20.1%

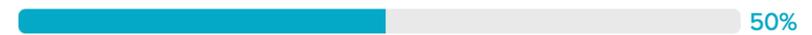
TAG

POINTS

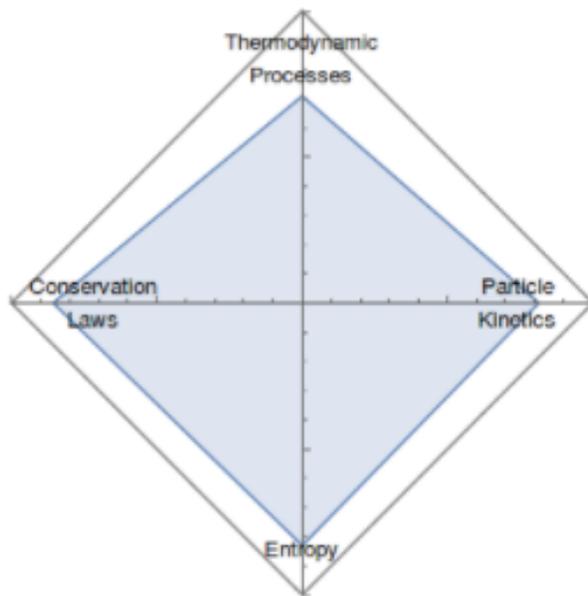
MEAN

WWPP

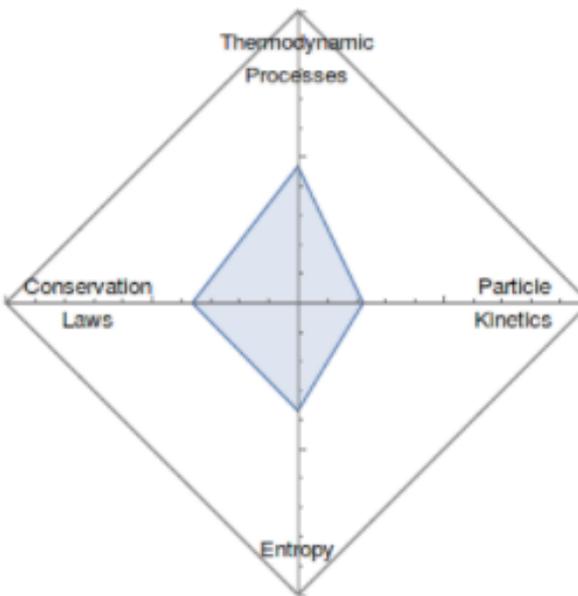
24 points



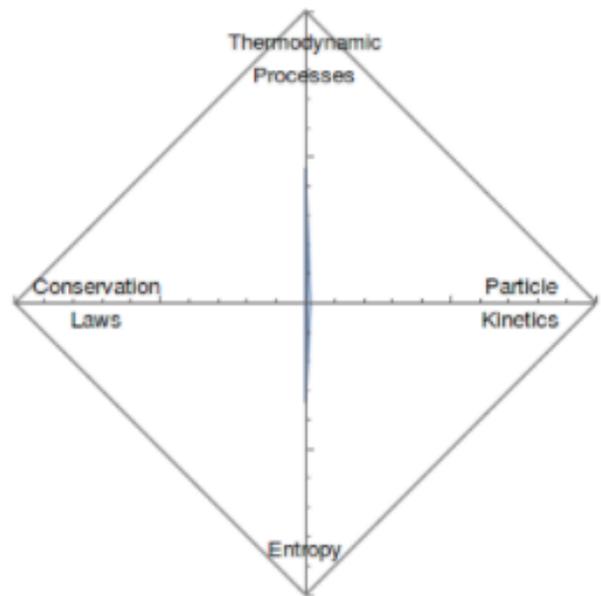
- We would like to infer concepts automatically.
- One way forward: Gradescope user Lenny Evans used topic modeling on rubric item applications.



Best-scoring
students



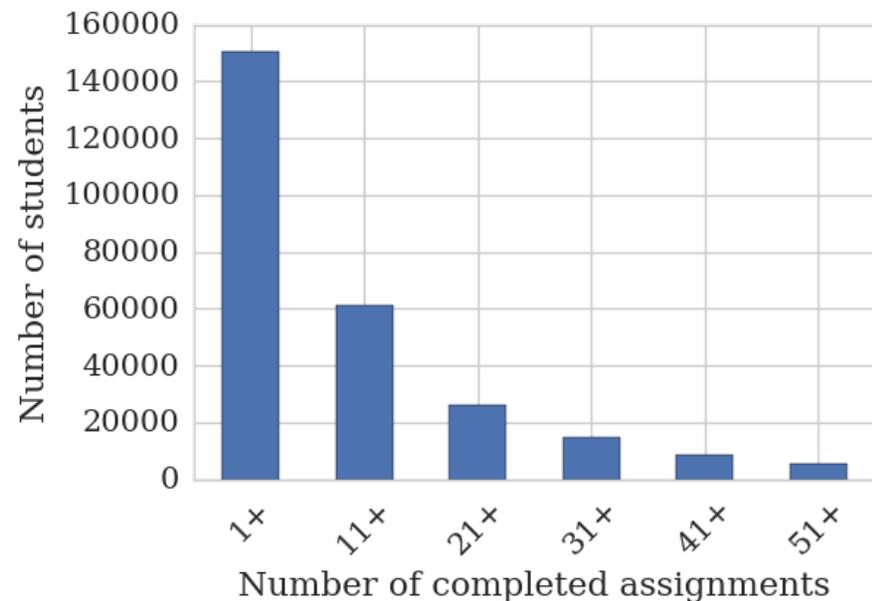
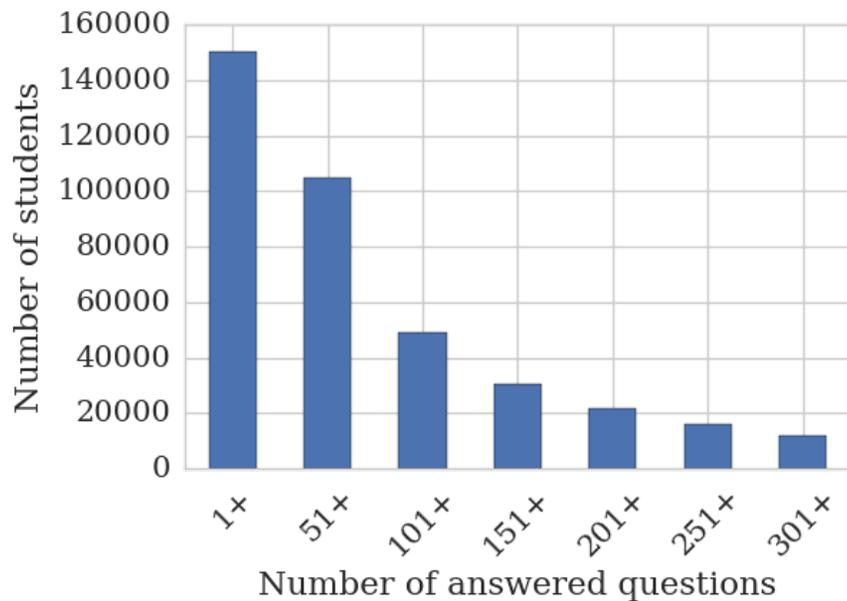
Medium-scoring
students



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

- | | | |
|---|---|---|
| <input checked="" type="radio"/> $f(s) + g(s) + h(s)$ | <input type="radio"/> $f(s) + g(s) + h(s)$ | <input checked="" type="radio"/> $f(s) + g(s) + h(s)$ |
| <input checked="" type="radio"/> $f(s)/6 + g(s)/3 + h(s)/2$ | <input type="radio"/> $f(s)/6 + g(s)/3 + h(s)/2$ | <input type="radio"/> $f(s)/6 + g(s)/3 + h(s)/2$ |
| <input checked="" type="radio"/> $\min(f(s), g(s), h(s))$ | <input checked="" type="radio"/> $\min(f(s), g(s), h(s))$ | <input checked="" type="radio"/> $\min(f(s), g(s), h(s))$ |
| <input checked="" type="radio"/> $\max(f(s), g(s), h(s))$ | <input checked="" type="radio"/> $\max(f(s), g(s), h(s))$ | <input checked="" type="radio"/> $\max(f(s), g(s), h(s))$ |

Freeform M.C.

- | | |
|--|--|
| Circle <i>True</i> or <u>False</u> : meow is cat[0] | Circle <u>True</u> or <i>False</i> : meow is cat[0] |
| Circle <i>True</i> or <u>False</u> : meow[0][0] is cat[0][0] | Circle <i>True</i> or <u>False</u> : meow[0][0] is cat[0][0] |

Constrained text

- | | |
|-----------------|-----------------|
| (a) <u>ISN</u> | (a) <u>ACK</u> |
| (b) <u>DHCP</u> | (b) <u>DHCP</u> |

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.

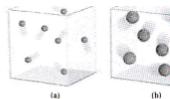
select title from ratings
where rating > 7
order by title;

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

select title from ratings
where rating > 7
order by title

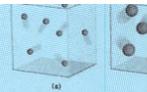
Short answer

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.



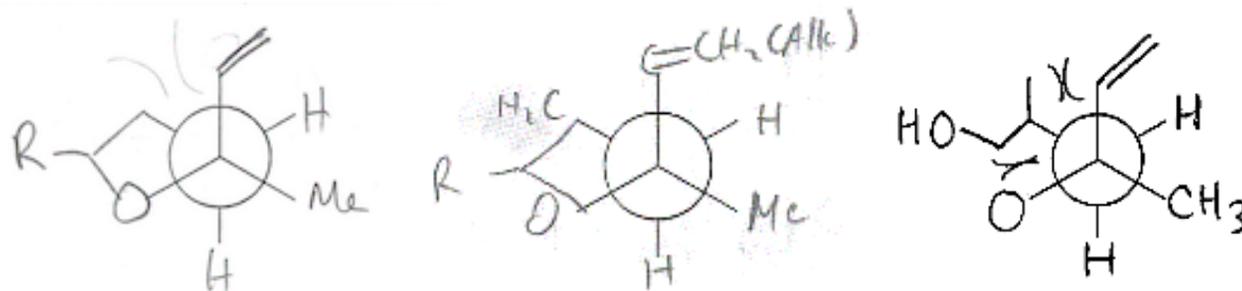
Sample A, because more open space allows to move more freely & b/c particles are moving faster.

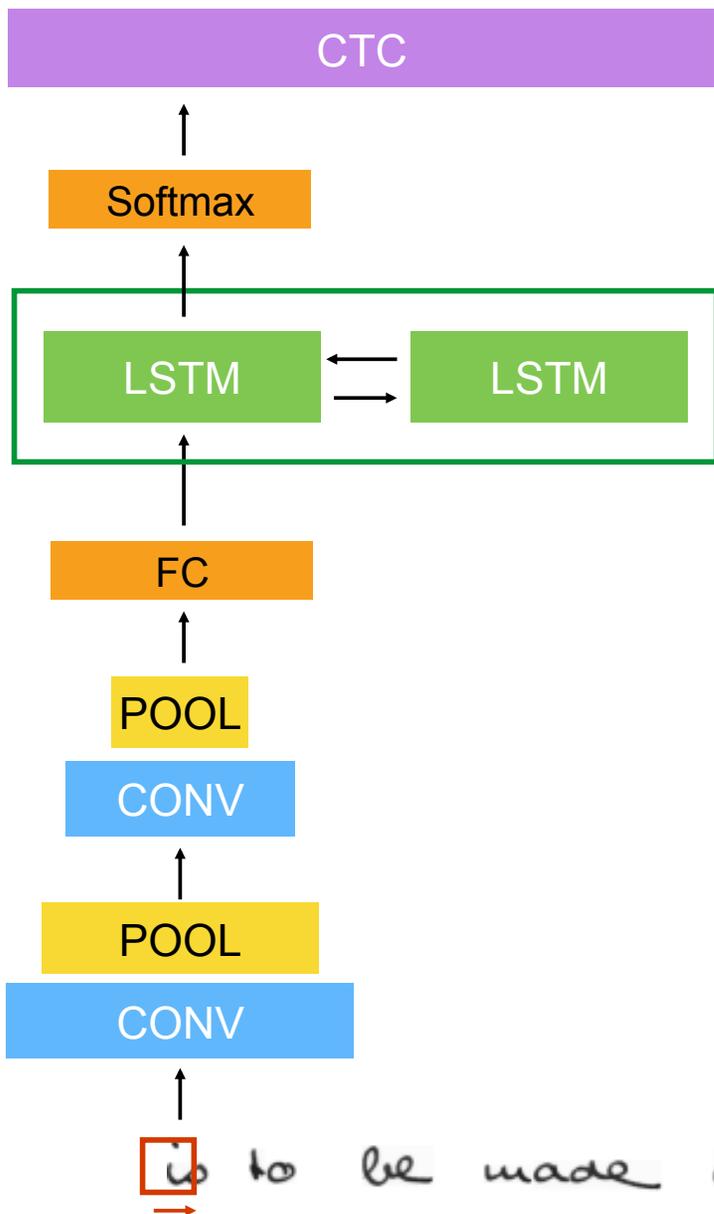
2. (5 points) The figure to the right represents three ideal gas samples (a, b, c). Assume that the mass of each particle is proportional to its size, and that all the gas samples are at the same temperature, in which sample is the pressure highest? Explain your reasoning.



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

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 becch

He love frou his breakfait -nook becch

was impatient far it. This afternoon, then, we climbed

wars impatient far it. Teis affroon, then, wechinbed

83.5% character accuracy
(no language model)

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

< 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

? Explanation ⚙ Settings

0/18 Confirmed Answers

2 Ungrouped Answers >

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

[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

☰ ○ ⊞ Settings

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 = <input style="border: 1px solid gray; padding: 5px; width: 80px; height: 30px;" type="text" value="20"/>	x = <input style="border: 1px solid gray; padding: 5px; width: 80px; height: 30px;" type="text" value="20"/>	x = <input style="border: 1px solid gray; padding: 5px; width: 80px; height: 30px;" type="text" value="20"/>	x = <input style="border: 1px solid gray; padding: 5px; width: 80px; height: 30px;" type="text" value="20"/>
x = <input style="border: 1px solid gray; padding: 5px; width: 80px; height: 30px;" type="text" value="20"/>	x = <input style="border: 1px solid gray; padding: 5px; width: 80px; height: 30px;" type="text" value="20"/>	x = <input style="border: 1px solid gray; padding: 5px; width: 80px; height: 30px;" type="text" value="20"/>	x = <input style="border: 1px solid gray; padding: 5px; width: 80px; height: 30px;" type="text" value="20"/>

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



11/19 Confirmed Answers

1 Ungrouped Answers >

☰ ○ ⊞ Settings

x =

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 =

x =

x =

x =

REMOVE

x =

<☰

< Back to Course

Demo Quiz

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

- Regrade Requests
- Statistics
- Settings

Q 2 Answer Groups > Ungrouped Answers

4/4 Groups Confirmed ⊞ Settings

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 ☰

▶ 20 (1)

2 $x =$ 80

80 (4)

3 $x =$

Blank (2)

4 $x =$ 2

2 (1)

Admin User ^

2 answers selected
⊞ View Groups
Grade >

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

Thank you!

NIPS 2016 Education Workshop

Pieter Abbeel

pabbeel@cs.berkeley.edu

Professor UC Berkeley
Co-founder Gradescope

http://dsp.rice.edu/ml4ed_nips2016