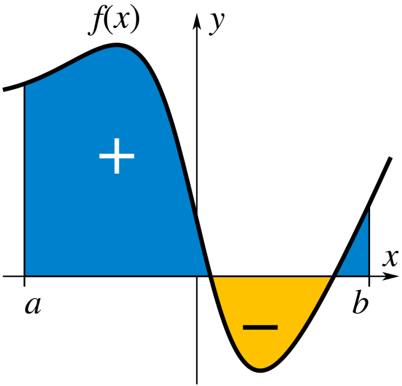
What is...machine learning in mathematics - part 12?

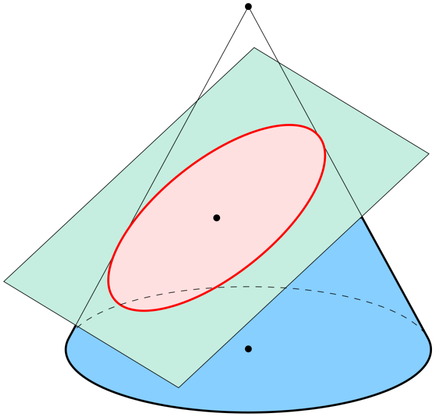
Or: AI and symbolic mathematics

Integration



- Integration = calculation of area as a function
- Solving integral or differential equations is a key problem
- Its also very difficult and many different methods are needed

A first nonexample



- An ellipse is a rather simple shape
- There is no simple way to express the perimeter of the ellipse in terms of elementary functions (algebraic functions, exponential functions *etc.*)
- So we cannot hope to solve integrals in general

Hopeless most of the time

```
Integrate[(5 + 6 x + 4 x^2 + 3 x^3 + 9 x^4) / (1 + x + x^2), x] //
 FullSimplify
```

$$x + 3 \cdot (-1 + x)\, x^2 - \sqrt{3}\, \mathrm{ArcTan}\left[\frac{1 + 2 x}{\sqrt{3}}\right] + \frac{11}{2} \mathrm{Log}\left[1 + x + x^2\right]$$

```
Integrate[(5 + 6 x + 4 x^2 + 3 x^3 + 9 x^4) / Sqrt[(1 + x + x^(1/2))],
  x] // FullSimplify
```

$$\frac{1}{3\,440\,640}\Bigg(2 \sqrt{1 + \sqrt{x} + x}\, \Big(13\,238\,191 + 2 \cdot \big(-5\,902\,549 \sqrt{x} + 5\,418\,164\, x - 580\,104\, x^{3/2} - 975\,552\, x^2 + 1\,342\,080\, x^{5/2} + 729\,600\, x^3 - 1\,827\,840\, x^{7/2} + 1\,720\,320\, x^4\big)\Big) + 10\,372\,005\, \mathrm{ArcSinh}\left[\frac{1 + 2\sqrt{x}}{\sqrt{3}}\right]\Bigg)$$

- Risch's algorithm and friends can compute a variety of integrals
- However, there are huge limitations
- Wannabe theorem Almost no integral has a nice solution

Enter, the theorem

A (transformer) neural network (NN) found

solutions to integrals with ≈95% accuracy

The results were compared with several computer algebra systems and outperformed them by quite some extend

- Here are the results:

	Integration (FWD)	Integration (BWD)	Integration (IBP)	ODE (order 1)	ODE (order 2)
Beam size 1	93.6	98.4	96.8	77.6	43.0
Beam size 10	95.6	99.4	99.2	90.5	73.0
Beam size 50	96.2	99.7	99.5	94.0	81.2

Table 2: **Accuracy of our models on integration and differential equation solving.** Results are reported on a held out test set of 5000 equations. For differential equations, using beam search decoding significantly improves the accuracy of the model.

	Integration (BWD)	ODE (order 1)	ODE (order 2)
Mathematica (30s)	84.0	77.2	61.6
Matlab	65.2	-	-
Maple	67.4	-	-
Beam size 1	98.4	81.2	40.8
Beam size 10	99.6	94.0	73.2
Beam size 50	99.6	97.0	81.0

Table 3: **Comparison of our model with Mathematica, Maple and Matlab on a test set of 500 equations.** For Mathematica we report results by setting a timeout of 30 seconds per equation. On a given equation, our model typically finds the solution in less than a second.

- Something similar works for ODEs

Many solutions!

Hypothesis	Score	Hypothesis	Score
$\frac{9\sqrt{x}\sqrt{\frac{1}{\log(x)}}}{\sqrt{c+2x}}$	-0.047	$\frac{9}{\sqrt{\frac{c\log(x)}{x}+2\log(x)}}$	-0.124
$\frac{9\sqrt{x}}{\sqrt{c+2x}\sqrt{\log(x)}}$	-0.056	$\frac{9\sqrt{x}}{\sqrt{c\log(x)+2x\log(x)}}$	-0.139
$\frac{9\sqrt{2}\sqrt{x}\sqrt{\frac{1}{\log(x)}}}{2\sqrt{c+x}}$	-0.115	$\frac{9}{\sqrt{\frac{c}{x}+2}\sqrt{\log(x)}}$	-0.144
$9\sqrt{x}\sqrt{\frac{1}{c\log(x)+2x\log(x)}}$	-0.117	$9\sqrt{\frac{1}{\frac{c\log(x)}{x}+2\log(x)}}$	-0.205
$\frac{9\sqrt{2}\sqrt{x}}{2\sqrt{c+x}\sqrt{\log(x)}}$	-0.124	$9\sqrt{x}\sqrt{\frac{1}{c\log(x)+2x\log(x)+\log(x)}}$	-0.232

Table 5: Top 10 generations of our model for the first order differential equation $162x\log(x)y' + 2y^3\log(x)^2 - 81y\log(x) + 81y = 0$, generated with a beam search. All hypotheses are valid solutions, and are equivalent up to a change of the variable c. Scores are log-probabilities normalized by sequence lengths.

- Integration algorithms always produce the same answer
- Integration NN can produce equivalent answers (that are differently written)
- This, depending on the context, could be a huge advantage

Thank you for your attention!

I hope that was of some help.