Why you should still do math in "~~the age~~ THIS STAGE of AI"

BMC Advanced
Handout for 240925 – day 1 of 2
Chris Overton

<warmup question>
What are some of the stupidest, scariest, and most useful things you have heard about AI?

(We'll discuss these in a few minutes)

Outline

A view of AI for "math people"
We'll include some actual math, but will also allow ourselves a bit of opinion and polemic...

- Preview of AI as commonly used in 2024
- How reality works (with emphasis on math)
- How AI works
  - gradients, backpropagation
  - vector embeddings (contrast with Taylor & Fourier series)
- Conclusions for Day 1

(In Day 2 next week, we'll cover some techniques in more detail)

Stupid? Scary? Useful?

- As we did in class, please try to think of each of these before reading on for some suggestions

Stupid?

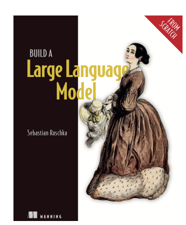
- "What artificial intelligence will never be able to do"
  - Not that unlike "why humans are unlike all the animals"
- Conversely: "we will have self-driving cars in 5 years" (e.g. as said in 1995)
- "Democratizing AI for everyone"
- "AI's effects on the job market will be small/incremental"
- What's stupid about the title for today's talk? (FIXED THIS ONE)
- "Open" AI ~ Google's earlier motto "don't be evil"

Scary?

- X% of news articles are generated by bots
- X% of high school students use AI to cheat on their homework
- X% of people are in love with a bot...
- Claims of "democratizing AI for everyone"
- It's amazing how far AI has come in the past few years
- Fake news, fake reality, weakening links to actual reality
- AI often lies (or at least is biased) - just like news, but cleverer
  - It is being used by those who want to control populations

Useful?

- We're in an "AI revolution" that is starting to be well-documented (despite lots of noise) in culture, news, nonfiction writing, and CODE
- Arxiv papers (besides all the posturing) do let you see how many techniques work, and often how you could replicate their work
- Platforms like HuggingFace let you load, run, and even reverse-engineer many models
- Emerging distilled knowledge like Raschka's "How to build an LLM from scratch" (cover at right, used in some of our examples)
- "AI playgrounds" and "open" models



Preview of AI as commonly used in 2024

- Old stuff: sentiment analysis, image recognition
- New advances are especially in **generative** AI
- LLM's: next token prediction – dramatically better than older RNN's (recurrent neural nets)
- Multi-modal: e.g. stable diffusion, audio & video in & out

Stable diffusion example

Prompt: (Pope Francis) wearing leather jacket is a DJ in a nightclub, mixing live on stage, giant mixing table, 4k resolution, a masterpiece
Negative prompt: white robe, easynegative, bad-hands-5, grainy, low-res, extra limb, poorly drawn hands, missing limb, blurry, malformed hands, blur
Parameters: Steps: 40, Sampler: DDIM, CFG scale: 8.0, Seed: 1639299662, Face restoration, Size: 480x512

for more, see: https://stablediffusion.fr/prompts

Progression of AI models:

1)

- Pass tests like GED high school equivalency, SAT
- Move on to professional tests like medical and law

2)

- Clone voices of famous people (typically without their consent)
- Make fake videos of famous people (often fraudulent or pornographic
- Begin to assist in authoring of movies (hence a subject of recent writer's strike)

How reality works
(with some emphasis on math)

- Approximation 1: Newtonian mechanics
- Approximation 2: quantum mechanics
  - Why an approximation? Because a) inherently probabilistic,
    b) not yet reconciled with other theories (e.g. gravity),
    c) we don't know what most of the universe is made of
- And yet "logic" seems to work perfectly
  - After revised axiom systems "fixed" contradictions early last century
  - A single contradiction threatens to destroy all of math, but yet math seems to survive!
  - Is this just an emergent probabilistic illusion like Newtonian physics? If so, it's a very good one!

Gödel incompleteness theorems

- Essentially: in a "reasonable" logic, "almost all" statements are undecidable
- We won't outline a proof today, but as an analogy could consider the "diagonalization" proof that the cardinality of R (reals) is greater than that of Q (rationals)
- What this could mean: the "vast majority" of math remains to be discovered.
- In particular: even for AI, there's a long way to go in math

How AI works: four levels of optimization

- Calculus 1: max of single-variable function
- Multivariate calculus: gradients
- Neural nets (NN): backpropagation
- Large Language Models (LLM): next token prediction

Backpropagation: the main "secret sauce" of neural nets

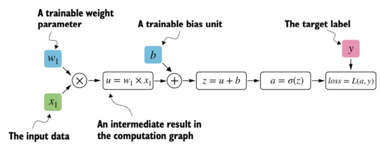
- Neural nets are built with up to billions of trainable weights (as w_1 here.)
- In training, these are adjusted according to how they affect "loss."

Listing A.2 A logistic regression forward pass

```
import torch.nn.functional as F

y = torch.tensor([1.0])
x1 = torch.tensor([1.1])
w1 = torch.tensor([2.2])
b = torch.tensor([0.0])
z = x1 * w1 + b
a = torch.sigmoid(z)
loss = F.binary_cross_entropy(a, y)
```

This import statement is a common convention in PyTorch to prevent long lines of code.
True label
Input feature
Weight parameter
Bias unit
Net input
Activation and output

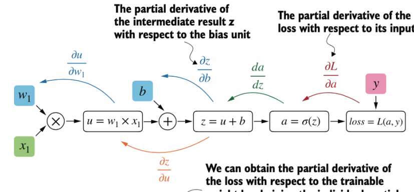


Backpropagation: how weights update

```
loss.backward()
print(w1.grad)
print(b.grad)
```

The outputs are

```
(tensor([-0.0898]),)
(tensor([-0.0817]),)
```



$$\frac{\partial L}{\partial w_1} = \frac{\partial u}{\partial w_1} \times \frac{\partial z}{\partial u} \times \frac{da}{dz} \times \frac{\partial L}{\partial a}$$

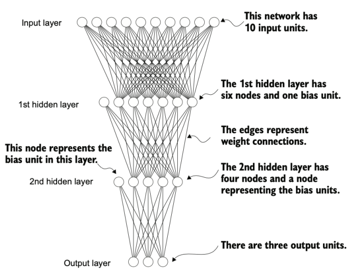
We can obtain the partial derivative of the loss with respect to the trainable weight by chaining the individual partial derivative in the graph.

$$\frac{\partial L}{\partial b} = \frac{\partial z}{\partial b} \times \frac{da}{dz} \times \frac{\partial L}{\partial a}$$

Similar to above, we can compute the partial derivative of the trainable derivative by applying the chain rule.

"Deep" neural nets

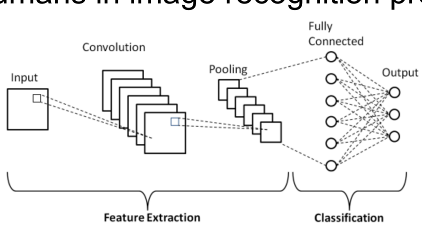
- Training (slow) vs. inference (a single forward pass)
- Basic neural nets can fit basic patterns, after lots of training, which doesn't scale that well with subject complexity



How AI works: vector embeddings, tensors

- Consider how Taylor & Fourier series try to model functions through linear combinations of (up to countably many) similar pieces
- In AI, "pieces" are just collected in vectors (or higher-dimensional "tensors"), often without much internal structure

One major triumph: CNN's ("convolutional") beat humans in image recognition problems



Schematic diagram of a basic convolutional neural network (CNN) architecture [26].

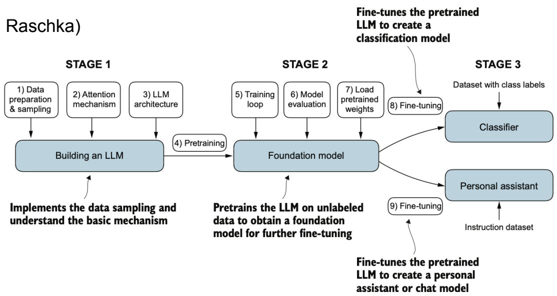
- <application example: triplet loss for biometric identification>

This didn't work so well for RNN's (recurrent)

- One reason: errors can compound, and signals can attentuate as one tries to forecast several steps out.
- But as we discuss next week, the key idea that got LLM's working better than RNN's was "transformers" (involving "attention")
- Currently, some large competitors spend over $100M in model iterations, built roughly as on the next slide

One overview for how to build an LLM

(from Raschka)



Current weak points of LLM's:

what they "couldn't" vs "don't" tell you

- Attempts to safeguard against undesired use adds opaque "safety" layers
- Since weights of very deep models interacts in opaque ways, even model creators often have to retrain largely from scratch for key improvements
- "big AI" is running up against limits in material for training (e.g. Wikipedia, all printed books, collections of user interactions, ...)
- NN outputs are still probabilistic –these must be balanced against compute where the latter is appropriate
  - One key test for models is in "reasoning" and math problems
- <examples discussed in class: twin primes, 1% prime diff, "honesty" metric, LLM's irregular refusal to answer questions>

Conclusions for Day I

- NN are built on several math ideas
  - Understanding these can put you ahead of much of the large (and growing) populations of users & developers
- Recent NN's have so much changed people's thinking that many are tempted to ignore what humanity learned previously
  - Converseley, many NN papers are just new recipe mixtures without even pretense of solid understanding of what is working
- You have both capabilities and strong incentives to understand what current models are producing

Conclusions for Day 1 (II)

- New AI capabilities would seem to increase social wealth. But:
- The "market value" of mediocre thinking has decreased dramatically

  (Because you can just have an AI do it)
- Mediocre thinking is increasing dramatically

  (via consumption of AI answers
  instead of causal understanding)

Does anyone see a problem with this?

Conclusions for Day 1 (III)

- Math appears to work perfectly (i.e. its logic is solid and free of contradictions)
- With current "proof technology", we are doomed to not know "almost all" of math
  - So there is still an infinite amount left to learn & figure out
- NN's can be good at replicating and even using our common logic, but its results are largely probabilistic (unless verified explicitly)
  - So math is infinitely many orders of magnitude more rigid
- As most of the population becomes enslaved to the "convenience" of AI's fictitious reality, interacting with math is one way you can continue to experience actual reality

Topics for next week

- Proof software systems
- Important pieces in NN's:
  attention and transformers
- More on interesting techniques we mentioned today