Announcements

- **Homework 9** due **today** (Tuesday, Nov 29) at 11:59pm PT
- **Project 5** due **Thursday, Dec 1** at 11:59pm PT
  - Q1 unchanged, Q2 & Q3 now have hyperparameters, Q4 is **optional**
- **Homework 10** will be released this week and is **optional**
- Online office hours:
  - 5:00-6:30pm M/W
  - 6:30-8:00pm Tu/Th
- **Final exam** will be held as **planned**
- Thursday lecture will be review session
CS 188: Artificial Intelligence
Applications & Conclusion

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at http://ai.berkeley.edu.]
Recap of Key Neural Network Ideas

Optimize probability of label given input
\[
\max_w \ ll(w) = \max_w \sum_i \log P(y^{(i)} | x^{(i)}; w)
\]

Continuous optimization

**Gradient ascent**: take step in steepest uphill direction
Backpropagation computes gradient efficiently (out of scope)

Deep neural nets

Last layer = logistic regression
Many layers before this last layer: learns “features” of input
Universal function approximation theorem:

a large enough 2 layer neural network can represent any continuous function with arbitrary accuracy
Today

- Applications of supervised deep learning
- Brief overview of unsupervised learning
- Frontier applications of AI
  - Language and text models
  - Image-language models
  - Reinforcement learning (Go and Robotics)
- AI ethics considerations
How well does deep learning work?
Computer Vision
Object Detection and Image Classification
Traditional Computer Vision

- Manual Feature Design

Image

Histogram of Gradients (HoG)

[HoG: Dalal and Triggs, 2005]
Image Classification: Performance

ImageNet Error Rate 2010-2014

Error Rate

2010 2011 2012 2013 2014

Traditional CV

graph credit Matt Zeiler, Clarifai
Image Classification: Performance

ImageNet Error Rate 2010-2014

Graph credit Matt Zeiler, Clarifai
Image Classification: Performance

ImageNet Error Rate 2010-2014

- Traditional CV
- Deep Learning

Error Rate:
- 79%
- 60%
- 40%
- 20%
- 7%

Year:
- 2010
- 2011
- 2012
- 2013
- 2014

AlexNet

Graph credit Matt Zeiler, Clarifai
Image Classification: Performance

ImageNet Error Rate 2010-2014

- Traditional CV
- Deep Learning

Error Rate

2010 2011 2012 2013 2014

AlexNet

graph credit Matt Zeiler, Clarifai
Image Classification: Performance

ImageNet Error Rate 2010-2014

- Traditional CV
- Deep Learning

Graph credit: Matt Zeiler, Clarifai

AlexNet
Image Classification: Performance

Credit: paperswithcode.com
Image Segmentation
Speech Recognition

TIMIT Speech Recognition

Error Rate vs. Year

- Traditional
- Deep Learning

Graph credit: Matt Zeiler, Clarifai
Machine Translation

Google Neural Machine Translation

Encoder

\[ e_0 \rightarrow e_1 \rightarrow e_2 \rightarrow e_3 \rightarrow e_4 \rightarrow e_5 \rightarrow e_6 \]

Decoder

\[ d_0 \rightarrow d_1 \rightarrow d_2 \rightarrow d_3 \]

Knowledge \rightarrow is \rightarrow power \rightarrow <end>
Google announced today that it has made energy produced by wind farms more viable using the artificial intelligence software of its London-based subsidiary DeepMind. By using DeepMind’s machine learning algorithms to predict the wind output from the farms, Google uses for its green energy initiatives, the company says it can now schedule set deliveries of energy output, which are more valuable to the grid than standard, non-time-based deliveries.
‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures

Google’s deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Ewen Callaway
T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

T1049 / 6y4f
93.3 GDT
(adhesin tip)
CASP 2020 Competition

Median Free-Modelling Accuracy

Credit: DeepMind
Change in Programming Paradigm?

**Traditional Programming:**
program by writing lines of code

```plaintext
If Number = 1 Then
  Count1 = Count1 + 1
Else
  If Number = 2 Then
    Count2 = Count2 + 1
  Else
    If Number = 3 Then
      Count3 = Count3 + 1
    Else
      CountX = CountX + 1
    End If
  End If
End If
```

Poor performance on AI problems

**Deep Learning (“Software 2.0”):**
program by providing data

Success!
Today

- Applications of supervised deep learning
- Brief overview of unsupervised learning
- Frontier applications of AI
  - Language and text models
  - Image-language models
  - Reinforcement learning (Go and Robotics)
- AI ethics considerations
Deep Unsupervised Learning

- **Key hypothesis:**
  
  **Task 1** IF neural network smart enough to predict:
  - Next frame in video
  - Next word in sentence
  - Generate realistic images
  - "Translate" images
  - ...

  **Task 2** THEN same neural network is ready to do Deep Supervised Learning from very small data-set
Transfer from Unsupervised Learning

Task 1 = unsupervised

Task 2 = real task
Example Setting

Task 1 = predict next word

Task 2 = predict sentiment
A train carriage containing controlled nuclear materials was stolen in Cincinnati today. Its whereabouts are unknown.

The incident occurred on the downtown train line, which runs from Covington and Ashland stations. In an email to Ohio news outlets, the U.S. Department of Energy said it is working with the Federal Railroad Administration to find the thief.

‘The theft of this nuclear material will have significant negative consequences on public and environmental health, our workforce and the economy of our nation,’ said Tom Hicks, the U.S. Energy Secretary, in a statement. ‘Our top priority is to secure the theft and ensure it doesn’t happen again.’

The stolen material was taken from the University of Cincinnati’s Research Triangle Park nuclear research site, according to a news release from Department officials.

The Nuclear Regulatory Commission did not immediately release any information. According to the release, the U.S. Department of Energy’s Office of Nuclear Material Safety and Security is leading that team’s investigation.

‘The safety of people, the environment and the nation’s nuclear stockpile is our highest priority,’ Hicks said. ‘We will get to the bottom of this and make no excuses.’
Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I’m not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a tremendous price for the privilege of having the world’s most advanced and efficient recycling system. Recycling is a huge, colossal waste of time, energy, money, and resources.
Unsupervised Sentiment Neuron

This is one of Crichton's best books. The characters of Karen Ross, Peter Elliot, Munro, and Amy are beautifully developed and their interactions are exciting, complex, and fast-paced throughout this impressive novel. And about 99.8 percent of that got lost in the film. Seriously, the screenplay AND the directing were horrendous and clearly done by people who could not fathom what was good about the novel. I can't fault the actors because frankly, they never had a chance to make this turkey live up to Crichton's original work. I know good novels, especially those with a science fiction edge, are hard to bring to the screen in a way that lives up to the original. But this may be the absolute worst disparity in quality between novel and screen adaptation ever. The book is really, really good. The movie is just dreadful.

[Radford et al, 2017]
<table>
<thead>
<tr>
<th>DATASET</th>
<th>METRIC</th>
<th>OUR RESULT</th>
<th>PREVIOUS RECORD</th>
<th>HUMAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winograd Schema Challenge</td>
<td>accuracy (+)</td>
<td>70.70%</td>
<td>63.7%</td>
<td>92%+</td>
</tr>
<tr>
<td>LAMBADA</td>
<td>accuracy (+)</td>
<td>63.24%</td>
<td>59.23%</td>
<td>95%+</td>
</tr>
<tr>
<td>LAMBADA</td>
<td>perplexity (-)</td>
<td>8.6</td>
<td>99</td>
<td>~1-2</td>
</tr>
<tr>
<td>Children’s Book Test Common Nouns</td>
<td>accuracy (+)</td>
<td>93.30%</td>
<td>85.7%</td>
<td>96%</td>
</tr>
<tr>
<td>(validation accuracy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children’s Book Test Named Entities</td>
<td>accuracy (+)</td>
<td>89.05%</td>
<td>82.3%</td>
<td>92%</td>
</tr>
<tr>
<td>(validation accuracy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penn Tree Bank</td>
<td>perplexity (-)</td>
<td>35.76</td>
<td>46.54</td>
<td>unknown</td>
</tr>
<tr>
<td>WikiText-2</td>
<td>perplexity (-)</td>
<td>18.34</td>
<td>39.14</td>
<td>unknown</td>
</tr>
</tbody>
</table>
Unsupervised Learning in Vision

Task 1 = fill in a patch

Task 2 = predict cat vs. dog
Pre-Trained Model (SimCLR) + Linear Classifier

[Chen et al, 2020]
Pre-Training and Fine-Tuning

1. **Pre-Train**: train a large model with a lot of data on a supervised *pretext* task
   - Predict next word / patch of image
   - Predict missing word / patch of image
   - Predict if two images are related (contrastive learning)

2a. **Fine-Tune**: continue training the same model on task you care about

2b. **Prompt**: Or directly ask the model?
The three settings we explore for in-context learning

**Zero-shot**
The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
1 Translate English to French: task description
2 cheese => .................................. prompt
```

**One-shot**
In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

```
1 Translate English to French: task description
2 sea otter => loutre de mer example
3 cheese => .................................. prompt
```

**Few-shot**
In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

```
1 Translate English to French: task description
2 sea otter => loutre de mer examples
3 peppermint => menthe poivrée
4 plush giraffe => girafe peluche
5 cheese => .................................. prompt
```

**Fine-tuning (not used for GPT-3)**
The model is trained via repeated gradient updates using a large corpus of example tasks.

```
1 sea otter => loutre de mer example #1
2 peppermint => menthe poivrée example #2
3 plush giraffe => girafe peluche example #N
```

[Language Models are Few-Shot Learners, Brown et al, 2020]
### Prompt without training:

- **Control desired output via examples**

<table>
<thead>
<tr>
<th>Standard Prompting</th>
<th>Chain of thought prompting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Example Input</strong></td>
<td>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</td>
</tr>
<tr>
<td><strong>Example Output</strong></td>
<td>Q: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.</td>
</tr>
<tr>
<td><strong>Prompt</strong></td>
<td>The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</td>
</tr>
<tr>
<td><strong>Model Response</strong></td>
<td>The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9.</td>
</tr>
</tbody>
</table>

**Question:** A line parallel to $y = 4x + 6$ passes through $(5, 10)$. What is the $y$-coordinate of the point where this line crosses the $y$-axis?

**Model output:** The line is parallel to $y = 4x + 6$, which means that it has the same slope as $y = 4x + 6$. Since the line passes through $(5, 10)$, we can use the point-slope form of a line to find the equation of the line:

$$y - 10 = 4(x - 5)$$
$$y - 10 = 4x - 20$$
$$y = 4x - 10$$

Now that we have the equation of the line, we can find the $y$-coordinate of the point where the line crosses the $y$-axis by substituting $x = 0$ into the equation:

$$y = 4 \cdot 0 - 10 = -10.$$
Applications of Text Models: Mathematics

QUESTION Two simple pendulums $A$ and $B$ consist of identical masses suspended from strings of length $L_A$ and $L_B$, respectively. The two pendulums oscillate in equal gravitational fields. If the period of pendulum $B$ is twice the period of pendulum $A$, which of the following is true of the lengths of the two pendulums? Which of the following is the right choice? Explain your answer: (A) $L_B = 4L_A$ (B) $L_B = 2L_A$ (C) $L_B = L_A$ (D) $L_B = 1/2L_A$

MODEL ANSWER The period of a simple pendulum is given by the formula

$$T = 2\pi \sqrt{\frac{L}{g}}.$$

If the period of pendulum $B$ is twice the period of pendulum $A$, we have

$$2T_A = T_B \Rightarrow 2(2\pi \sqrt{\frac{L_A}{g}}) = 2\pi \sqrt{\frac{L_B}{g}} \Rightarrow \frac{L_B}{L_A} = 4.$$

Final Answer: The final answer is (A).

REFERENCE ANSWER A

PROBLEM SOURCE MMLU college physics

EVALUATION RESULT Correct

[https://minerva-demo.github.io/]

[Solving Quantitative Reasoning Problems with Language Models, Lewkowycz et al, 2022]
Applications of Text Models: Mathematics

- **Incorrect Answer:** reasoning mistake (but looks worryingly plausible!)

**Question:** At each basketball practice last week, Jenny made twice as many free throws as she made at the previous practice. At her fifth practice she made 48 free throws. How many free throws did she make at the first practice?

**Model output:** At the first practice, Jenny made \(\frac{48}{2} = 24\) free throws. Final Answer: The final answer is 24.
Applications of Text Models: Programming

```python
def incr_list(l: list):
    """Return list with elements incremented by 1."
    >>> incr_list([1, 2, 3])
    [2, 3, 4]
    >>> incr_list([5, 3, 5, 2, 3, 3, 9, 0, 123])
    [6, 4, 6, 3, 4, 4, 10, 1, 124]
    """
    return [i + 1 for i in l]

def solution(lst):
    """Given a non-empty list of integers, return the sum of all of the odd elements that are in even positions."

    Examples
    solution([5, 8, 7, 1]) => 12
    solution([3, 3, 3, 3]) => 9
    solution([30, 13, 24, 321]) => 0
    """
    return sum(lst[i] for i in range(0, len(lst)) if i % 2 == 0 and lst[i] % 2 == 1)
```

[Evaluating Large Language Models Trained on Code, Chen et al, 2021]
Applications of Image-Text Models

[Hierarchical Text-Conditional Image Generation with CLIP Latents, Ramesh et al, 2022]
Applications of Image-Text Models

Hierarchical Text-Conditional Image Generation with CLIP Latents, Ramesh et al, 2022
Applications of Image-Text Models

Hierarchical Text-Conditional Image Generation with CLIP Latents, Ramesh et al, 2022

TEXT DESCRIPTION
An astronaut Teddy bears A bowl of soup
mixing sparkling chemicals as mad scientists shopping for groceries working on new AI research
as a 1990s Saturday morning cartoon as digital art in a steampunk style

DALL·E 2
Applications of Image-Text Models

Hierarchical Text-Conditional Image Generation with CLIP Latents, Ramesh et al, 2022

An astronaut Teddy bears A bowl of soup mixing sparkling chemicals as mad scientists shopping for groceries working on new AI research as kids' crayon art on the moon in the 1980s underwater with 1990s technology
Applications of Image-Text Models

Hierarchical Text-Conditional Image Generation with CLIP Latents, Ramesh et al, 2022

An astronaut Teddy bears A bowl of soup mixing sparkling chemicals as mad scientists shopping for groceries working on new AI research as kids' crayon art on the moon in the 1980s underwater with 1990s technology
Applications of Image-Text Models

Hierarchical Text-Conditional Image Generation with CLIP Latents, Ramesh et al, 2022

TEXT DESCRIPTION

An astronaut  Teddy bears  A bowl of soup

that is a portal to another dimension  that looks like a monster  as a planet in the universe

as digital art  in the style of Basquiat  drawn on a cave wall

DALL·E 2
Applications of Image-Text Models: Visual Dialogue
Alayrac et al, 2022

This is a picture of two teddy bears on the moon.
- What are they doing?
- They are having a conversation.
- What object are they using?
- It looks like a computer.
- Is this surprising?
- Yes, it is surprising.
- Why is this picture surprising to you?
- I think it is surprising because teddy bears are not usually found on the moon.

What is the common thing about these three images?
- They are all flamingos.

What is the difference between these three images?
- The first one is a cartoon, the second one is a real flamingo, and the third one is a 3D model of a flamingo.

This is an apple with a sticker on it.
- What does the sticker say?
- The sticker says "iPod".
- Where is the photo taken?
- It looks like it's taken in a backyard.
- Do you think it is printed or handwritten?
- It looks like it's handwritten.
- What color is the sticker?
- It's white.
Applications of Deep Reinforcement Learning: Go
Applications of Deep Reinforcement Learning: Go

- Just MiniMax Search?
Exhaustive Search?

Credit: DeepMind
Reducing depth with value network
Value network

Credit: DeepMind
Reducing breadth with policy network

Credit: DeepMind
Policy network

Move probabilities

Position

\[ p(a|s) \]

Credit: DeepMind
AlphaGo: neural network training pipeline

Human expert positions → Supervised Learning policy network → Reinforcement Learning policy network → Self-play data → Value network

N-Layer Neural Network

Policy Search

- Simplest policy search:
  - Start with an initial linear value function or Q-function
  - Nudge each feature weight up and down and see if your policy is better than before

- Problems:
  - How do we tell the policy got better?
  - Need to run many sample episodes!
  - If there are a lot of features, this can be impractical

- Better methods exploit lookahead structure, sample wisely, change multiple parameters...

Credit: DeepMind
Personal Robotics
Reinforcement Learning in Robotics

[Levine*, Finn*, Darrell, Abbeel, JMLR 2016]
Robotics + Language

[Do As I Can, Not As I Say: Grounding Language in Robotic Affordances, Ahn et al, 2022]
Highly structured environments
Pre-programmed, deterministic

Traditional Robotic Automation

Next Generation: AI Robotic Automation
Less structured environments
Intelligent, reactive behaviors
Today

- Applications of supervised deep learning
- Brief overview of unsupervised learning
- Frontier applications of AI
  - Language and text models
  - Image-language models
  - Reinforcement learning (Go and Robotics)
- AI ethics considerations
AI Ethics Ever More Important

- **Why?**
  - AI is making decisions, at scale
  - Any kind of issues (e.g. bias or malignant use) could significantly affect people

- **Many open questions:**
  - Who is responsible?
  - How to diagnose and prevent?
Some experts you’ll want to learn from

Prof. Rediet Abebe
UC Berkeley

Prof. Rachel Thomas
Queensland Univ.

Prof. Ruha Benjamin
Princeton Univ.

Dr. Timnit Gebru
DAIR

Prof. Moritz Hardt
MPI / UC Berkeley
Some Key AI Ethics Topics

- Disinformation
- Bias and fairness
- Privacy and surveillance
- Metrics
- Algorithmic colonialism

Source: Rachel Thomas (@math_rachel)
What will be AI’s impact in the future?

- You get to determine that!
- As you apply AI
- As researchers / developers
- As auditors and regulators
- As informed public voices
Where to Go Next?
Congratulations, you’ve seen the basics of modern AI

... and done some amazing work putting it to use!

How to continue:

- Machine learning: cs189, cs182, stat154
- Data Science: data 100, data 102
- Data / Ethics: data c104
- Probability: ee126, stat134
- Optimization: ee127
- Cognitive modeling: cog sci 131
- Machine learning theory: cs281a/b
- Computer vision: cs280
- Reinforcement Learning: cs285
- Robotics: cs287, cs287h
- NLP: cs288

... and more; ask if you’re interested
Lightweight Opportunities to Keep Learning

- Andrew Ng weekly newsletter:
  The Batch: [https://www.deeplearning.ai/thebatch/](https://www.deeplearning.ai/thebatch/)

- Jack Clark weekly newsletter:
  Import AI: [https://jack-clark.net/](https://jack-clark.net/)

- Rachel Thomas AI Ethics course:
  Course website: [ethics.fast.ai](https://ethics.fast.ai)

- Pieter Abbeel podcast:
  The Robot Brains Podcast: [https://therobotbrains.ai](https://therobotbrains.ai)
That’s It!

- Help us out with some course evaluations
- Good luck on the exam and have a great winter break!