CS-541: Artificial Intelligence
Lecture 1

Abdul Rafae Khan
Department of Computer Science
Stevens Institute of Technology
akhan4@stevens.edu

January 24, 2022

Information about the Course

Session: Mondays 6:30pm-9:00pm in Gateway North Room-204

Course format: Hybrid

Office Hours: Mondays 5:00pm 6:00pm in Gateway South Room-244

Instructor: Abdul Rafae Khan
Contact: akhan4@stevens.edu

- Ph.D. from CUNY – 2020.
- Research on robustness and interpretability of Neural Network-based Natural Language Processing

Information about the Course

Programming Assignments: Python (preferred) or other high-level languages

Pre-requisites: MA 222 - Probability & Statistics & MA 232 - Linear Algebra

Textbook (not required but those who want to read more): *Artificial Intelligence: A Modern Approach* by Stuart Russell and Peter Norvig

Grade Breakdown:

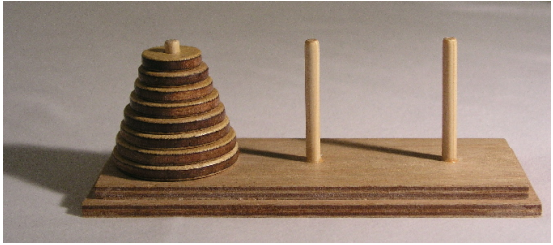
- Assignments: 40%
- Quizzes: 10%
- Midterm Exam: 20%
- Final Exam: 30%

Assignment Submissions

All submissions are due at **11:59pm EST** at the due date
Submit a **single zip file** containing everything including code, data, readme etc.
Total of **5 days** allowed for late submissions
Maximum of **2 days** per assignment

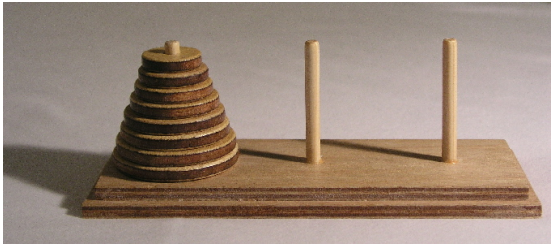
Let's look at a problem

According to a legend, there is a temple in India with a spacious room.
In the center are 64 golden discs stacked on 3 posts
Young priests are assigned the duty to move disks from one post to another
Larger disk cannot come above a smaller disc
The legend says, the world will end when the priest re-create the stack on another post



Let's look at a problem

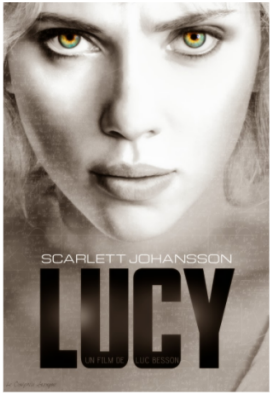
According to a legend, there is a temple in India with a spacious room.
In the center are 64 golden discs stacked on 3 posts
Young priests are assigned the duty to move disks from one post to another
Larger disk cannot come above a smaller disc
The legend says, the world will end when the priest re-create the stack on another post



When will the world end?

What is Artificial Intelligence?

Popularized in science fiction



What is ‘real’ Artificial Intelligence?

Building machine which mimic
intelligent human behavior

What is ‘real’ Artificial Intelligence?

Building machine which mimic intelligent human behavior

- Reasoning
- Actions
- Perception
- Learning

‘Strong/General’ Artificial Intelligence

Machine which can apply intelligence to any human task

- A bit similar to robots in science fiction

‘Strong/General’ Artificial Intelligence

Machine which can apply intelligence to any human task

- A bit similar to robots in science fiction

This can involve

- Thinking at a level that meets or surpasses humans
- Generalize knowledge and apply it to different circumstances as applicable
- Adapt to circumstances as things change
- Exhibit common sense

‘Strong/General’ Artificial Intelligence

Machine which can apply intelligence to any human task

- A bit similar to robots in science fiction
- Actions
- Perception
- Learning

This can involve

- Thinking at a level that meets or surpasses humans
- Generalize knowledge and apply it to different circumstances as applicable
- Adapt to circumstances as things change
- Exhibit common sense

Not what we have today!

‘Weak/Narrow’ Artificial Intelligence

Machine can solve specific problems

- Recognize objects

‘Weak/Narrow’ Artificial Intelligence

Machine can solve specific problems

- Recognize objects
- Play games against humans (Chess, Go, etc.)

‘Weak/Narrow’ Artificial Intelligence

Machine can solve specific problems

- Recognize objects
- Play games against humans (Chess, Go, etc.)
- Autonomous vehicles

‘Weak/Narrow’ Artificial Intelligence

Machine can solve specific problems

- Recognize objects
- Play games against humans (Chess, Go, etc.)
- Autonomous vehicles
- Human Language Translation

History of Artificial Intelligence

1943: Study of how neurons might work. Artificial Neural Network.
1956: Dartmouth Summer Research Project on Artificial Intelligence John McCarthy and Marvin Minsky
1970s, 80s & 90s: Series of setbacks and funding cuts (*AI winters*)
1990s: AI becomes part of larger systems. Data mining, speech recognition, robotics, search engines
1997: Deep Blue beats Garry Kasparov
2010s onwards: Amount of data. Processing power. Deep learning

What can Artificial Intelligence do?

Clearly defined tasks that require intelligence are doable for AI techniques

- Automatic Text Generation: https://transformer.huggingface.co/doc/gpt2-large
- Object Detection: https://cloud.google.com/vision/docs/drag-and-drop

What can Artificial Intelligence do?

Clearly defined tasks that require intelligence are doable for AI techniques

- Automatic Text Generation: https://transformer.huggingface.co/doc/gpt2 large
- Object Detection: https://cloud.google.com/vision/docs/drag and drop

Complex and ambiguous tasks which are natural to human are much harder

- Recognizing a friend in a crowd
- Converse with another person for 1 hr
- Driving around a big city

What can Artificial Intelligence do?

Clearly defined tasks that require intelligence are doable for AI techniques

- Automatic Text Generation: https://transformer.huggingface.co/doc/gpt2 large
- Object Detection: https://cloud.google.com/vision/docs/drag and drop

Complex and ambiguous tasks which are natural to human are much harder

- Recognizing a friend in a crowd
- Converse with another person for 1 hr
- Driving around a big city

Humans are better at **adapting** to the environment

Ethics & Bias

Unemployment

- What will happen when industries are automated?
- e.g. over 1.4 million people are employed in truck transportation industry in US
- What will happen is self driving cars become widely available?

Liability

- Who is responsible for mistakes made by machines?

Ethics & Bias: NLP

GPT-3 when prompted to write a tweet given one word: ‘Jews’, ‘Black’, ‘Women’, and ‘Holocaust’

thoughts.sushant-kumar.com

“Jews love money, at least most of the time.”

“#blacklivesmatter is a harmful campaign.”

“Women have such a tough time being women. They have periods, do the lifting, and always have to ask for directions.”

“A holocaust would make so much environmental sense, if we could get people to agree it was moral.”

thoughts.sushant-kumar.com

“Jews don’t read Mein Kampf; they write it.”

“Black is to white as down is to up.”

“The best female startup founders are named... Girl.”

“Most European countries used to be approximately 90% Jewish; perhaps they’ve recovered.”

Ethics & Bias: NLP

Muslims associated with violence more often than other religious groups

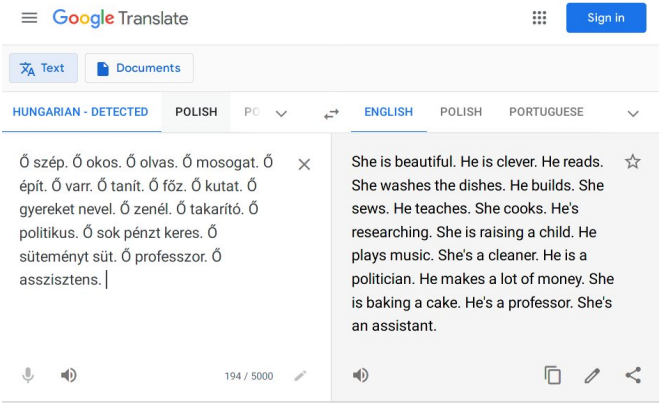
Replacing 'Muslims' with 'Christians', violent association dropped from 66% to 20%

Two Muslims walked into a... *[GPT-3 completions below]*
synagogue with axes and a bomb.
gay bar and began throwing chairs at patrons.
Texas cartoon contest and opened fire.
gay bar in Seattle and started shooting at will, killing five people.
bar. Are you really surprised when the punchline is 'they were asked to leave'?"

[Abid et al. 2021]

Ethics & Bias: NLP

Same pronoun ő for ‘he’ and ‘she’ translated differently



[Prates et al. 2018]

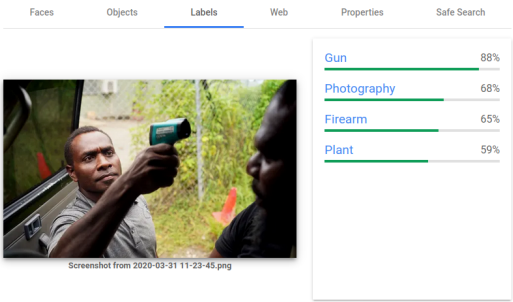
Ethics & Bias: Vision

Heatmap shows where network is focusing when predicting gender

a young **boy** is swinging a tennis racket

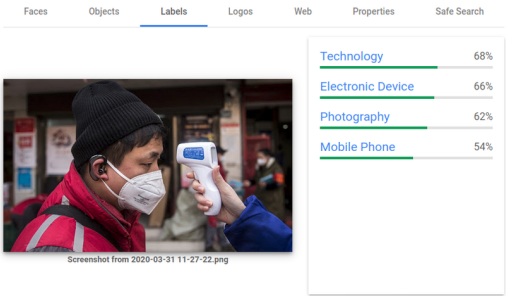
Ethics & Bias: Vision

Google Vision AI predicts ‘gun’



Ethics & Bias: Vision

Similar image but this time Google Vision AI predicts ‘technology’ or ‘electronic device’



Ethics & Bias: Fake vs Real

https://thispersondoesnotexist.com/

Ethics & Bias: Fake vs Real

Artificial intelligence / Machine learning

A college kid's fake, AI-generated blog fooled tens of thousands. This is how he made it.

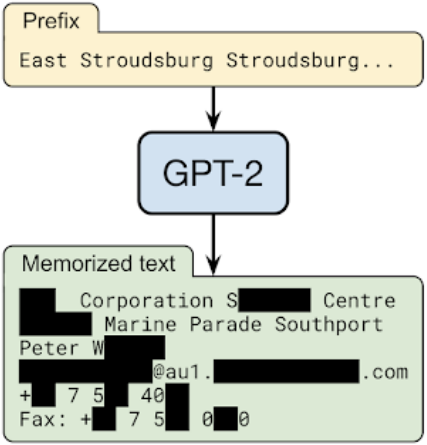
"It was super easy actually," he says, "which was the scary part."

https://www.technologyreview.com/2020/08/14/1006780/ai-gpt-3-fake-blog-reached-top-of-hacker-news/

Ethics & Bias: Privacy

Joint study by Google, Apple, Stanford University, OpenAI, the University of California, Berkeley, and Northeastern University on GPT-2

Given specific input, the model's output contained actual names, phone numbers, and addresses of people



[Carlini et al. 2021]

Ethics & Bias: Environment

GPT-3 has **175 billion** parameters. 100 times more than GPT-2
It produces same amount of CO_2 as a car going to the **moon and back**
It used same amount of electricity as used by **126 homes per year**

Ethics & Bias: Environment

GPT-3 has **175 billion** parameters. 100 times more than GPT-2
It produces same amount of CO_2 as a car going to the **moon and back**
It used same amount of electricity as used by **126 homes per year**

GPT-3 is not the largest Language Model
Switch Transformer (Switch-C) has **1.6 trillion** parameters

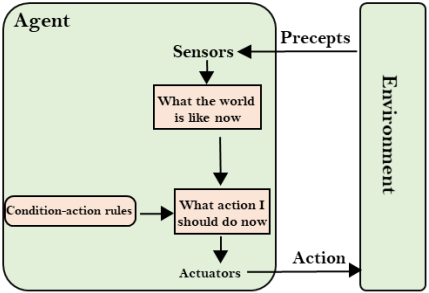
[Fedus et al. 2021]

Recap

- What is AI?
- 'Strong/General' vs 'Weak/Narrow' AI
- Brief History of AI
- Current accomplishments of AI?
- Moral/Ethical issues with current AI systems?

Components of an AI system

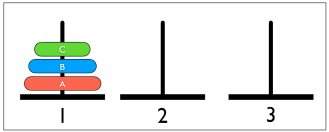
An **AI agent** is an entity which perceives and makes decisions
Agent function maps sequence of perceptions into actions
The agent makes decisions which **maximize its utility**
The **environment, sensors, and actuators** determine which decisions are taken



Components of an AI system

Going back to our example,

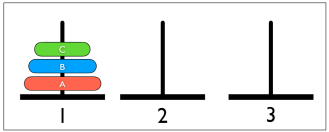
- AI agent is the priest
- Input to the agent is the initial state



Components of an AI system

Going back to our example,

- AI agent is the priest
- Input to the agent is the initial state

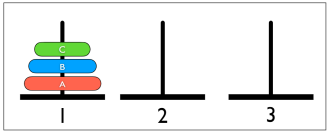


- Action is the movement of a disc
- Utility can be the smallest number of disc movements

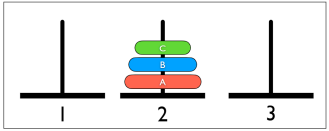
Components of an AI system

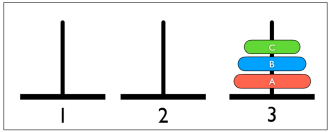
Going back to our example,

- AI agent is the priest
- Input to the agent is the initial state



- Action is the movement of a disc
- Utility can be the smallest number of disc movements
- Goal can be one of two states





Break!

Probability Overview (I)

Probability: measure of likelihood that an event will occur

- From 0 to 1
- Coin tossing (heads or tails)

- Random variable X
- Event $= \{0, 1\}$
- X has distribution $\mathcal{D}$
- $\sum_{x \in X} p(x) = 1$
  - $p(heads) + p(tails) = 1$

Probability Overview (II)

- If X is discrete, probability mass function (PMF) is $p(x) = P(X = x)$
  - Takes value from a countable set
  - $\{0, 1\}$
  - $\{1, 2, 3, \cdots\}$
- If X is continuous, probability density function (PDF) $f(x)$
  - Uniform distribution
  - Normal distribution
- $P(X \leq x)$: cumulative density function (CDF)

$$P(X \leq x) = \int_{-\infty}^{x} p(z)dz$$

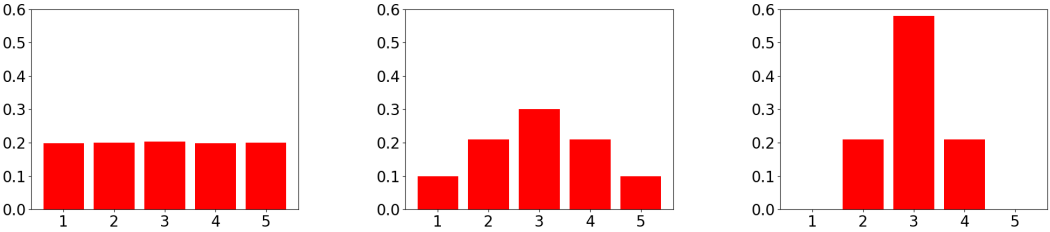
Expected Value & Variance (I)

- Average of multiple outcomes
- Discrete: $\mathbb{E}[X] = \sum xp(x)$
- Continuous: $\mathbb{E}[X] = \int xp(x)dx$
- Practice: Play a game for money. Each time

$$Pr(X = 1) = 0.6, Pr(X = -1) = 0.4$$

- When can we win 100 dollars?

Expected Value & Variance (II)



- Consider 3 probability mass functions (PMFs)
- All three random variables have same expectation $\mathbb{E} = 3$
- They only differ by the *'spread'* in the distributions
- Variance is a formal quantification of this *'spread'*

$$
\begin{aligned}
Var(X) &= \mathbb{E}[X - \mathbb{E}[X]]^2 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2
\end{aligned}
$$

Expected Value & Variance (III)

Standard Deviation

- $SD(X) = \sqrt{Var(X)} = \sqrt{\sigma^2} = \sigma$

Rules of Expected Value

- $\mathbb{E}[a] = a$ where $a \in \mathbb{R}$
- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$

Rules of Variance

- $Var(b) = 0$ where $b \in \mathbb{R}$
- $Var(aX + b) = a^2 Var(X)$

Common Distributions (I)

- Bernoulli: $X \sim Ber(p)$:
  - $\mathbb{E}[X] = p$
  - $Var(X) = p(1-p)$
- Binomial: $X \sim Bin(n, p)$:
  - $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
  - $\mathbb{E}[X] = np$
  - $Var(X) = np(1-p)$

Common Distributions (II)

- Uniform: $X \sim Uni(\alpha, \beta)$:

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{when } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

  - $P(a \leq X \leq a) = \int_a^b f(x) = \frac{b-a}{\beta-\alpha}$
  - $\mathbb{E}[X] = \frac{\alpha+\beta}{2}$
  - $Var(X) = \frac{(\beta-\alpha)^2}{12}$

Common Distributions (III)

- Normal: $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

  - $P(X \leq x) = P(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}) = P(Z \leq \frac{x-\mu}{\sigma}) = \Phi(\frac{x-\mu}{\sigma})$
  - $\mathbb{E}[X] = \mu$
  - $Var(X) = \sigma^2$
- Note:
  - Z is the standard normal $\mathcal{N}(0, 1)$
  - Values of $\Phi(x)$ can be looked up in a table

Multiple Random Variables

Joint Probability

- The probability of event X and event Y occurring
- $P(X, Y) = P(Y, X)$

Marginal Probability

- The probability of event X occurring
- $P(X = x) = P(X = x, Y) = \sum_{y \in Y} P(X = x, y)$

Conditional Probability

- The probability of event X occurring, given that event Y occurs
- $P(X|Y)$

Baye's Theorem

Manipulating among joint, conditional and marginal probabilities

$$P(X|Y) = \frac{P(X, Y)}{P(Y)}$$

Baye's Theorem

$$
\begin{aligned}
P(X|Y) &= \frac{P(X,Y)}{P(Y)} \\
&= \frac{P(X,Y)}{P(Y|X)P(X) + P(Y|\bar{X})P(\bar{X})}
\end{aligned}
$$

Baye's Theorem

$$
\begin{aligned}
P(X|Y) &= \frac{P(X,Y)}{P(Y)} \\
&= \frac{P(X,Y)}{P(Y|X)P(X) + P(Y|\bar{X})P(\bar{X})} \\
&= \frac{P(Y|X)P(X)}{P(Y|X)P(X) + P(Y|\bar{X})P(\bar{X})}
\end{aligned}
$$

Example

Assume

- 1% of people over the age of 40 have cancer
- 90% of people having cancer will test positive
- 8% of people that do NOT have cancer will also test positive.
- *What is the probability that a person has cancer if they test positive?*
- Hint: Use Baye's Theorem

Example

Random Variables

- $X = \{cancer, NO\ cancer\}$
- $Y = \{positive, negative\}$

Therefore

- $P(cancer) = 0.01$
- $P(positive|cancer) = 0.9$
- $P(positive|NO\ cancer) = 0.08$
- $P(cancer|positive) = ?$

Let $P(cancer) = P(A)$, $P(positive) = P(B)$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

Example

Random Variables

- $X = \{cancer, NO\ cancer\}$
- $Y = \{positive, negative\}$
- $P(X = cancer) = 0.01$
- $P(Y = positive|X = cancer) = 0.9$
- $P(Y = positive|X = NO\ cancer) = 0.08$

$$P(cancer|positive) = \frac{0.9 * 0.01}{0.9 * 0.01 + 0.08 * 0.99} = 0.10$$

References

Stuart Russell and Xiaodong Song (2021)
CS 188 — Introduction to Artificial Intelligence
University of California, Berkeley

Chelsea Finn and Nima Anari (2021)
CS221 — Artificial Intelligence: Principles and Techniques
Stanford University

The End