CS 540-1: Introduction to Artificial Intelligence
Homework Assignment # 1

Assigned: 9/6
Due: 9/13 before class

Hand in your homework:

This homework includes only written questions, but you will need to submit **two** files:

- For questions 1 and 2 please submit a single **hw1.pdf** file – we recommend latex, but you can use anything that can render nice math.
- For the rejoinder please write your essay in **plain text format** and name it **rejoinder.txt**.

Go to UW Canvas, choose your CS540-1 course, choose Assignment, click on Homework 1: this is where you submit your files.

Question 1: Math Preparation Self Test [45 points, 9 each]

Modern AI requires a certain degree of math maturity. While it is impractical for CS 540 to require too many math courses as prerequisites, you are expected to have good working knowledge of elementary concepts in calculus, probability, statistics, and linear algebra. Solve the following questions. **Be sure to show your steps.** If most questions sound foreign to you, you may have difficulty understanding the lectures.

Separately, you need to be fluent in Java programming to do some future homeworks.

1. Let a, m, n be positive integers. Express

$$\sum_{n=0}^{m} \sum_{i=0}^{n} a^i$$

in closed form, without the summations. This is the complexity of iterative deepening search.

A useful identity here is for $b \neq 1$:

$$\sum_{i=j}^{n} b^i = \frac{b^j - b^{n+1}}{1 - b}$$

How to see that? Let $F = \sum_{i=j}^{n} b^i$ and notice that $F - bF = b^j - b^{n+1}$ and rearrange.

From there, the computation proceeds by applying it twice:

$$\begin{aligned}\sum_{n=0}^{m}\sum_{i=0}^{n} a^i &= \sum_{n=0}^{m}\frac{1-a^{n+1}}{1-a}\\ &= \frac{1}{1-a}\left(m+1-\sum_{n=0}^{m} a^{n+1}\right)\\ &= \frac{1}{1-a}\left(m+1-\sum_{n=1}^{m+1} a^{n}\right)\\ &= \frac{1}{1-a}\left(m+1-\frac{a-a^{m+2}}{1-a}\right)\\ &= \frac{1-2a+m(1-a)+a^{m+2}}{(1-a)^2}\end{aligned}$$

This however is undefined if $a = 1$. A complete answer would compute that case separately:

$$\sum_{n=0}^{m}\sum_{i=0}^{n} a^i = \sum_{n=0}^{m}(n+1) = \sum_{n=1}^{m+1} n = \frac{(m+1)(m+2)}{2}$$

.

2. Let $\Delta, T > 0$. At a 3-way intersection a robot turns left with probability p, and turns right with probability $1-p$, where
$$p = e^{-\Delta/T}.$$
What kind of Δ, T makes the robot strictly prefer a left turn?

The robot strictly prefers a left turn when $p > 0.5$. This occurs when $-(\Delta/T) > \log 0.5$ or more simply $\Delta < T \log 2$.

3. Consider matrix
$$A = \begin{bmatrix}1 & 3\\ 2 & 4\end{bmatrix}.$$
Compute $trace(A^{-1})$ by hand. Matrix operations are important in visualizing big data.

For $A = \begin{pmatrix}1 & 3\\ 2 & 4\end{pmatrix}$, $\det(A) = (4)(1) - (3)(2) = -2$.

$$\begin{aligned}A^{-1} &= \frac{1}{\det(A)}\begin{pmatrix}4 & -3\\ -2 & 1\end{pmatrix}\\ &= \begin{pmatrix}-2 & 1.5\\ 1 & -0.5\end{pmatrix}\end{aligned}$$

So $trace(A^{-1}) = -2.5$.

4. Let
$$f(x) = e^x + e^{-x}.$$
Find the x where the derivative vanishes: $f'(x) = 0$. Derivatives and their high-dimensional cousins the gradients are fundamental in deep learning.

$$f'(x) = e^x - e^{-x}$$

If $f'(x) = 0$ then $e^x = e^{-x}$. Since the exponential function is strictly monotone increasing, this is only true where $x = -x$, and that's only true when $x = 0$.

5. Mary has a book and she won't tell you the number of pages N it has. However, Mary agrees to randomly (i.e., uniformly in $\{1, \dots, N\}$) flip to a page and tell you that page number x. Mary repeats this m times, so you have random page numbers $x_1, \dots, x_m$. Given $x_1, \dots, x_m$, what is your best guess of N in the sense of maximum likelihood? Hint: write down the joint probablity of observing $x_1, \dots, x_m$ as a function of N. This will be the product of probability of observing each x_i, which can be easily expressed due to the uniform distribution. Then find N that maximizes the joint probability. This is the concept behind statistical machine learning.

We will use the notation $X_1, \dots, X_m$ to denote the random variables and $x_1, \dots, x_m$ to denote values they take on. For each $i \in 1, \dots, m$
$$\mathbb{P}(X_i = x_i) = \frac{1}{N}\mathbb{I}[x_i \in \{1, \dots, N\}]$$
Here $\mathbb{I}[P]$ is the indicator function takes the value 1 if the statement P is true, 0 otherwise. To put the above equation in words: given a guess N, any observation $X_i = x_i$ has probability $\frac{1}{N}$ if x_i is a number from 1 to N, and probability 0 otherwise.

The joint probability of m observation can be written as:
$$\mathbb{P}(X_1 = x_1, \dots, X_m = x_m) = \left(\frac{1}{N}\right)^m \mathbb{I}[x_1 \in \{1, \dots, N\} \wedge \dots \wedge x_m \in \{1, \dots, N\}]$$

This function is maximized when N is the smallest value such that
$$\mathbb{I}[x_1 \in \{1, \dots, N\} \wedge \dots \wedge x_m \in \{1, \dots, N\}] = 1$$

And that is achieved by choosing $N = \max(x_1, \dots, x_m)$.

Question 2: Syllabus [21 points total, 3 each]

Read the course syllabus at `http://pages.cs.wisc.edu/~jerryzhu/cs540.html` and answer the following questions:

1. Which groups hold office hours: the instructor, TAs, peer mentors, graders?

Instructor, TAs, and peer mentors hold office hours.

2. Where do we post announcements and clarifications?

The instructors and TAs will post announcements, clarifications, hints, etc. on Piazza.

3. What time of day (hour:minute) are all homeworks due?

Homework is always due the minute before class starts on the due date.

4. Will late homework be accepted?

Late submissions will not be accepted.

5. Tom received the following scores on his 11 homeworks: 0, 59, 92, 93, 94, 95, 96, 97, 98, 99, 100. According to the homework policies, what is Tom's final average homework score?

The lowest two scores are dropped according to the homework policy. The average of the remaining scores is 96.

6. How can you discuss homework questions with fellow students while avoiding the impression of cheating?

All examinations, programming assignments, and written homeworks must be written up individually. For example, code for programming assignments must not be developed in groups, nor should code be shared. Make sure you work through all problems yourself, and that your final write-up is your own.

If peer discussions are too deep for comfort, declare it in the homework solution: "I discussed with X,Y,Z the following specific ideas: A, B, C; therefore our solutions may have similarities on D, E, F..."

7. Imagine one plausible "poison berry" an instructor may plant in an old question+solution to catch cheating, or state "I don't know."

Question 3: Rejoinder [34 points]

Read the Stanford One Hundred Year Study on Artificial Intelligence 2016 Report (linked on course webpage) in its entirety. It will take perhaps a whole morning – plan ahead!

Then, write a short essay to challenge one or more aspects of the Stanford One Hundred Year Study on Artificial Intelligence. Requirements with point distribution:

- (3) Your essay should be between 500 and 600 words.
- (3) The essay should be a single plain text file in English. No pdf, word, rtf, etc., and no attachments of audio, video, images, etc. please. Name your file **rejoinder.txt**
- (3) Do not include your name, your family, friends, or any private and sensitive information in the essay. Your essay will be read by others, including your fellow students.
- (3) Make a clear statement of your challenge.
- (22) General scholarly writing. Other than the above requirements, you have complete freedom in format and content.