**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.

$$\sum_{n=0}^{m} \sum_{i=0}^{n} a^i = \sum_{n=0}^{m} \left(a^0 + a^1 + a^2 + a^3 + ..... + a^n\right)$$

$$= \sum_{n=0}^{m} \left(1 + a + a^2 + a^3 + ..... + a^{n-1} + a^n\right)$$

$$= \sum_{n=0}^{m} \left(\frac{a^{n+1} - 1}{a - 1}\right) \text{ // From Solution 1 solved below}$$

$$= \frac{1}{a-1} \sum_{n=0}^{m} \left(a^{n+1} - 1\right)$$

$$= \frac{1}{a-1}\left[\sum_{n=0}^{m} a^{n+1} - \sum_{n=0}^{m} 1\right]$$

$$= \frac{1}{a-1}\left[\left(a^1 + a^2 + a^3 + \ldots\ldots\ldots\ldots + a^m + a^{m+1}\right) - (m+1)\right]$$

$$= \frac{1}{a-1}\left[a\left(1 + a + a^2 + \ldots\ldots\ldots\ldots + a^{m-1} + a^m\right) - (m+1)\right]$$

$$= \frac{1}{a-1}\left[a\left(\frac{a^{m+1}-1}{a-1}\right) - (m+1)\right] \text{ // From Solution 1 solved below}$$

$$= \frac{1}{(a-1)^2}\left[a\left(a^{m+1}-1\right) - (m+1)(a-1)\right]$$

$$= \frac{1}{(a-1)^2}\left[a^{m+2} - a - ma + m - a + 1\right]$$

$$= \frac{1}{(a-1)^2}\left[a^{m+2} - m(a-1) - 2a + 1\right]$$

Solution 1:

$$\text{Consider } X = \left(1 + a + a^2 + a^3 + \ldots\ldots\ldots\ldots + a^{n-1} + a^n\right)$$

Multiplying 'a' on both the side

$$aX = \left(a + a^2 + a^3 + \ldots\ldots\ldots\ldots + a^{n-1} + a^n + a^{n+1}\right)$$

Subtracting X from both the sides

$$aX - X = \left(a + a^2 + a^3 + \ldots\ldots\ldots\ldots + a^{n-1} + a^n + a^{n+1}\right) - X$$

$$(a-1)X = \left(a + a^2 + a^3 + \ldots\ldots\ldots\ldots + a^{n-1} + a^n + a^{n+1}\right) - \left(1 + a + a^2 + a^3 + \ldots\ldots\ldots\ldots + a^{n-1} + a^n\right)$$

$$X = \frac{a^{n+1}-1}{a-1}$$

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 $\Delta, T$ makes the robot strictly prefer a left turn?

$$=> p = e^{-\frac{\Delta}{T}}$$

Taking log to the base e on both the sides,

$$\ln p = \ln e^{-\frac{\Delta}{T}}$$

$$\ln p = -\frac{\Delta}{T} \ln e$$

$$\ln p = -\frac{\Delta}{T} \text{ // Exponential Property } \ln e = 1$$

For the robot to strictly prefer left turn, the probability value would be in the range

$$0.5 < p \leqslant 1$$

Applying log to the base e

$$\ln 0.5 < \ln p \leqslant ln1$$

$$-0.6931 < -\frac{\Delta}{T} \leqslant 0$$

Multiplying by -1 (Comparison operator changes accordingly)

$$0.6931 > \frac{\Delta}{T} \geqslant 0$$

This shows,

$$\frac{\Delta}{T} \geqslant 0 \; and \; \frac{\Delta}{T} < 0.6931$$

$$\text{Given } \Delta,\ T\ > 0$$

We have our solution:

$$\Delta\ >\ 0.6931\ T$$

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.

$$Trace\left(A^{-1}\right), given A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}.$$

Mathematically Inverse of any matrix, say A is defined as below

$$A^{-1} = \frac{1}{|A|} Adj\,(A)$$

Computing determinant of Matrix A

$$|A| = \begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} = (1 * 4 - 3 * 2) = 4 - 6 = -2$$

Computing Adjugate of Matrix A

$$Adj(A) = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}$$

Computing Inverse of A

$$A^{-1} = \frac{1}{-2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 3/2 \\ 1 & -1/2 \end{bmatrix}$$

Now determining Trace of Inverse of A

$$Trace(A^{-1}) = -2 - \frac{1}{2} = -\frac{5}{2}$$

$$Thus, Trace(A^{-1}) = -\frac{5}{2}$$

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.

$$f\,(x) = e^x + e^{-x}, f^{'}\,(x) = 0$$

Differentiating f(x),

$$f^{'}\,(x) = 0 => \frac{de^x}{dx} + \frac{de^{-x}}{dx} = 0$$

$$e^x + e^{-x}\,(-1) = 0$$

$$e^x - e^{-x} = 0$$

$$e^x - \frac{1}{e^x} = 0$$

$$e^{2x} - 1 = 0$$

$$e^{2x} = 1$$

$$x = 0$$

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 $\{1, \ldots, N\}$) flip to a page and tell you that page number $x$. Mary repeats this $m$ times, so you have random page numbers $x_1, \ldots, x_m$. Given $x_1, \ldots, x_m$, what is your best guess of $N$ in the sense of maximum likelihood? Hint: write down the joint probability of observing $x_1, \ldots, x_m$ 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.

$$\text{Suppose} x_1, x_2, x_3, x_4 \ldots\ldots\ldots x_m$$

are random samples of a uniform distribution function with an unknown maximum value N, then each random sample/variable has a density (probability)

$$f(\frac{x}{N}) = \begin{cases} \frac{1}{N} & 0 < x \leqslant N \\ 0 & otherwise \end{cases}$$

The joint probability or the likelihood is given by the product of density (probability) of each random variable

$$f(\frac{x}{N}) = L(\frac{N}{x}) \begin{cases} \frac{1}{N^m} & 0 < x_i \leqslant N, \ \forall\ 0 < i < m \\ 0 & otherwise \end{cases}$$

$$L(\frac{N}{x}) = N^{-m}$$

$$L\left(\frac{N}{x}\right) = N^{-m} \ //\text{Likelihood function}$$

Likelihood function now has to be maximized which is represented by,

$$\widehat{N(x)} = N^{-m}$$

This is a reducing function, By that means with more instances of m random samples the function reduces and the maximum value of unknown parameter N for all m ranging between 0 and N would be given by the highest among the random samples provided

$$\text{i.e } \widehat{N(x)} = Max(x_1, x_2, x_3, .....x_m)$$

# 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?

   Office hours are held by the Instructor, TAs and Peer Mentors.

2. Where do we post announcements and clarifications?

   Announcements and clarifications are posted in Piazza.

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

   It's a minute before the class starts on the due date. (10.59 AM If all assignments due is every Thrusday)

4. Will late homework be accepted?

   Last submissions are not acceptable.

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?

   Best 9 out of 11 scores are considered for final average homework score (0 and 59 are omitted)

   The average of remaining 9 scores is given by:

$$\frac{92 + 93 + 94 + 95 + 96 + 97 + 98 + 99 + 100}{9} = \frac{864}{9} = 96$$

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

   I will make sure that I work through all problems individually, and that my final write-up is my own. If at the discussion were too deep for comfort, I 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."

   I will not be tempted to use such old questions+solutions. If I were caught copying such solutions, then I would be automatically failed.

## 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.