CS 601.471/671 NLP: Self-supervised Models

Homework 1: Background Review + Word Representations

For homework deadline, check the calendar on the course website*

Name: _____

Collaborators: _____

Sources used for your homework: _____

This assignment it combines knowledge and skills across several disciplines. The purpose of this assignment is to make sure you are prepared for this course. We anticipate that each of you will have different strengths and weaknesses, so don't be worried if you struggle with *some* aspects of the assignment. But if you find this assignment to be very difficult overall, that is an early warning sign that you may not be prepared to take this course at this time.

To succeed in the course, you will need to know or very quickly get up to speed on:

- Math to the level of the course prerequisites: linear algebra, multivariable calculus, some probability.
- Statistics, algorithms, and data structures to the level of the course prerequisites.
- Python programming, and the ability to translate from math or algorithms to programming and back.
- Some basic LaTeX skills so that you can typeset equations and submit your assignments.

How to hand in your written work: via Gradescope.

Collaboration: Make certain that you understand the course collaboration policy, described on the course website. You may discuss the homework to understand the problems and the mathematics behind the various learning algorithms, but you are **not allowed to share problem solutions with any other students**. You must write the solutions **individually**.

Typesetting: We strongly recommend typesetting your homework, especially if you have sloppy handwriting. We will provide a LaTeX template for homework solutions.

1 Linear Algebra Review

For these questions, you may find it helpful to review these notes on linear algebra:
http://www.cs.ubc.ca/~schmidtm/Documents/2009_Notes_LinearAlgebra.pdf

1.1 Basic Operations

Use the definitions below,

$$\alpha = 2, \quad \mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad A = \begin{bmatrix} 3 & 2 & 2 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix},$$

and use x_i to denote element i of vector $\mathbf{x}$. Evaluate the following expressions:

1. $\sum_{i=1}^{n} x_i y_i$ (inner product).

   Answer: *TBD*

*https://self-supervised.cs.jhu.edu/sp2023/

2. $\sum_{i=1}^{n} x_i z_i$ (inner product between orthogonal vectors).

   Answer: *TBD*

3. $\alpha(\mathbf{x} + \mathbf{y})$ (vector addition and scalar multiplication).

   Answer: *TBD*

4. $\|\mathbf{x}\|$ (Euclidean norm of x).

   Answer: *TBD*

5. $\mathbf{x}^\top$ (vector transpose).

   Answer: *TBD*

6. $A\mathbf{x}$ (matrix-vector multiplication).

   Answer: *TBD*

7. $\mathbf{x}^\top A\mathbf{x}$ (quadratic form).

   Answer: *TBD*

Note, you do not need to show your work.

1.2 Matrix Algebra Rules

Assume that $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ are $n \times 1$ column vectors and $\{A, B, C\}$ are $n \times n$ real-valued matrices, and $\mathbf{I}$ is the identity matrix of appropriate size. State whether each of the below is true in general (you do not need to show your work).

1. $\mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{n} x_i y_i$.

   Answer: *TBD*

2. $\mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|^2$.

   Answer: *TBD*

3. $\mathbf{x}^\top \mathbf{x} = \mathbf{x}\mathbf{x}^\top$.

   Answer: *TBD*

4. $(\mathbf{x} - \mathbf{y})^\top (\mathbf{y} - \mathbf{x}) = \|\mathbf{x}\|^2 - 2\mathbf{x}^\top \mathbf{y} + \|\mathbf{y}\|^2$.

   Answer: *TBD*

5. $AB = BA$.

   Answer: *TBD*

6. $A(B + C) = AB + AC$.

   Answer: *TBD*

7. $(AB)^\top = A^\top B^\top$.

   Answer: *TBD*

8. $\mathbf{x}^\top A\mathbf{y} = \mathbf{y}^\top A^\top \mathbf{x}$.

   Answer: *TBD*

9. $A^\top A = \mathbf{I}$ is the columns of A are orthonormal.

   Answer: *TBD*

2 Probability Review

For these questions you may find it helpful to review these notes on probability:
http://www.cs.ubc.ca/~schmidtm/Courses/Notes/probability.pdf
http://www.cs.ubc.ca/~schmidtm/Courses/Notes/probabilitySlides.pdf

Answer the following questions. You do not need to show your work.

1. You are offered the opportunity to play the following game: your opponent rolls 2 regular 6-sided dice. If the difference between the two rolls is at least 3, you win $15. Otherwise, you get nothing. What is a fair price for a ticket to play this game once? In other words, what is the expected value of playing the game?

   Answer: *TBD*

2. Consider two events A and B such that $\mathbf{P}(A, B) = 0$ (they are mutually exclusive). If $\mathbf{P}(A) = 0.4$ and $\mathbf{P}(A \cup B) = 0.95$, what is $\mathbf{P}(B)$? Note: $p(A, B)$ means "probability of A and B" while $p(A \cup B)$ means "probability of A or B". It may be helpful to draw a Venn diagram.

   Answer: *TBD*

3. Instead of assuming that A and B are mutually exclusive ($\mathbf{P}(A, B) = 0$), what is the answer to the previous question if we assume that A and B are independent?

   Answer: *TBD*

3 Calculus Review

For these questions you may find it helpful to review these notes on calculus:
http://www.cs.ubc.ca/~schmidtm/Courses/Notes/calculus.pdf

3.1 One-variable derivatives

Answer the following questions. You do not need to show your work.

1. Find the derivative of the function $f(x) = 3x^2 - 2x + 5$.

   Answer: *TBD*

2. Find the derivative of the function $f(x) = x(1 - x)$.

   Answer: *TBD*

3. Let $p(x) = \frac{1}{1+\exp(-x)}$ for $x \in \mathbb{R}$. Compute the derivative of the function $f(x) = x - \log(p(x))$ and simplify it by using the function $p(x)$.

   Answer: *TBD*

Note that in this course we will use $\log(x)$ to mean the "natural" logarithm of x, so that $\log(\exp(1)) = 1$. Also, observe that $p(x) = 1 - p(-x)$ for the final part.

3.2 Multi-variable derivative

Compute the gradient $\nabla f(\mathbf{x})$ of each of the following functions. You do not need to show your work.

1. $f(\mathbf{x}) = x_1^2 + \exp(x_2)$ where $\mathbf{x} = [x_1, x_2] \in \mathbb{R}^2$.

   Answer: *TBD*

2. $f(\mathbf{x}) = \exp(x_1 + x_2 x_3)$ where $\mathbf{x} = [x_1, x_2, x_3] \in \mathbb{R}^3$.

   Answer: *TBD*

3. $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$ where $\mathbf{x} \in \mathbb{R}^2$ and $\mathbf{a} \in \mathbb{R}^2$.

   Answer: *TBD*

4. $f(\mathbf{x}) = \mathbf{x}^\top Ax$ where $A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ and $\mathbf{x} \in \mathbb{R}^2$.

   Answer: *TBD*

5. $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|^2$ where $\mathbf{x} \in \mathbb{R}^d$.

   Answer: *TBD*

Hint: it is helpful to write out the linear algebra expressions in terms of summations.

4 Algorithms and Data Structures Review

For these questions you may find it helpful to review these notes or this Wiki page on big-O notation. Now, answer the following questions using big-O notation You do not need to show your work.

1. What is the cost of running the merge-sort algorithm to sort a list of n numbers?

   Answer: *TBD*

2. What is the cost of finding the third-largest element of an unsorted list of n numbers?

   Answer: *TBD*

3. What is the cost of finding the smallest element greater than o in a *sorted* list with n numbers?

   Answer: *TBD*

4. What is the cost of finding the value associated with a key in a hash table with n numbers? (Assume the values and keys are both scalars.)

   Answer: *TBD*

5. What is the cost of computing the matrix-vector product $A\mathbf{x}$ when A is $n \times d$ and x is $d \times 1$?

   Answer: *TBD*

6. What is the cost of computing the quadratic form $\mathbf{x}^\top A\mathbf{x}$ when A is $d \times d$ and $\mathbf{x}$ is $d \times 1$?

   Answer: *TBD*

7. How does the answer to the previous question change if A has only z non-zeroes elements? (You can assume $z \geq d$)

   Answer: *TBD*

5 Programming

See the course website for the link to Google Colab.