COS 484: Natural Language Processing **(Due: 02/06/23)**

Assignment #0

Instructor: Danqi Chen

Course Policy: Read all the instructions below carefully before you start working on the assignment and before you make a submission. The course assignment policy is available at http://nlp.cs.princeton.edu/cos484.

- This assignment contains 2 parts, a theoretical and a programming part. The former consists of 4 problems, and the latter has 2, for a total of **6 problems**.
- We *highly* recommended that you typeset your submissions in LaTeX. Use the template provided on the website for your answers. If you have never used LaTeX, you can refer to the short guide here: http://bit.ly/WorkingWithLaTeX. Include your name and NetIDs with your submission. If you wish to submit hand-written answers, you can scan and upload the pdf.
- Assignments must be uploaded to Gradescope **before class (9:30am Eastern)** on the due date mentioned above.
- As per the late-day policy outlined on the course website, you have 96 allowed late hours (about 4 days) to use at your discretion throughout the semester. Once you run out of late hours, late submissions will incur a penalty of 10% for each day, up to a maximum of 3 days beyond which submissions will not be accepted.
- One of the goals of this assignment is to assess whether you have adequate preparation for the course. It's fine to not be familiar with *every* concept here. However, if you find yourself struggling with much of this assignment (especially Problem 1: Math Review), you should ask the course staff whether this course is appropriate for you given your background.

Part 1: Math Review

(10 + 7 + 3 + 5 = 25 points)

Throughout this course, we'll be constantly referencing concepts from linear algebra (e.g. vectors and matrices), probability (e.g. distributions and conditional probability), and calculus (e.g. partial derivatives). The following questions cover some of the core fundamentals that we expect you to be familiar with from the start.

1.1 Linear Algebra

(a) (5 points) Provide answers to the following operations or write "invalid" if not possible:

(i) $\begin{bmatrix} -4 & 2 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} -4 \\ 4 \end{bmatrix}$

(ii) $\begin{bmatrix} -2 & 4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 \end{bmatrix}$

(iii) $\begin{bmatrix} 3 & 4 & -3 \\ 5 & 2 & 1 \end{bmatrix} \begin{bmatrix} -4 & 3 & 5 \\ 2 & -1 & -1 \end{bmatrix}^\intercal$

(iv) $\begin{bmatrix} 4 & 5 & 6 \end{bmatrix}^\intercal \begin{bmatrix} 2 & 1 \end{bmatrix}^\intercal$

(v) $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^\intercal \begin{bmatrix} 5 & 4 \end{bmatrix}$

(b) (5 points) Suppose that we're given two matrices: $\mathbf{w} \in \mathbb{R}^{D \times H}$ and $\mathbf{x} \in \mathbb{R}^{D \times N}$, where $H \neq D \neq N$. What are the output dimension of the following expressions? Write "Invalid" if they cannot be computed.

(i) $\mathbf{x}^\intercal \mathbf{x}$

(ii) $\mathbf{x}\mathbf{x}^\intercal$

(iii) $\mathbf{w}^\intercal \mathbf{x}$

(iv) $\mathbf{x}\mathbf{w}^\intercal$

(v) $\mathbf{w}^\intercal \mathbf{x}\mathbf{x}^\intercal \mathbf{w}$

1.2 Expectation and entropy

(a) (1 point) Given a categorical distribution $P(X)$ as follows, compute the value of $\mathbb{E}[X]$.

x	1	2	3
$P(X = x)$	0.75	0.1	0.15

(b) (1 point) Given a uniform distribution $P(X = x) = \frac{1}{m}, \forall x \in \{1, 2, \ldots, m\}$, compute the value of $\mathbb{E}[X]$.

(c) (1 point) The entropy of a discrete random variable X is defined as (use base e for all log operations unless otherwise specified):

$$H(X) = -\sum_{x \in X} P(x) \log P(x).$$

For the categorical distribution $P(X)$ in problem (a) above, compute the entropy of the distribution.

(d) (1 point) Compute the entropy of the uniform distribution $P(X = x) = \frac{1}{m}, \forall x \in \{1, 2, \ldots, m\}$.

(e) (3 points) Consider the entropy of the joint distribution $P(X, Y)$:

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} P(x, y) \log P(x, y)$$

How does this entropy relate to H(X) and H(Y), (i.e. the entropies of the marginal distributions) when X and Y are independent?

1.3 Bayes' Rule

Suppose that you have two coins in your pocket, one fair coin: $P(H) = P(T) = 0.5$, and one unfair coin: $P(H) = 0.8$ and $P(T) = 0.2$. Suppose that you take one coin from your pocket (each coin equally likely) and flip it.

(a) (1 point) What's the probability that it turns up heads (H)?

(b) (2 points) Suppose that it comes up heads. Given this fact, what's the probability that we picked the unfair coin?

1.4 Calculus

Suppose that we define 4 functions:

$$f(x) = 5x - 1$$

$$h(x) = \frac{f(x)}{2} + 2$$

$$\mathcal{L}(x) = -\log(h(x)^2)$$

$$S(x) = \frac{1}{1 + e^{-x}}$$

(a) (1 point) Compute $\frac{df}{dx}$.
(b) (1 point) Compute $\frac{dh}{dx}$.
(c) (1 point) Compute $\frac{d\mathcal{L}}{dx}$.
(d) (2 points) Show that $\frac{dS}{dx} = S(x)(1 - S(x))$.

Part 2: Programming with Python (10 + 10 = 20 points)

This second problem will be done in Google Colab. It consists of two questions, one to practice using NumPy and another to practice using Matplotlib. Navigate to the Colab document and follow the directions there.

1. This problem will be completed in this Colab notebook.
2. If you've never worked with Google Colab before, take a look through this introduction guide: `http://bit.ly/WorkingWithColab`
3. When you're ready to submit your assignment, you can follow the submission instructions found here: `http://bit.ly/COS_NLP_Submission`