CS 7650: Natural Language Processing
Fall 2022
Problem Set 0

Instruction

1. Problem Set 0 (together with the Programming Project 0) is meant to serve as a background preparation test, and to help students decide whether they have enough math/programming skills to succeed in this class. Collaboration and discussion are all **NOT** allowed. All of the questions represent materials that students are expected to be familiar with before they take this class.

2. As one of the most advanced AI classes, CS 7650 covers deep learning and other machine learning models in the context of processing text data. The lectures will contain a lot of math (i.e., linear algebra, probability, and multivariate calculus) and the actual programming assignments will be of larger scale and much more challenging to debug than Project 0. You are strongly recommended to take CS 4641 and (Math 2550 or Math 2551 or Math 2561 or Math 2401 or Math 24X1 or 2X51) before taking this class.

3. Submit your answers as a pdf file on Gradescope. When submitting to Gradescope, make sure to mark page(s) correspond to each problem or subproblem. We recommend students type answers with LaTeX or word processors. A scanned handwritten copy would also be acceptable (but hard copies will not be accepted). If writing by hand, write as clearly as possible. No credit may be given to unreadable handwriting. It may take about a day to be added to Gradescope after you enrolled into the class. If you still cannot access Gradescope the day when Problem Set 0 (PS0) is due, please email your submission to the instructor with "CS 7650 - PS0" in title.

4. Write out all steps required to find the solutions so that partial credit may be awarded.

5. You may find the lecture note on matrix calculus by Randal J. Barnes helpful: https://atmos.washington.edu/~dennis/MatrixCalculus.pdf.

6. **For students on the wait list**: we don't have any additional information on whether you will be able to enroll in the course, but if you plan to take the class, please complete and submit Problem Set 0 by the due time (late penalty will apply if submitted later). If you get off the wait list, you will be automatically added to Gradescope after about a day. You can also post a message on Piazza to get the access code to Gradescope. If you cannot access Gradescope by the due date, please email your submission to the instructor as mentioned above in (3).

1 Linear Algebra

(a) (**3 pts**) Compute the l_1 norm, l_2 norm, and l_∞ norm of the vector $\mathbf{x} = \begin{bmatrix} -5 \\ -1 \\ 6 \end{bmatrix}$.

(b) (**1 pts**) Compute vector $\mathbf{x}$ as the solution to the following linear equation: $\begin{bmatrix} 1 & 5 \\ -2 & 3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 7 \\ -1 \end{bmatrix}$

(c) (**3 pts**) Provide answers to the following operations ($^\top$ transposes a vector or matrix):

(i) Dot product: $\mathbf{w} = \begin{bmatrix} 2 \\ 0 \\ 6 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} -5 \\ -1 \\ 3 \end{bmatrix}$, then $\mathbf{w}^\top \mathbf{x} =$?

(ii) Matrix product: $\mathbf{A} = \begin{bmatrix} 1 & -1 & -4 \\ 2 & 6 & 0 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} -3 & 7 \\ 5 & 2 \\ 3 & -4 \end{bmatrix}$, then $\mathbf{AB} = $?

(iii) Elementwise product: same $\mathbf{A}$ and $\mathbf{B}$ as above, what is $\mathbf{A} \odot \mathbf{B}^\top = $?

2 Geometry

(a) (**2 pts**) True or False (if false, explain why)? $\|\alpha \mathbf{u} + \mathbf{v}\|^2 = \alpha^2 \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$, where $\|\cdot\|$ denotes Euclidean norm, α is a scalar, $\mathbf{u}$ and $\mathbf{v}$ are vectors.

(b) (**2 pts**) Show that the vector $\mathbf{w}$ is orthogonal to the line $\mathbf{w}^\top \mathbf{x} + b = 0$. (*Hint*: consider two points $\mathbf{x}_1$ and $\mathbf{x}_2$ that lie on the line.)

3 Multivariate Calculus

(a) (**2 pts**) The number of members of a gym in Midtown Atlanta grows approximately as a function of the number of weeks, t, in the first year it is opened: $f(t) = 100(60 + 5t)^{2/3}$. How fast was the membership increasing initially (i.e., what is the gradient of $f(t)$ when $t = 0$)?

(b) (**2 pts**) Consider the equations $L = (1 - z)^2$, $z = w_2 y + b$, and $y = w_1 x$.
Compute the gradients $\frac{\partial L}{\partial w_1}$, $\frac{\partial L}{\partial w_2}$, and $\frac{\partial L}{\partial b}$.

(c) Let $\mathbf{c}$ be a column vector. Let $\mathbf{x}$ be another column vector of the same dimension.

   (i) (**1 pts**) Consider a linear function $f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x}$. Compute the gradient $\frac{\partial}{\partial \mathbf{x}} f(\mathbf{x})$.

   (ii) (**1 pts**) Consider a quadratic function $g(\mathbf{x}) = \frac{1}{2} \mathbf{x}^\top \mathbf{H} \mathbf{x}$ where $\mathbf{H}$ is a square matrix of dimensions compatible with $\mathbf{x}$. Compute the gradient $\frac{\partial}{\partial \mathbf{x}} g(\mathbf{x})$.

   (iii) (**2 pts**) Let $h(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{H} \mathbf{x} + \mathbf{c}^T \mathbf{x}$, where $\mathbf{H} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$ and $\mathbf{c} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$. When the gradient $\frac{\partial}{\partial \mathbf{x}} h(\mathbf{x}) = 0$, what is $\mathbf{x} = ?$ Is it a local minimum, maximum or saddle point?

4 Probability

(a) (**2 pts**) Georgia Tech's Robotics Lab has designed a robot that either takes one step forward or backward. The probability that it takes a forward step is 0.3. Find the probability that at end of 8 steps it is 2 steps away from the starting point?

(b) (**2 pts**) Let A be the event that a patient has contracted the flu, and let B be the event that a patient has a fever. Assume $P(A) = 0.3$ and $P(B) = 0.1$. The probability that a patient has the flu, given that they also have a fever is $P(A \mid B) = 0.9$. A new patient has arrived in the hospital, and they have a fever. What is the probability of them having the flu?

(c) A probability density function is defined by

$$f(x) = \begin{cases} Ce^{-x} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

    (i) (**1 pts**) Find the value of C that makes $f(x)$ a valid probability density function.

    (ii) (**1 pts**) Compute the expected value of x, i.e., $E(x)$.