CS 7650: Natural Language Processing (Fall 2025)
Problem Set 0

Instructor: Wei Xu
TAs: Duong Minh Le, Jerry Zheng, Joseph Thomas, Rohan Phadnis
Course website: https://cocoxu.github.io/CS7650_fall2025/
Gradescope: https://www.gradescope.com/courses/1086056

Due: Thursday, August 21, 11:59 PM ET

Instruction

1. This Problem Set 0 (together with 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. As PS0 is a test, collaboration and discussion are **NOT** allowed. All of the questions represent materials that students are expected to be familiar with before they take this class. PS0 is part of the coursework and will be counted towards 2% of the final grade.

2. CS 4650 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 a larger scale and much more challenging to debug than Project 0. You are strongly required/recommended to have taken the CS 4641/7641 (or equivalent) Machine Learning class **before** taking this class (not in the same semester while taking this class). Taking CS 4644/7643 Deep Learning class first will also help you understand the class material better, and complete and debug the programming homework in this course with more ease.

3. You may find the lecture notes on matrix calculus by Randal J. Barnes and Zico Kolter helpful: (1) https://atmos.washington.edu/~dennis/MatrixCalculus.pdf; (2) http://cs229.stanford.edu/section/cs229-linalg.pdf

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

5. 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. If you cannot access Gradescope when homework is due, please email your submission to the instructor with "CS 7650 - PS0" in the title.

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 (largely depend on when/whether students currently enrolled will drop the class before registration closes). If you are on the waitlist and plan to take the class, please make sure to complete and submit Problem Set 0 by the due time. If you get off the waitlist, you will be automatically added to Gradescope after about a day. If you still cannot access Gradescope when homework is due, please email your submission to the instructor (see 5. above).

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 $\frac{\partial}{\partial \mathbf{x}} h(\mathbf{x}) = \mathbf{0}$, 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 a fever, and let B be the event that a patient has contracted the flu. Assume $P(A) = 0.3$ and $P(B) = 0.1$. The probability that a patient has a fever, given that they have the flu 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., $E(x)$.