CS 1678/2078 Homework Prerequisite Review

Abstract

This assignment is to serve as a review of some of the relevant background required to do well in this course. It is also an introduction to typesetting math equations in LaTeX, and to practice the relevant the relevant concepts we will use in the course. I recommend first working out the solutions below by hand then typing them up. To submit this assignment, upload a `.pdf` to Gradescope containing your responses to the questions below. You are required to use LaTeX for your write up. We suggest using overleaf.com since they do not require installing anything on your computer and provide free accounts. When submitting your answers, use the template LaTeX code provided and put your answers in relevant blocks. When uploading your solutions to gradescope make sure to tag each question with the correct page in the pdf.

To submit the assignment's coding portion, upload a zip folder to gradescope containing the files used for the project.

1 Math Review (25 points)

Write the answers to the questions below showing your work in LaTeX. You do not have to show every single step, but the reason for each jump should be obvious. When applicable, you may reuse the answers in previous questions to answer the question.

1. (3 points) Consider the function $f\colon \mathbb{R}\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, that takes as input three real numbers x, y, z and returns a scalar $f(x,y,z) \doteq 3x^2+\sin(y)z$. What are partial derivatives for each x, y, z?

   Answer:

$$\begin{aligned}\frac{\partial f(x,y,z)}{\partial x} &= \frac{\partial}{\partial x}\left(3x^2+\sin(y)z\right)\\ &= \frac{\partial 3x^2}{\partial x} + \underbrace{\frac{\partial \sin(y)z}{\partial x}}_{=0}\\ &= 6x\end{aligned}$$

$$\begin{aligned}\frac{\partial f(x,y,z)}{\partial y} &= \frac{\partial}{\partial y}\left(3x^2+\sin(y)z\right)\\ &= \frac{\partial 3x^2}{\partial y} + \frac{\partial \sin(y)z}{\partial y}\\ &= z\cos(y)\end{aligned}$$

$$\begin{aligned}\frac{\partial f(x,y,z)}{\partial z} &= \frac{\partial}{\partial z}\left(3x^2+\sin(y)z\right)\\ &= \frac{\partial 3x^2}{\partial z} + \frac{\partial \sin(y)z}{\partial z}\\ &= \sin(y)\end{aligned}$$

2. (1 points) What is the gradient of f? You may reuse the results of the above equations.

   Answer:

$$\nabla f(x,y,z) = \begin{bmatrix}\frac{\partial f(x,y,z)}{\partial x}\\ \frac{\partial f(x,y,z)}{\partial y}\\ \frac{\partial f(x,y,z)}{\partial z}\end{bmatrix} = \begin{bmatrix}6x\\ z\cos(y)\\ \sin(y)\end{bmatrix}$$

3. (1 point) We could redefine the function above to take as input a vector of length three instead of x, y, z, i.e., $f : \mathbb{R}^3 \to \mathbb{R}$ where $f(x) \doteq 3x_1^2 + \sin(x_2)x_3$, with $x \in \mathbb{R}^3$. What is the gradient of this new f?

   Answer:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \frac{\partial f(x)}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 6x \\ x_3 \cos(x_2) \\ \sin(x_2) \end{bmatrix}$$

4. (8 points) Write the derivatives for the following vector norms

$$\begin{aligned} \|x\|_2^2 &\doteq \sum_{i=1}^{n} x_i^2 \\ \|x\|_2 &\doteq \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\|x\|_2^2} \\ \|x\|_1 &\doteq \sum_{i=1}^{n} |x_i|, \text{ and} \\ \|x\|_\infty &\doteq \max_i |x_i|. \end{aligned}$$

   HINT: break the problem into partial derivatives for each element x_k, e.g.,

$$\frac{\partial \|x\|_2^2}{\partial x} = \begin{bmatrix} \frac{\partial \|x\|_2^2}{\partial x_1} & \frac{\partial \|x\|_2^2}{\partial x_2} & \dots & \frac{\partial \|x\|_2^2}{\partial x_n} \end{bmatrix}^\top .$$

   You can reuse part of a previous answer in the derivation for the next answer.

   (a) (2 points) $\frac{\partial \|x\|_2^2}{\partial x}$
   Answer:

$$\begin{aligned} \frac{\partial \|x\|_2^2}{\partial x} &= \frac{\partial (x_1^2 + x_2^2 + \ldots + x_n^2)}{\partial x} \\ &= \begin{bmatrix} \frac{\partial (x_1^2 + x_2^2 + \ldots + x_n^2)}{\partial x_1} \\ \frac{\partial (x_1^2 + x_2^2 + \ldots + x_n^2)}{\partial x_2} \\ \vdots \\ \frac{\partial (x_1^2 + x_2^2 + \ldots + x_n^2)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial x_1^2}{\partial x_1} \\ \frac{\partial x_2^2}{\partial x_2} \\ \vdots \\ \frac{\partial x_n^2}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} \\ &= 2x \end{aligned}$$

   (b) (2 points) $\frac{\partial \|x\|_2}{\partial x}$

Answer:

$$
\begin{aligned}
\text{Let } y &= \|x\|_2^2 \\
\frac{\partial \|x\|_2}{\partial x} &= \frac{\partial \sqrt{y}}{\partial x} \\
&= \frac{\partial \sqrt{y}}{\partial y}\frac{\partial y}{\partial x} \\
&= \frac{\partial y^{\frac{1}{2}}}{\partial y}\frac{\partial y}{\partial x} \\
&= \frac{1}{2}y^{-\frac{1}{2}}\frac{\partial \|x\|_2^2}{\partial x} \\
&= \frac{1}{2}\frac{1}{\sqrt{y}}\frac{\partial \|x\|_2^2}{\partial x} \\
&= \frac{1}{2}\frac{1}{\sqrt{y}}2x \\
&= \frac{x}{\|x\|_2}
\end{aligned}
$$

(c) (2 points) $\frac{\partial \|x\|_1}{\partial x}$
Answer:

$$
\begin{aligned}
\frac{\partial \|x\|_1}{\partial x} &= \begin{bmatrix} \frac{\partial}{\partial x_1}\sum_{i=1}^n |x_i| \\ \frac{\partial}{\partial x_2}\sum_{i=1}^n |x_i| \\ \vdots \\ \frac{\partial}{\partial x_n}\sum_{i=1}^n |x_i| \end{bmatrix} = \begin{bmatrix} \frac{\partial |x_1|}{\partial x_1} + \frac{\partial |x_2|}{\partial x_1} + \cdots + \frac{\partial |x_n|}{\partial x_1} \\ \frac{\partial |x_1|}{\partial x_2} + \frac{\partial |x_2|}{\partial x_2} + \cdots + \frac{\partial |x_n|}{\partial x_n} \\ \vdots \\ \frac{\partial |x_1|}{\partial x_n} + \frac{\partial |x_2|}{\partial x_n} + \cdots + \frac{\partial |x_n|}{\partial x_n} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial |x_1|}{\partial x_1} \\ \frac{\partial |x_2|}{\partial x_2} \\ \vdots \\ \frac{\partial |x_n|}{\partial x_n} \end{bmatrix} \\
&= \begin{bmatrix} \operatorname{sgn}(x_1) \\ \operatorname{sgn}(x_2) \\ \vdots \\ \operatorname{sgn}(x_n) \end{bmatrix}
\end{aligned}
$$

(d) (2 points) $\frac{\partial \|x\|_\infty}{\partial x}$
Answer: Let $x^* = \max_i |x_i|$.

$$
\begin{aligned}
\frac{\partial \|x\|_\infty}{\partial x} &= \frac{\partial \max(|x_1|, |x_2|, ..., |x_n|)}{\partial x} \\
&= \begin{bmatrix} \operatorname{sgn}(x_1)\mathbf{1}_{|x_1|=x^*} \\ \operatorname{sgn}(x_2)\mathbf{1}_{|x_2|=x^*} \\ \vdots \\ \operatorname{sgn}(x_n)\mathbf{1}_{|x_n|=x^*} \end{bmatrix}
\end{aligned}
$$

5. (3 points) Consider the function $f(x) = e^{-\frac{1}{2}\|x\|_2^2}$, whose output decays quickly towards zero as the input vector $x \in \mathbb{R}^n$ gets further away from the origin. What is the derivative of $f(x)$ with respect to x? Hint: use the chain rule.

Answer:

$$
\begin{aligned}
\frac{\partial f(x)}{\partial x} &= \frac{\partial}{\partial x} e^{-\frac{1}{2}\|x\|_2^2} \\
&= e^{-\frac{1}{2}\|x\|_2^2} \frac{\partial - \frac{1}{2}\|x\|_2^2}{\partial x} \\
&= -\frac{1}{2} e^{-\frac{1}{2}\|x\|_2^2} \frac{\partial \|x\|_2^2}{\partial x} \\
&= -e^{-\frac{1}{2}\|x\|_2^2} x
\end{aligned}
$$

6. (2 points) Consider the function $f\colon \mathbb{R}^{2\times 3} \times \mathbb{R}^3 \to \mathbb{R}^2$ with the output $f(A, x) = Ax$, where $A \in \mathbb{R}^{2\times 3}$ and $x \in \mathbb{R}^3$. Give an expression for $f(x)$ using the elements of x and A, e.g., x_1,x_2, $A_{1,2}$, $A_{2,1}$. Note that the output of f is a vector of length two. Let $f(x, A)_1$ and $f(x, A)_2$ be the two outputs of $f(A, x)$.

   Answer:

$$
\begin{aligned}
f(A, x) &= \begin{bmatrix} f(A, x)_1 \\ f(A, x)_2 \end{bmatrix} \\
&= \begin{bmatrix} A_{1,1}x_1 + A_{1,2}x_2 + A_{1,3}x_3 \\ A_{2,1}x_1 + A_{2,2}x_2 + A_{2,3}x_3 \end{bmatrix}
\end{aligned}
$$

7. (2 points) What are the partial derivatives of each output of f with respect to x?

   Answer:

$$
\begin{aligned}
\frac{\partial f(A, x)_1}{\partial x} &= \begin{bmatrix} \frac{\partial f(A,x)_1}{\partial x_1} \\ \frac{\partial f(A,x)_1}{\partial x_2} \\ \frac{\partial f(A,x)_1}{\partial x_3} \end{bmatrix} \\
&= \begin{bmatrix} A_{1,1} \\ A_{1,2} \\ A_{1,3} \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f(A, x)_2}{\partial x} &= \begin{bmatrix} \frac{\partial f(A,x)_2}{\partial x_1} \\ \frac{\partial f(A,x)_2}{\partial x_2} \\ \frac{\partial f(A,x)_2}{\partial x_3} \end{bmatrix} \\
&= \begin{bmatrix} A_{2,1} \\ A_{2,2} \\ A_{2,3} \end{bmatrix}
\end{aligned}
$$

8. (2 points) What about $\frac{\partial f(A,x)}{\partial x}$? Since f has two outputs and x is three dimensional, the derivative must reflect the change of each output of f for each input in x. So this requires us to choose a matrix of dimension 2×3 or 3×2. To make this choice recall that for some scalar function $f(x)$ and some small input change to x, Δx, we can create the approximation $f(x + \Delta x) \approx f(x) + \frac{df(x)}{dx}\Delta x$. When working with vectors and matrices we have the approximation $f(x + \Delta x) \approx f(x) + \frac{df(x)}{dx}^\top x$. For $x \in \mathbb{R}^n$, the inner product $\frac{df(x)}{dx}^\top x$ only works if $\frac{df(x)}{dx} \in \mathbb{R}^n$. In our case we have $f(A, x + \Delta x) \approx f(A, x) + \frac{\partial f(A,x)}{\partial x}^\top \Delta x$. Thus, $\frac{\partial f(A,x)}{\partial x} \in \mathbb{R}^{3\times 2}$ for the approximation to be valid. Meaning

$$
\frac{\partial f(A, x)}{\partial x} = \begin{bmatrix} \frac{\partial f(A,x)_1}{\partial x_1} & \frac{\partial f(A,x)_2}{\partial x_1} \\ \frac{\partial f(A,x)_1}{\partial x_2} & \frac{\partial f(A,x)_2}{\partial x_2} \\ \frac{\partial f(A,x)_1}{\partial x_3} & \frac{\partial f(A,x)_2}{\partial x_3} \end{bmatrix}.
$$

   Give an expression for this derivative.

Answer:

$$\frac{\partial f(A,x)}{\partial x} = \begin{bmatrix} A_{1,1} & A_{2,1} \\ A_{1,2} & A_{2,2} \\ A_{1,3} & A_{2,3} \end{bmatrix}$$

9. (3 points) Recall that the expectation (mean or average) of a discrete random variable X is $\mathbf{E}[X] = \sum_{x \in \mathcal{X}} \Pr(X = x)x$. Similarly, the expectation of a function $f(x)$ is

$$\mathbf{E}[f(X)] = \sum_{x \in \mathcal{X}} \Pr(X = x) f(x).$$

Consider the function $f\colon \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ that takes as input x and some scalar w. The derivative $\frac{\partial}{\partial w}\mathbf{E}[f(X,w)]$ says in which direction should w be changed to increase the output of f in expectation over the random variable X. Give an expression for this derivative.

Answer:

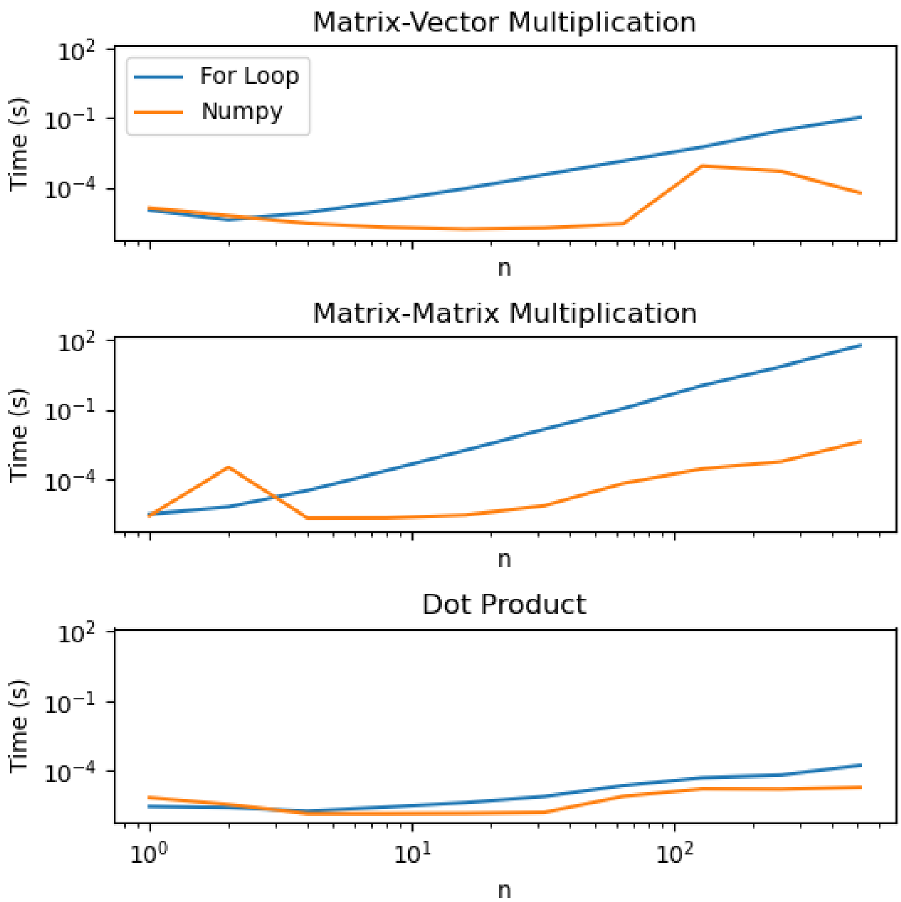
$$\begin{aligned}
\frac{\partial}{\partial w}\mathbf{E}[f(X,w)] &= \frac{\partial}{\partial w}\sum_{x \in \mathcal{X}} \Pr(X = x) f(x,w) \\
&= \sum_{x \in \mathcal{X}} \frac{\partial \Pr(X = x) f(x,w)}{\partial w} \\
&= \sum_{x \in \mathcal{X}} f(x,w) \underbrace{\frac{\partial \Pr(X = x)}{\partial w}}_{=0} + \Pr(X = x)\frac{\partial f(x,w)}{\partial w} \\
&= \sum_{x \in \mathcal{X}} \Pr(X = x)\frac{\partial f(x,w)}{\partial w} \\
&= \mathbf{E}\left[\frac{\partial}{\partial w} f(X,w)\right]
\end{aligned}$$

2 Linear Algebra (15 points)

Complete the linear algebra file, `linear_algebra.py`. This file serves as review of the linear algebra basics you will need to use in this course, e.g., dot products, matrix-vector products, and matrix-matrix products. You will implement a custom version of these functions written in python using for loops. These functions will not be vectorized code meaning it uses scalar operations of each element of the vector, e.g., you will implement the dot product of x and y as $\sum_{i=1}^{n} x_i y_i$. You will then compare the speed of these functions to their corresponding versions in the numpy package.

The goal for this question is for you to build intuition about the basic linear algebra operations and observe just how slow python is. The numpy package runs optimized C code. Python also compiles to C code, but it is unable to perform any code optimization, which makes it incredibly slow. While you could correctly implement deep learning models using pure python functions and for loops, it would be so slow that you would not be able to complete the assignments on time. As such you will need to use the optimized libraries.

After successfully implementing the code and running the file, it will produce a plot showing the timings for each operation with different sizes for the vectors and matrices. Include this plot in your submission pdf.


Matrix-Vector Multiplication
For Loop
Numpy
Time (s)
n
Matrix-Matrix Multiplication
Time (s)
n
Dot Product
Time (s)
n