CSE 446 Section 1

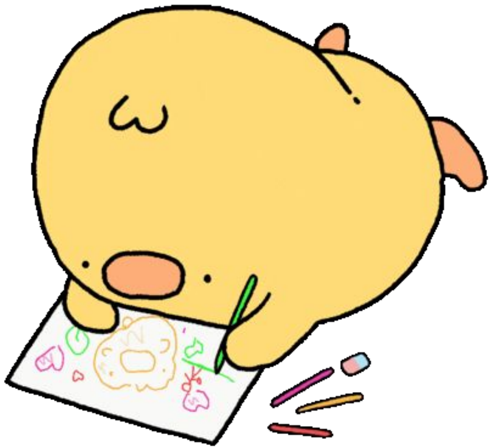
TA: Yufei Zhang

Agenda

- Intros
- Icebreaker (I promise this won't be that bad)
- Handout
  - Problem 2.1
  - Problem 1

What to expect in Section

- Casual setting
- Content review
- Ask questions
- Practice problems

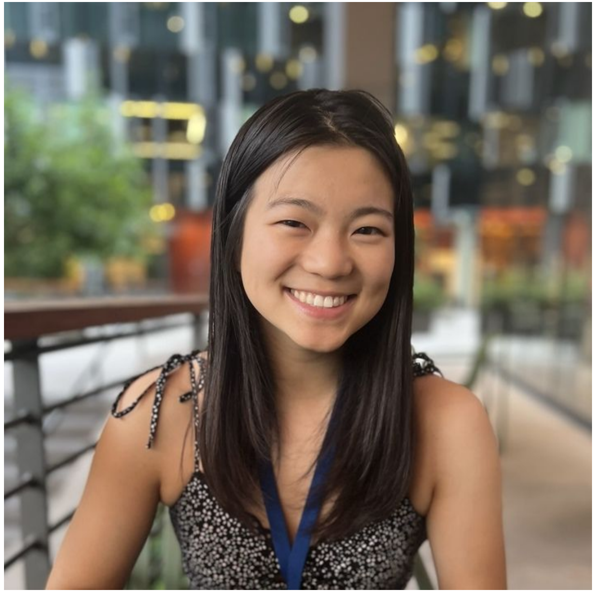


Meet your TA

Yufei Zhang

BS/MS Student, NLP Research

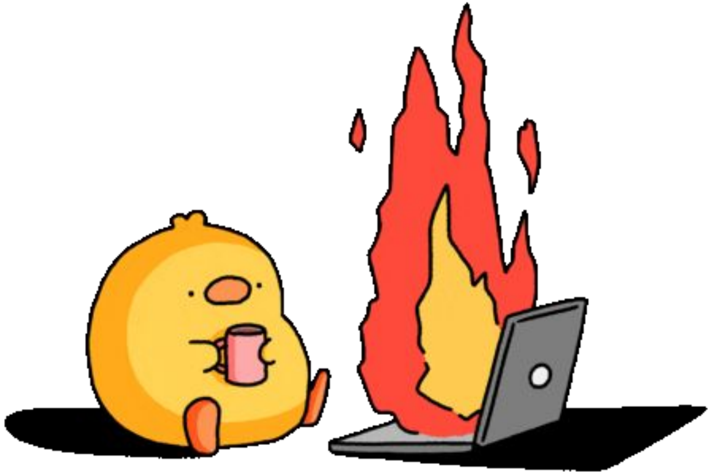
OH: Mon 4-5 PM, Gates 152



IN-N-OUT BURGER

Say your name,
your **favorite food/drink**
or an **interesting hobby**
you have

I’m here to help you succeed!



Practice Time!

Problem 2

Quick Matrix Algebra/Calculus Refresher

Unsure if this has been taught before to you all (when I took 208, I definitely did not learn it)

Note: For the matrix calculus section to the right, **B** is a constant matrix.

Thank you: Kirsty McNaught

Rule	Comments
$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$	order is reversed, everything is transposed
$(\mathbf{a}^T\mathbf{Bc})^T = \mathbf{c}^T\mathbf{B}^T\mathbf{a}$	as above
$\mathbf{a}^T\mathbf{b} = \mathbf{b}^T\mathbf{a}$	(the result is a scalar, and the transpose of a scalar is itself)
$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$	multiplication is distributive
$(\mathbf{a}+\mathbf{b})^T\mathbf{C} = \mathbf{a}^T\mathbf{C}+\mathbf{b}^T\mathbf{C}$	as above, with vectors
$\mathbf{AB} \neq \mathbf{BA}$	multiplication is **not** commutative

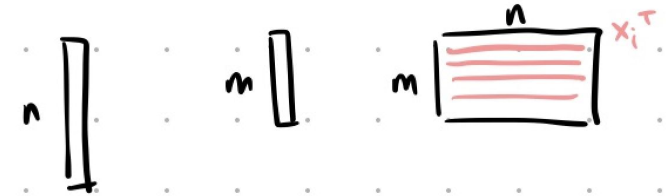
Scalar derivative			Vector derivative		
$f(x)$	$\rightarrow$	$\frac{\mathrm{d}f}{\mathrm{d}x}$	$f(\mathbf{x})$	$\rightarrow$	$\frac{\mathrm{d}f}{\mathrm{d}\mathbf{x}}$
bx	$\rightarrow$	b	$\mathbf{x}^T\mathbf{B}$	$\rightarrow$	$\mathbf{B}$
bx	$\rightarrow$	b	$\mathbf{x}^T\mathbf{b}$	$\rightarrow$	$\mathbf{b}$
x^2	$\rightarrow$	$2x$	$\mathbf{x}^T\mathbf{x}$	$\rightarrow$	$2\mathbf{x}$
bx^2	$\rightarrow$	$2bx$	$\mathbf{x}^T\mathbf{B}\mathbf{x}$	$\rightarrow$	$2\mathbf{B}\mathbf{x}$

Let $X \in \mathbb{R}^{m \times n}$. X may not have full rank. We explore properties about the four fundamental subspaces of X.

2.1. Summation form v.s. Matrix form

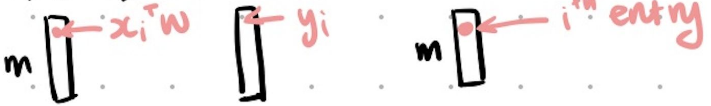
(a) Let $w \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$. Let $x_i^\top$ denote each row in X and y_i in Y. Show $\|Xw - Y\|_2^2 = \sum_{i=1}^m (x_i^\top w - y_i)^2$.

$w \in \mathbb{R}^n$ $\quad Y \in \mathbb{R}^n$ $\quad X \in \mathbb{R}^{m \times n}$



let $r := Xw - Y \in \mathbb{R}^m$

$(m \times n)(n \times 1) \Rightarrow (m \times 1) \Rightarrow Xw - Y$



then $r_i := (Xw - Y)_i = x_i^T w - y_i$

Apply def ℓ_2 norm:

$$\|r\|_2 = \sqrt{r^T r} \xrightarrow{\text{square}} \|r\|_2^2 = r^T r = \sum_{i=1}^m r_i^2$$

↑ just plug in

$$= \sum_{i=1}^m (x_i^T w - y_i)^2$$

Let $X \in \mathbb{R}^{m \times n}$. X may not have full rank. We explore properties about the four fundamental subspaces of X.

2.1. Summation form v.s. Matrix form

(b) Let $L(w) = \|Xw - Y\|_2^2$. What is $\nabla_w L(w)$? (Hint: You can use either summation or matrix form from first sub-problem).

Rule
$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
$(\mathbf{a}^T\mathbf{B}\mathbf{c})^T = \mathbf{c}^T\mathbf{B}^T\mathbf{a}$
$\mathbf{a}^T\mathbf{b} = \mathbf{b}^T\mathbf{a}$
$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$
$(\mathbf{a}+\mathbf{b})^T\mathbf{C} = \mathbf{a}^T\mathbf{C}+\mathbf{b}^T\mathbf{C}$
$\mathbf{AB} \neq \mathbf{BA}$

Scalar derivative			Vector derivative		
$f(x)$	$\rightarrow$	$\frac{df}{dx}$	$f(\mathbf{x})$	$\rightarrow$	$\frac{df}{d\mathbf{x}}$
bx	$\rightarrow$	b	$\mathbf{x}^T\mathbf{B}$	$\rightarrow$	$\mathbf{B}$
bx	$\rightarrow$	b	$\mathbf{x}^T\mathbf{b}$	$\rightarrow$	$\mathbf{b}$
x^2	$\rightarrow$	$2x$	$\mathbf{x}^T\mathbf{x}$	$\rightarrow$	$2\mathbf{x}$
bx^2	$\rightarrow$	$2bx$	$\mathbf{x}^T\mathbf{B}\mathbf{x}$	$\rightarrow$	$2\mathbf{B}\mathbf{x}$

Let $X \in \mathbb{R}^{m \times n}$. X may not have full rank. We explore properties about the four fundamental subspaces of X.

2.1. Summation form v.s. Matrix form

(b) Let $L(w) = \|Xw - Y\|_2^2$. What is $\nabla_w L(w)$? (Hint: You can use either summation or matrix form from first sub-problem).

$$Lw = \|Xw - Y\|_2^2$$

$$\nabla_w Lw = \nabla_w \left(\sqrt{(Xw-Y)^T(Xw-Y)}\right)^2$$

$$= \nabla_w (Xw-Y)^T(Xw-Y)$$

$$= \nabla_w (w^TX^T - Y^T)(Xw-Y)$$

constant - transpose

$$= \nabla_w (w^TX^TXw - w^TX^TY - Y^TXw + Y^TY)$$

$$= \nabla_w (w^TX^TXw - 2w^TX^TY + Y^TY) \quad 0$$

$$= 2X^TXw - 2X^TY$$

$$= X^T(2Xw - 2Y)$$

can take derivative wrt to either one - just be consistent

Problem 1

Problem 1

(a) You've just started a new exercise regimen. You start on the 2nd floor of CSE1, and then make a random choice:

- With probability p_1 you run up 2 flights of stairs.
- With probability p_2 you run up 1 flight of stairs.
- With probability p_3 you walk down 1 flight of stairs.

Where $p_1 + p_2 + p_3 = 1$.

You will do two iterations of your exercise scheme (with each draw being independent). Let X be the floor you're on at the end of your exercise routine. Recall you start on floor 2.

Problem 1

(i) Let Y be the difference between your ending floor and your starting floor in one iteration. What is $\mathbb{E}[Y]$ (in terms of p_1, p_2, p_3)?

Problem 1

(i) Let Y be the difference between your ending floor and your starting floor in one iteration. What is $\mathbb{E}[Y]$ (in terms of p_1, p_2, p_3)?

Solution:

Recall for a random variable $X, \mathbb{E}[X] = \sum_i x_i \cdot p_i$.
So $\mathbb{E}[Y] = 2 \cdot p_1 + 1 \cdot p_2 + (-1) \cdot p_3$

Problem 1

(i) Let Y be the difference between your ending floor and your starting floor in one iteration. What is $\mathbb{E}[Y]$ (in terms of p_1, p_2, p_3)?

Solution:

Recall for a random variable $X, \mathbb{E}[X] = \sum_i x_i \cdot p_i$.
So $\mathbb{E}[Y] = 2 \cdot p_1 + 1 \cdot p_2 + (-1) \cdot p_3$

(ii) What is $\mathbb{E}[X]$ (use your answer from the previous part)

Problem 1

(i) Let Y be the difference between your ending floor and your starting floor in one iteration. What is $\mathbb{E}[Y]$ (in terms of p_1, p_2, p_3)?

Solution:

Recall for a random variable $X, \mathbb{E}[X] = \sum_i x_i \cdot p_i$.
So $\mathbb{E}[Y] = 2 \cdot p_1 + 1 \cdot p_2 + (-1) \cdot p_3$

(ii) What is $\mathbb{E}[X]$ (use your answer from the previous part)

Solution:

Since we start at floor 2, we can take 2 and add the difference ($\mathbb{E}[Y]$) twice to get our expected floor at the end of the routine.
$\mathbb{E}[X] = 2 + \mathbb{E}[Y] + \mathbb{E}[Y] = 2 + 2\mathbb{E}[Y]$

Problem 1

(iii) You change your scheme: instead of doing two independent iterations, you decide the second iteration of your regimen will just use the same random choice as your first (in particular they are no longer independent!). Does $\mathbb{E}[X]$ change? (Optional)

Problem 1

(iii) You change your scheme: instead of doing two independent iterations, you decide the second iteration of your regimen will just use the same random choice as your first (in particular they are no longer independent!). Does $\mathbb{E}[X]$ change? (Optional)

Solution:

No! We can say using the same choice as the first will effectively double Y, thus by linearity of expectation, $\mathbb{E}[X] = 2 + \mathbb{E}[2Y] = 2 + 2\mathbb{E}[Y]$