- The goal of written HW1 is to give you an idea of the level of mathematical knowledge and maturity expected in this course. You should have seen all this material before; the goal of this homework is to encourage you to go back to some of the material and refresh your memory. You will find the primers released on Canvas helpful.

- CIS 5190 is an Applied Machine Learning course and aims to build strong practitioners of machine learning. A good ML practitioner should know the mathematics that drives ML algorithms, and your performance in this class will depend significantly on your ability to understand mathematical descriptions of ML algorithms in the lectures. **If you find that you are not familiar with more than 30% of the material in this problem set, you should strongly reconsider taking the course.** Speak to the instructor / TAs if you have concerns. If you've seen this material before but just need a refresher, the primers will help you.

- While we encourage discussion within and outside the class such as on Ed Discussion and in your cohort sessions, in the specific case of HW1, you are on your own. It's a test of your level of readiness to take this course, so please work on it independently. Please use Ed Discussion if you have clarifications about this homework.

- Keep all solutions brief and clear.

- You are encouraged to format your solutions using LaTeX. You'll find some pointers to resources for learning LaTeX among the primers. Handwritten solutions are permitted, but remember that you bear the risk that we may not be able to read your work and grade it properly — do not count on providing post hoc explanations for illegible work. You will submit your solution manuscript for written HW1 as a single PDF file.

- The homework is **due at 8:00 PM** on the due date. We will be using Gradescope for collecting the homework assignments. Please submit your solution manuscript as a PDF file via Gradescope. Post on Piazza and contact the TAs if you are having technical difficulties in submitting the assignment.

- **If you are on the waiting list**, we recommend that you still work on and submit the assignment.

1 Multiple Choice & Written Questions

Note: You do not need to show work for multiple choice questions. If formatting your answer in LaTeX, use our LaTeX template `hw1_template.tex` (This is a read-only link. You'll need to make a copy before you can edit. Make sure you make only private copies.).

1. [Calculus] Let $f(x_1, ..., x_n) = \sum_{i=1}^{n} x_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j$

   a. (4pts) Find $\frac{\partial f}{\partial x_k}$, the partial derivative of f with respect to x_k.

   A) $\frac{\partial f}{\partial x_k} = 1 - \sum_{i=1}^{n} x_i$

   B) $\frac{\partial f}{\partial x_k} = 1 - 2x_k - \sum_{i=1, i \neq k}^{n} x_i$

   C) $\frac{\partial f}{\partial x_k} = n - \sum_{i=1}^{n} x_i$

   D) $\frac{\partial f}{\partial x_k} = n - 2x_k - \sum_{i=1, i \neq k}^{n} x_i$

   b. (4pts) Find $\vec{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ that maximizes f.

   A) $\vec{x} = (\frac{1}{n+1}, ..., \frac{1}{n+1})$

   B) $\vec{x} = (\frac{1}{n}, ..., \frac{1}{n})$

   C) $\vec{x} = (1, ..., 1)$

   D) f is unbounded

2. [Probability] Consider an n sided die with sides labeled $1, 2, ..., n$

   a. (4pts) What is the probability of rolling the first 1 on the $(k+1)$-th toss?

   A) $(1 - \frac{1}{n}) \cdot \frac{1}{n^k}$

   B) $(1 - \frac{1}{n})^k \cdot \frac{1}{n}$

   C) $(1 - \frac{1}{n})^k \cdot \frac{1}{n^k}$

   D) $(1 - \frac{1}{n}) \cdot \frac{1}{n}$

   b. (4pts) What is the expected number of times of rolling 1 if the die is rolled k times?

   A) $\frac{k}{n}$

   B) k^n

   C) n^k

   D) $(1 - \frac{1}{n})^k$

3. [Probability] Assume we know the following facts about COVID-19 testing: Among the population of people who are tested:

   - 99% of infected people test positive (this is called the true positive rate)
   - 1% of uninfected people test positive (this is the false positive rate)
   - 10% are infected in this population

   With this information, please answer the two questions below:

a. (4pts) John tests positive. What is the probability that he is infected?

A) ≈ 0.99

B) ≈ 0.98

C) ≈ 0.92

D) ≈ 0.11

b. (4pts) With all other values remaining the same, what should the false positive rate change to, for John's probability of being infected to change to 0.95?

A) $\approx 0.058\%$

B) $\approx 0.58\%$

C) $\approx 0.1\%$

D) $\approx 5.8\%$

4. [Probability] The variance of a random variable X is defined as

$$Var(X) = E[(X - E[X])^2] \ .$$

a. (4pts) If $E[X] = 1$ and $E[X^2] = 2$, and $Y = a + bX$, where a and b are constants, then what is the variance of Y?

A) $Var(Y) = 2b^2$

B) $Var(Y) = a^2 + b^2$

C) $Var(Y) = b$

D) $Var(Y) = b^2$

b. (9pts) Prove that $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. Please show your work in detail.

5. [Linear Algebra] Let

$$v = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

$$M = \begin{bmatrix} 3 & 6 \\ 4 & 7 \\ 5 & 8 \end{bmatrix}$$

a. (4pts) What is the length $\|v\|_2$ of the vector v (sometimes written as $\|v\|$)?

A) $\sqrt{3}$

B) $\sqrt{5}$

C) 3

D) 5

b. (4pts) What is the result of the vector-matrix product $v^\top M$, where $\top$ indicates the transpose operation?

A) Undefined, sizes do not align.

B) $\begin{bmatrix} -3 & 12 \\ -4 & 14 \\ -5 & 16 \end{bmatrix}$

C) $\begin{bmatrix} -12 & 42 \end{bmatrix}$

D) $\begin{bmatrix} -4 & 14 \end{bmatrix}$

c. (4pts) What is the result of the matrix-vector product Mv?

A) Undefined, sizes do not align.

B) $\begin{bmatrix} 12 \\ 14 \\ 16 \end{bmatrix}$

C) $\begin{bmatrix} 12 & 14 & 16 \end{bmatrix}$

D) $\begin{bmatrix} 9 \\ 10 \\ 11 \end{bmatrix}$

2 Python Programming Questions

A Google Colab notebook is linked in the "HW1 Coding" assignment on Canvas. This will tell you everything you need to do, and provide starter code. See the recording from the first day of class for how to use Colab and how to submit homeworks.

- Submit the notebook as .ipynb file to the coding portion of the Gradescope submission.

- Remember to download and embed the figure for question 4 into the *written problem set solution*.