# CMSC 678 Machine Learning - Fall 2022
# Homework Assignment 1
# Due September $22^{nd}$ by 2:30pm

Submit a single PDF file to the TA via slack with your answers with a direct message to Mohammad Eskandari.

1. Suppose we have a dataset of the form $[(x_1, y_1) \dots (x_n, y_n)]$ where $x_i \in \mathbb{R}^d$ and $y \in \mathbb{R}$. We will learn a linear function of $x$ of the form $\hat{y} = \theta x$ where $\theta \in \mathbb{R}^d$. If the loss function is the absolute difference $|y - \hat{y}|$, write an expression for $F(\theta)$, the total loss over the entire datase.

2. Consider the following loss function on vectors $w \in \mathbb{R}^4$:

$$L(w) = w_1^2 + 2w_2^2 + w_3^2 - 2w_3w_4 + w_4^2 + 2w_1 - 4w_2 + 4$$

   - What is $\nabla L(w)$?
   - Suppose we use gradient descent to minimize this function and that the current estimate is $(0, 0, 0, 0)$. If the step size is $\eta$, what is the next estimate?

3. These questions are taken from
   http://users.umiacs.umd.edu/~hal/courses/2013S_ML/math4ml.pdf
   which is a great refresher on math for machine learning practitioners. It is a good resource if you are having difficulty with any of these questions.

   - What is the derivative of $f(x) = \exp(-\frac{1}{2}x^2)$?
   - What is the derivative of $f(x) = \log(x^2 + x - 1)$?
   - Given $n$ points of the form $(x_i, y_i)$ where both $x_i$ and $y_i$ are real numbers, compute the partial derivative $\frac{\partial J}{\partial b}$, where $J(m, b) = \sum_{i=1}^{n}((mx_i + b) - y_i)^2$.

4. Consider the following function:

$$g(z_1, z_2) = (a - z_1)^2 + b(z_2 - z_1^2)^2$$

   Write a program to numerically minimize this function. Turn in all of your code. Experiment and report on your experiments. Specifically:

- Derive the partial derivatives for the function and include the derivations in your write-up.
- Implement the function and its gradient.
- Write code to minimize $g$ using gradient descent with a fixed learning rate.
- Run your code using $(a = 1, b = 100)$ from a few different initial values for $(z_1, z_2)$ and a few different learning rates, and repeat that process using one other set of values for $a$ and $b$.
- Write up your results in a report using prose, graphs, and tables as needed to explain the optimal values $(z_1^*, z_2^*)$ found for each run. Discuss the impact of the starting point and the learning rate.