CS 577 – Fall 2025 – Homework 1

Throughout this homework, we let $g(t) = \max\{0, t\}$ denote the ReLU activation function. The derivative of g is $\frac{dg}{dt} = 1$ if $t > 0$ and $= 0$ otherwise.

Problem 1.a – [2] point(s). Plot over $x \in [-2, 2]$ the function

$$f(x;\theta) := g(x + b_1) - g(-(x + b_2)) + b_3 \text{ where } \theta = [b_1, b_2, b_3] = [-1, 1, 1]$$

(Hint: It is piecewise linear function with 3 linear pieces.)

Answer for 1a.

Problem 1.b – [4] point(s). (Continuation of Problem 1.a) The empirical risk is $J(\theta) = \frac{1}{m}\sum_i^m J_i(\theta)$ where the i-th contribution to the empirical risk with the square error loss is defined as

$$J_i(\theta) = \left(y^{(i)} - f\left(x^{(i)};\theta\right)\right)^2.$$

Consider the following dataset with $m = 3$:

$$x^{(1)} = 2, \quad x^{(2)} = -2, \quad x^{(3)} = 0 \tag{1}$$

$$y^{(1)} = 5, \quad y^{(2)} = -1, \quad y^{(3)} = 2 \tag{2}$$

Compute the numerical values of $\frac{\partial J_i}{\partial \theta}(\theta)$ for each $i = 1, 2, 3$, where θ is as in part a. Show all work. (Hint: if your answer is correct, then you should have

$$\frac{\partial J}{\partial b_1}(\theta) = -2, \quad \frac{\partial J}{\partial b_2}(\theta) = \frac{2}{3}, \quad \frac{\partial J}{\partial b_3}(\theta) = -2 \tag{3}$$

where $J(\theta) = \frac{1}{3}\sum_{i=1}^n J_i(\theta)$.)

Answer for 1b.

Problem 1.c – [4] point(s). (Continuation of Problem 1.b) Calculate, showing work, θ after 1 gradient descent update with step size $\eta = \frac{3}{4}$. Plot over $x \in [-2, 2]$ the function $f(x;\theta)$ with the updated parameters. (Hint: you should get [0.5, 0.5, 2.5])

Answer for 1c.

Problem 2.a – [1] point(s). Plot over $x \in [-1, 1]$ the function

$$f(x;\theta) := g\left(w_2 \cdot g(w_1 \cdot x + b_1) + b_2\right) \text{ where } \theta = [b_1, w_1, b_2, w_2] = [0.5, 1, 1, -1].$$

Answer for 2a.

Problem 2.b – [3] point(s). (Continuation of Problem 2.a) Suppose that $x^{(i)}, y^{(i)} \in \mathbb{R} \times \mathbb{R}$ is a dataset where $i = 1, \ldots, m$. Let $J_i(\theta) = (y_i - f(x_i;\theta))^2$. Compute all partial derivatives (algebraically):

$$\frac{\partial J_i}{\partial b_1}(\theta), \frac{\partial J_i}{\partial w_1}(\theta), \frac{\partial J_i}{\partial b_2}(\theta), \frac{\partial J_i}{\partial w_2}(\theta).$$

Answer for 2b.

Problem 2.c – [6] point(s). Implement gradient descent in numpy in the Python notebook `hw1.ipynb` provided by filling in missing code block at ``` ``YOUR CODE HERE'' ```.