MATH 629 Homework 1

Paul Zhang

October 24, 2024

Problem 1

The weights and biases are given by

$$\begin{aligned}
\mathbf{W}^{(1)} &= \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \\
\mathbf{b}^{(1)} &= \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \\
\mathbf{W}^{(2)} &= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \\
b^{(2)} &= -2.4
\end{aligned}$$

Problem 2

The dimensions of X, Y, W, B are as follows (assuming that the number of features is d):

- $X \in \mathbb{R}^{N \times d}$
- $Y = [y_1, y_2, \cdots, y_N]^T \in \mathbb{R}^{N \times 1}$
- $W \in \mathbb{R}^{d \times 1}$
- $B \in \mathbb{R}^{N \times 1}$

With the notations above, we can define $\mathcal{L}(Y, T)$ as

$$\mathcal{L}(Y, T) := [\mathcal{L}(y_1, t_1), \mathcal{L}(y_2, t_2), \cdots, \mathcal{L}(y_N, t_N)]^T \in \mathbb{R}^{N \times 1}$$

Accoringly, we have that

$$\mathcal{E} = \frac{1}{N} \mathcal{L}(Y, T)^T \cdot e^{N \times 1}$$

and, obviously,

$$Y = XW + B$$

Therefore, the derivatives are computed as

$$\begin{aligned}
\frac{\partial \mathcal{E}}{\partial Y} &= \frac{1}{N} \sin(Y - T) \\
\frac{\partial \mathcal{E}}{\partial W} &= \frac{\partial \mathcal{E}}{\partial Y} \frac{\partial Y}{\partial W} = \frac{1}{N} X^T \sin(Y - T) \\
\frac{\partial \mathcal{E}}{\partial B} &= \frac{\partial \mathcal{E}}{\partial Y} \frac{\partial Y}{\partial B} = \frac{1}{N} \sin(Y - T)
\end{aligned}$$

Problem 3

We start with showing that the absolute loss is indeed a squared loss with reassigned weights.

$$\sum_{n=1}^{N} |y_n - \mathbf{X_n}\beta| = \sum_{n=1}^{N} \frac{1}{|y_n - \mathbf{X_n}\beta|} (y_n - \mathbf{X_n}\beta)^2 =: \sum_{n=1}^{N} s_n (y_n - \mathbf{X_n}\beta)^2 \tag{1}$$

In order to estimate the s_n's, we can utilize the following algorithm:

- Initialize OLS weights, β^0
- Compute residuals $|y_n - \mathbf{X_n}\beta|$
- Update weights w_n^i with $w_n^i = \frac{1}{|y_n - \mathbf{X_n}\beta|}$
- Solve the WLS problem, $\beta^{i+1} = \text{argmin} \sum_n w_n^i (y_n - X_n \beta^i)$
- Stop when $|\beta^{i+1} - \beta^i| < \epsilon$, where ϵ is a pre-determined threshold.

Problem 4

Figure 1: Implementation of XOR