CS 4501: Optimization - Assignment 7

Your name and email

Consider an n-layer feedforward neural network (FNN) parameterized by

$$W = (W_0, W_1, \dots, W_{n-1}),$$
$$b = (b_0, b_1, \dots, b_{n-1}),$$

where,

$$W_i \in \mathbb{R}^{d_i \times d_{i+1}},$$
$$b_i \in \mathbb{R}^{d_{i+1}}, \quad i = 0, 1, \dots, n-1.$$

Here the sequence $d \doteq (d_0, d_1, \dots, d_n)$ specifies the number of units in each layer, with d_0 being the size of the input x and d_n being the size of the output. Let $\sigma : \mathbb{R} \to \mathbb{R}$ be the activation function. Notably, we use the convention that if x is matrix / vector, $\sigma(x)$ denotes a matrix / vector obtained by elementwise application of σ in x.

Given an input $x \in \mathbb{R}^{d_0}$, the output $f(x; W, b) \in \mathbb{R}^{d_n}$ is given by

$$\begin{aligned} x_0 &\doteq x \\ z_{k+1} &\doteq W_k^\top x_k + b_k, \\ x_{k+1} &\doteq \sigma(z_{k+1}), \quad k = 0, 1, \dots n-1, \\ f(x; W, b) &\doteq x_n. \end{aligned}$$

Suppose the target output is $y \in \mathbb{R}^{d_n}$, the loss function is then defined as

$$L(x, y) = \frac{1}{2} \| f(x; W, b) - y \|_2^2.$$

Task 1 (5pt): Compute the gradients analytically

Proof. (SZ: You can use recursive expression in expressing the gradients.)

$$\frac{\mathrm{d}L(x,y)}{\mathrm{d}W_k} =$$
$$\frac{\mathrm{d}L(x,y)}{\mathrm{d}b_k} =$$

□

Task 2 (25pt): Compute the gradients in python

(SZ: You cannot use any auto-diff software in your submission. You can of course use auto-diff software to verify your computation. Just remember to delete them in your submission.)