Problem Set 4 - Gradients and Backpropagation

DS542 - DL4DS

Spring, 2025

Note: Refer to Chapter 7 in *Understanding Deep Learning*.

Problem 4.1 (3 points)

Consider the case where we use the logistic sigmoid function as an activation function, defined as:

$$h = \sigma(z) = \frac{1}{1 + e^{-z}}. \tag{1}$$

Compute the derivative $\frac{\partial h}{\partial z}$. What happens to the derivative when the input takes (i) a large positive value and (ii) a large negative value?

$$\frac{d\sigma(z)}{dz} = \frac{d}{dz}\left(\frac{1}{1 + e^{-z}}\right)$$

$$\frac{d\sigma(z)}{dz} = \sigma(z)(1 - \sigma(z))$$

$$\frac{\partial h}{\partial z} = \sigma(z)(1 - \sigma(z)). \tag{2}$$

- **For large positive z:** As $z \to \infty$, we have $\sigma(z) \to 1$. Thus, $\sigma(z)(1 - \sigma(z)) \to 1(1 - 1) = 0$.
- **For large negative z:** As $z \to -\infty$, we have $\sigma(z) \to 0$. Thus, $\sigma(z)(1 - \sigma(z)) \to 0(1 - 0) = 0$.

The derivative $\frac{\partial h}{\partial z}$ is maximized at $z = 0$, where it attains the value $\frac{1}{4}$. It decreases towards 0 as $z \to \pm\infty$, meaning that the sigmoid function saturates and has very small gradients for extremely large or small values of z.

Problem 4.2 (3 points)

Calculate the derivative $\frac{\partial \ell_i}{\partial f[x_i,\phi]}$ for the binary classification loss function:

$$\ell_i = -(1-y_i)\log[1-\sigma(f[x_i,\phi])] - y_i \log[\sigma(f[x_i,\phi])], \tag{3}$$

where the function $\sigma(\cdot)$ is the logistic sigmoid, defined as:

$$\sigma(z) = \frac{1}{1+\exp(-z)}. \tag{4}$$

$$\frac{\partial \ell_i}{\partial f(x_i,\phi)} = -(1-y_i)\frac{1}{1-\sigma(f(x_i,\phi))}\cdot\left(-\frac{\partial \sigma(f(x_i,\phi))}{\partial f(x_i,\phi)}\right) - y_i \frac{1}{\sigma(f(x_i,\phi))}\cdot\frac{\partial \sigma(f(x_i,\phi))}{\partial f(x_i,\phi)}$$

$$\frac{\partial \sigma(f(x_i,\phi))}{\partial f(x_i,\phi)} = \sigma(f(x_i,\phi))(1-\sigma(f(x_i,\phi))).$$

$$\frac{\partial \ell_i}{\partial f(x_i,\phi)} = (1-y_i)\frac{\sigma(f(x_i,\phi))(1-\sigma(f(x_i,\phi)))}{1-\sigma(f(x_i,\phi))} - y_i\frac{\sigma(f(x_i,\phi))(1-\sigma(f(x_i,\phi)))}{\sigma(f(x_i,\phi))}$$

$$\frac{\partial \ell_i}{\partial f(x_i,\phi)} = (1-y_i)\sigma(f(x_i,\phi)) - y_i(1-\sigma(f(x_i,\phi))).$$

$$\frac{\partial \ell_i}{\partial f(x_i,\phi)} = \sigma(f(x_i,\phi)) - y_i. \tag{5}$$

gradient represents the difference between the predicted probability and the true label