# 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?

Answer:

To compute the derivative of the logistic sigmoid function:

$$h = \sigma(z) = \frac{1}{1 + e^{-z}}, \tag{2}$$

Part (i): Compute the Derivative Differentiating $h$ with respect to $z$:

$$\frac{d}{dz}h = \frac{d}{dz}\left(\frac{1}{1 + e^{-z}}\right). \tag{3}$$

Using the quotient rule, let $f(z) = 1$ and $g(z) = 1 + e^{-z}$, so that:

$$\frac{d}{dz}h = \frac{f'g - fg'}{g^2}. \tag{4}$$

Since $f' = 0$ and $g' = -e^{-z}$, we get:

$$\frac{d}{dz}h = \frac{(0)(1 + e^{-z}) - (1)(-e^{-z})}{(1 + e^{-z})^2}. \tag{5}$$

Simplifying:

$$\frac{d}{dz}h = \frac{e^{-z}}{(1+e^{-z})^2}. \tag{6}$$

Rewriting $e^{-z}$ in terms of $h$:

$$e^{-z} = \frac{1}{e^z} = \frac{1}{\frac{1}{h}-1} = h(1-h) = \sigma(z)(1-\sigma(z)) \tag{7}$$

Thus, the derivative simplifies to:

$$\frac{\partial h}{\partial z} = h(1-h) = \sigma(z)(1-\sigma(z)) \tag{8}$$

Part (ii): Behavior for Large $z$

- **When** $z \to +\infty$:
$$h = \sigma(z) \approx 1. \tag{9}$$
$$h(1-h) \approx 1(1-1) = 0. \tag{10}$$

- **When** $z \to -\infty$:
$$h = \sigma(z) \approx 0. \tag{11}$$
$$h(1-h) \approx 0(1-0) = 0. \tag{12}$$

## Problem 4.2 (3 points)

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

$$\ell_i = -(1-y_i)\log[1-\sigma(f[x_i,\phi])] - y_i\log[\sigma(f[x_i,\phi])], \tag{13}$$

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

$$\sigma(z) = \frac{1}{1+\exp(-z)}. \tag{14}$$

Answer:

The binary classification loss function is defined as:

$$\ell_i = -(1-y_i)\log[1-\sigma(f[x_i,\phi])] - y_i\log[\sigma(f[x_i,\phi])]. \tag{15}$$

Taking the derivative with respect to $f[x_i,\phi]$, let $h = \sigma(f[x_i,\phi])$:

$$\frac{\partial \ell_i}{\partial f[x_i,\phi]} = -(1-y_i)\frac{1}{1-h}\cdot(-\sigma(f[x_i,\phi])(1-\sigma(f[x_i,\phi]))) - y_i\frac{1}{h}\cdot\sigma(f[x_i,\phi])(1-\sigma(f[x_i,\phi])). \tag{16}$$

Simplifying:

$$\frac{\partial \ell_i}{\partial f[x_i,\phi]} = \sigma(f[x_i,\phi])(1-\sigma(f[x_i,\phi]))\left(\frac{(1-y_i)}{1-\sigma(f[x_i,\phi])} - \frac{y_i}{\sigma(f[x_i,\phi])}\right). \tag{17}$$

Further simplification leads to:

$$\frac{\partial \ell_i}{\partial f[x_i,\phi]} = \sigma(f[x_i,\phi]) - y_i. \tag{18}$$

Conclusion: The gradient of the loss function with respect to $f[x_i,\phi]$ is simply $\sigma(f[x_i,\phi]) - y_i$. This result is useful in logistic regression and neural networks, where it serves as the basis for gradient-based optimization algorithms such as stochastic gradient descent.