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?

Solution. By chain rule,

$$\begin{aligned}\frac{\partial h}{\partial z} &= -(1 + e^{-z})^{-2} \cdot \frac{\partial}{\partial z}(1 + e^{-z}) \\ &= \frac{e^{-z}}{(1 + e^{-z})^2} = \frac{1}{e^z + e^{-z} + 2}.\end{aligned}$$

When the input z takes a large positive value, $e^z \to \infty$ and $e^{-z} \to 0$, and thus $\frac{\partial h}{\partial z}$ gets close to 0.

In contrast, when z takes a large negative value, e^{-z} is large and e^z is close to 0. The derivative also gets close to 0. □

Problem 4.2 (3 points)

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

$$\ell_i = -(1 - y_i) \log[1 - \sigma(f[x_i, \phi])] - y_i \log[\sigma(f[x_i, \phi])], \tag{2}$$

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

$$\sigma(z) = \frac{1}{1 + \exp(-z)}. \tag{3}$$

Solution. Let $z = f[x_i, \phi]$. Note that $1 - \sigma(z) = \dfrac{e^{-z}}{1 + e^{-z}}$, so

$$\frac{\partial \sigma}{\partial z} = \frac{e^{-z}}{(1 + e^{-z})^2} = \sigma(z)(1 - \sigma(z)).$$

Thus

$$\begin{aligned}
\frac{\partial \ell_i}{\partial z} &= -(1 - y_i)\frac{1}{1 - \sigma(z)}(-1)\frac{\partial \sigma}{\partial z} - y_i \frac{1}{\sigma(z)}\frac{\partial \sigma}{\partial z} \\
&= (1 - y_i)\sigma(z) - y_i(1 - \sigma(z)) \\
&= \sigma(z) - y_i \\
&= \frac{1}{1 + \exp(-f[x_i, \phi])} - y_i.
\end{aligned}$$

□