

Problem 7.2

From Definition in 7.9 we know:

$$\begin{aligned}\frac{\partial h_3}{\partial f_2} &= -\sin(f_2), \frac{\partial f_3}{\partial h_3} = \omega_3, \frac{\partial \ell_i}{\partial f_3} = 2(f_3 - y_i), \frac{\partial f_2}{\partial h_2} = \omega_2 \\ \frac{\partial h_2}{\partial f_1} &= \exp(f_1), \frac{\partial f_1}{\partial h_1} = \omega_1, \frac{\partial h_1}{\partial f_0} = \cos(f_0)\end{aligned}$$

For the following five chains we can give answers by chain backpropagation:

$$\begin{aligned}\frac{\partial \ell_i}{\partial f_2} &= \frac{\partial h_3}{\partial f_2} \left(\frac{\partial f_4}{\partial h_3} \frac{\partial \ell_i}{\partial f_3} \right) \\ &= -\sin(f_2) (\omega_3 2(f_3 - y_i)) \\ \frac{\partial \ell_i}{\partial h_2} &= \frac{\partial f_2}{\partial h_2} \left(\frac{\partial h_3}{\partial f_2} \frac{\partial f_3}{\partial h_3} \frac{\partial \ell_i}{\partial f_3} \right) \\ &= \omega_2 (-\sin(f_2) (\omega_3 2(f_3 - y_i))) \\ \frac{\partial \ell_i}{\partial f_1} &= \frac{\partial h_2}{\partial f_1} \left(\frac{\partial f_2}{\partial h_2} \frac{\partial h_3}{\partial f_2} \frac{\partial f_3}{\partial h_3} \frac{\partial \ell_i}{\partial f_3} \right) \\ &= \exp(f_1) (\omega_2 (-\sin(f_2) (\omega_3 2(f_3 - y_i)))) \\ \frac{\partial \ell_i}{\partial h_1} &= \frac{\partial f_1}{\partial h_1} \left(\frac{\partial h_2}{\partial f_1} \frac{\partial f_2}{\partial h_2} \frac{\partial h_3}{\partial f_2} \frac{\partial f_3}{\partial h_3} \frac{\partial \ell_i}{\partial f_3} \right) \\ &= \omega_1 (\exp(f_1) (\omega_2 (-\sin(f_2) (\omega_3 2(f_3 - y_i))))) \\ \frac{\partial \ell_i}{\partial f_0} &= \frac{\partial h_1}{\partial f_0} \left(\frac{\partial f_1}{\partial h_1} \frac{\partial h_2}{\partial f_1} \frac{\partial f_2}{\partial h_2} \frac{\partial h_3}{\partial f_2} \frac{\partial f_3}{\partial h_3} \frac{\partial \ell_i}{\partial f_3} \right) \\ &= \cos(f_0) (\omega_1 (\exp(f_1) (\omega_2 (-\sin(f_2) (\omega_3 2(f_3 - y_i))))))\end{aligned}$$

Problem 7.4

$$\frac{\partial \ell_i}{\partial f} = \frac{\partial}{\partial f} (y_i - f)^2 = 2(y_i - f) \frac{\partial f}{\partial f} = -2(y_i - f)$$

Problem 7.6

$$\frac{\partial \mathbf{z}}{\partial \mathbf{h}} = \frac{\partial}{\partial \mathbf{h}} (\beta + \mathbf{\Omega h}) = \mathbf{\Omega}^T$$

Reference: Matrix calculus

Problem 7.8

Heaviside function and rectangular function are problematic for network training with gradient-based methods because they are not differentiable at some points.

Problem 7.10

$$\frac{\partial a}{\partial z} = \begin{cases} \alpha & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \end{cases}$$

Problem 7.14

We are given a random variable "a" with $E[a] = 0$, $Var[a] = \sigma^2$, and a symmetric distribution around the mean. We want to prove that if we pass this variable through the ReLU function, $b = ReLU(a)$, then the second moment of the transformed variable is $E[b^2] = \frac{\sigma^2}{2}$.

By the definition of expectation, we have:

$$E[b^2] = E[ReLU(a)^2] = \int_{-\infty}^0 0^2 f(a) da + \int_0^{\infty} a^2 f(a) da$$

where "f(a)" is the probability density function (PDF) of the random variable "a".

The first integral vanishes because $ReLU(a) = 0$ when $a < 0$, so we have:

$$E[b^2] = \int_0^{\infty} a^2 f(a) da$$

Since the distribution of "a" is symmetric around 0, we know that:

$$P(a \geq 0) = P(a < 0) = \frac{1}{2}$$

and the distribution of a^2 is the same regardless of whether "a" is positive or negative. Thus, we can rewrite the expectation as:

$$E[b^2] = \int_0^{\infty} a^2 f(a) da = \frac{1}{2} E[a^2]$$

The variance of "a" is given by:

$$Var[a] = E[a^2] - (E[a])^2$$

Since $E[a] = 0$, we have:

$$Var[a] = \mathbb{E}[a^2]$$

Therefore, we can substitute $\mathbb{E}[a^2] = \sigma^2$ into the previous equation:

$$\mathbb{E}[b^2] = \frac{\sigma^2}{2}$$

Problem 7.15

If we initialize the weight and bias to zero, the gradient of the loss function with respect to the weight and bias will be zero. Therefore, the weight and bias will not be updated during the training process.