Neural Networks Homework 7

Sayeh Jarollahi (7073520, saja00006@stud.uni-saarland.de)
Mahsa Amani (7064006, maam00002@stud.uni-saarland.de)

December 15, 2024

Exercise 7.1

Proof. a)

$$h = ReLU(\begin{pmatrix} 0.1 & 0.4 \\ 0.2 & 0.5 \\ 0.3 & 0.6 \end{pmatrix} \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 0.1 \\ 0.2 \\ 0.3 \end{pmatrix}) = \begin{pmatrix} 0.23 \\ 0.4 \\ 0.57 \end{pmatrix}$$

$$y = sigmoid(\begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.5 & 0.3 & 0.4 \end{pmatrix} \begin{pmatrix} 0.23 \\ 0.4 \\ 0.57 \end{pmatrix} + \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}) = sigmoid(\begin{pmatrix} 0.398 \\ 0.663 \end{pmatrix}) = \begin{pmatrix} 0.5982 \\ 0.6599 \end{pmatrix}$$

b) We have to calculate the following term:

$$-\frac{1}{2}(log(0.5982) + log(0.3401)) = 0.79615$$

c) We use the chain rule for all of the derivatives. (Please note that we assumed we are using 1-index so $W_1(1,1)$ is the first left element of matrix):
We have to compute δ_2 (Error Signal at the Output Layer) The error signal is:

$$\delta_2 = \frac{1}{2}(y - t)$$

Substitute:

$$\delta_2 = \frac{1}{2}\begin{bmatrix} 0.5983 - 1 \\ 0.6591 - 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -0.4017 \\ 0.6591 \end{bmatrix}$$

$$\delta_2 = \begin{bmatrix} -0.20085 \\ 0.32955 \end{bmatrix}$$

The gradient with respect to W_2 is:

$$\frac{\partial L}{\partial W_2} = \delta_2 h^T$$

Substitute:

$$\frac{\partial L}{\partial W_2} = \begin{bmatrix} (-0.20085)(0.23) & (-0.20085)(0.4) & (-0.20085)(0.57) \\ (0.32955)(0.23) & (0.32955)(0.4) & (0.32955)(0.57) \end{bmatrix}$$

$$\frac{\partial L}{\partial W_2} = \begin{bmatrix} -0.0462 & -0.0803 & -0.1145 \\ 0.0758 & 0.1318 & 0.1878 \end{bmatrix}$$

Extracting $\frac{\partial L}{\partial W_2(1,1)}$:

$$\frac{\partial L}{\partial W_2(1,1)} = -0.0462$$

The gradient with respect to b_2 is:

$$\frac{\partial L}{\partial b_2} = \delta_2$$

$$\frac{\partial L}{\partial b_2(1)} = -0.20085$$

Now we go to the first layer gradients. The error signal for the first layer is:

$$\delta_1 = \frac{1}{2}\left(W_2^T \delta_2\right) \odot \text{ReLU}'(z_1)$$

Since $\text{ReLU}'(z_1) = 1$ (because $z_1 > 0$), we only compute $W_2^T \delta_2$:

$$W_2^T \delta_2 = \begin{bmatrix} (0.7)(-0.20085) + (0.5)(0.32955) \\ (0.2)(-0.20085) + (0.3)(0.32955) \\ (0.1)(-0.20085) + (0.4)(0.32955) \end{bmatrix}$$

$$W_2^T \delta_2 = \begin{bmatrix} 0.0242 \\ 0.0587 \\ 0.1117 \end{bmatrix}$$

$$\delta_1 = \frac{1}{2} \begin{bmatrix} 0.0242 \\ 0.0587 \\ 0.1117 \end{bmatrix} = \begin{bmatrix} 0.0121 \\ 0.0293 \\ 0.05585 \end{bmatrix}$$

The gradient with respect to W_1 is:

$$\frac{\partial L}{\partial W_1} = \delta_1 x^T$$

Substitute:

$$\frac{\partial L}{\partial W_1} = \begin{bmatrix} (0.0121)(0.5) & (0.0121)(0.2) \\ (0.0293)(0.5) & (0.0293)(0.2) \\ (0.05585)(0.5) & (0.05585)(0.2) \end{bmatrix}$$

$$\frac{\partial L}{\partial W_1} = \begin{bmatrix} 0.00605 & 0.00242 \\ 0.01465 & 0.00586 \\ 0.02793 & 0.01117 \end{bmatrix}$$

Extracting $\frac{\partial L}{\partial W_1(1,1)}$:

$$\frac{\partial L}{\partial W_1(1,1)} = 0.00605$$

The gradient with respect to b_1 is:

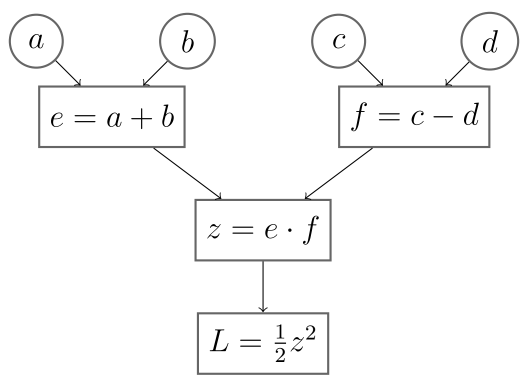
$$\frac{\partial L}{\partial b_1} = \delta_1$$

$$\frac{\partial L}{\partial b_1(2)} = 0.0293$$

$\square$

Exercise 7.2

Proof. a)



$\square$

b) Using the example values and doing a forward pass, we have:

$$\begin{aligned}
e &= a + b = 1 + 2 = 3 \\
f &= c - d = 3 - 4 = -1 \\
z &= e \cdot f = 3 \cdot (-1) = -3 \\
L &= \frac{1}{2}z^2 = \frac{1}{2}(-3)^2 = \frac{1}{2} \cdot 9 = 4.5
\end{aligned}$$

Now we compute the gradients of L w.r.t a, b, c, and d:

$$\frac{\partial L}{\partial z} = \frac{\partial}{\partial z}\left(\frac{1}{2}z^2\right) = z = -3$$

Using $z = e \cdot f$:

$$\frac{\partial z}{\partial e} = f = -1, \quad \frac{\partial z}{\partial f} = e = 3$$

Using $e = a + b$:

$$\frac{\partial e}{\partial a} = 1, \quad \frac{\partial e}{\partial b} = 1$$

Using $f = c - d$:

$$\frac{\partial f}{\partial c} = 1, \quad \frac{\partial f}{\partial d} = -1$$

The gradient of L w.r.t. a and b is:

$$\frac{\partial L}{\partial a} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial e} \cdot \frac{\partial e}{\partial a} = (-3) \cdot (-1) \cdot (1) = 3$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial e} \cdot \frac{\partial e}{\partial b} = (-3) \cdot (-1) \cdot (1) = 3$$

Similarly, the gradient of L w.r.t. c and d is:

$$\frac{\partial L}{\partial c} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial f} \cdot \frac{\partial f}{\partial c} = (-3) \cdot (3) \cdot (1) = -9$$

$$\frac{\partial L}{\partial d} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial f} \cdot \frac{\partial f}{\partial d} = (-3) \cdot (3) \cdot (-1) = 9$$

Exercise 7.3

Proof. a) WE have to use chain rul efor this problem. Compute $\frac{\partial L}{\partial x}$ for $F_1(x) = (x_2, x_4, x_3, x_0, x_1)$ The gradient $\frac{\partial L}{\partial x}$ is computed by mapping each component of $\frac{\partial L}{\partial F}$ back to the corresponding index in x:

$$\frac{\partial L}{\partial x_0} = d_3, \quad \frac{\partial L}{\partial x_1} = d_4, \quad \frac{\partial L}{\partial x_2} = d_0, \quad \frac{\partial L}{\partial x_3} = d_2, \quad \frac{\partial L}{\partial x_4} = d_1.$$

So, the result is:

$$\frac{\partial L}{\partial x} = (d_3, d_4, d_0, d_2, d_1).$$

b) Compute $\frac{\partial L}{\partial x}$ for $F_2(x) = x_0 \cdot x$
The function $F_2(x)$ computes the element-wise product $x_0 \cdot x = (x_0 \cdot x_0, x_0 \cdot x_1, x_0 \cdot x_2, x_0 \cdot x_3, x_0 \cdot x_4)$. and the upstream gradient is $\frac{\partial L}{\partial F} = (d_0, d_1, d_2, d_3, d_4)$.
For each component:

$$\frac{\partial L}{\partial x_0} = \sum_{i=0}^{4} d_i \cdot x_i,$$
$$\frac{\partial L}{\partial x_i} = d_i \cdot x_0, \quad \text{for } i = 1, 2, 3, 4.$$

This is because only in the first element which is x_0 we have multiplication. The result is:

$$\frac{\partial L}{\partial x} = \left(\sum_{i=0}^{4} d_i \cdot x_i,\ d_1 \cdot x_0,\ d_2 \cdot x_0,\ d_3 \cdot x_0,\ d_4 \cdot x_0 \right).$$

c) Compute $\frac{\partial L}{\partial x}$ for $F_3(x) = x_0 \cdot F_1(x)$

The function $F_3(x)$ computes the element-wise product $x_0 \cdot F_1(x)$, where $F_1(x) = (x_2, x_4, x_3, x_0, x_1)$. The upstream gradient is $\frac{\partial L}{\partial F} = (d_0, d_1, d_2, d_3, d_4)$.
For each component:

$$\frac{\partial L}{\partial x_0} = d_0 \cdot x_2 + d_1 \cdot x_4 + d_2 \cdot x_3 + d_3 \cdot x_0 + d_4 \cdot x_1,$$
$$\frac{\partial L}{\partial x_i} = d_j \cdot x_0, \quad \text{where } j \text{ is such that } F_1(x)_j = x_i.$$

Mapping $(x_2, x_4, x_3, x_0, x_1)$ back to indices:

$$j = \begin{cases} 0 & \text{if } i = 2, \\ 1 & \text{if } i = 4, \\ 2 & \text{if } i = 3, \\ 3 & \text{if } i = 0, \\ 4 & \text{if } i = 1. \end{cases}$$

$\square$