Submission for Deep Learning Exercise 2

Team: dl2024-sigma_learning
Students: Uralp Ergin (5975013), İsmail Karabaş (7654321), Ahmet Yaşar Ayfer (5986167)

November 6, 2024

1) Forward pass:

$$z_0 = W_0 X$$

$$h_0 = g_0(z_0)$$

$$z_1 = h_0 W_1$$

$$h_1 = g_1(z_1)$$

$$z_2 = h_1 W_2 + h_0 W_5$$

$$\hat{y} = g_2(z_2) = z_2$$

$$L(\hat{y}, y) = |y - \hat{y}|$$

Backward Pass:

$$\frac{\partial L}{\partial \hat{y}} = \begin{cases} -1 & , \hat{y} > y \\ +1 & , \hat{y} \leq y \end{cases}$$

$$\frac{\partial L}{\partial z_2} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z_2} = \frac{\partial L}{\partial \hat{y}} \cdot 1$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_2} = \boxed{\frac{\partial L}{\partial \hat{y}} \cdot 1 \cdot h_1}$$

$$\frac{\partial L}{\partial w_5} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_5} = \boxed{\frac{\partial L}{\partial \hat{y}} \cdot 1 \cdot h_0}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1} = \begin{cases} \frac{\partial L}{\partial \hat{y}} w_2 h_0 & , z_1 > 0 \\ 0 & , z_1 \leq 0 \end{cases}$$

$$\underbrace{\frac{\partial L}{\partial \hat{y}} \cdot 1}_{} \quad w_2 \quad \underbrace{\begin{cases} 1, & z_1 > 0 \\ 0, & z_1 \leq 0 \end{cases}}_{} \quad h_0$$

$$\frac{\partial L}{\partial z_1} = \begin{cases} \frac{\partial L}{\partial \hat{y}} w_2 & , z_1 > 0 \\ 0 & , z_1 \leq 0 \end{cases}$$

$$\frac{dL}{dw_0} = \frac{dL}{h_0} , \frac{dh_0^?}{dz_0} . \frac{dz_0}{dw_0} \searrow \times$$

$$\downarrow$$

$$\frac{dL}{dz_2} . \frac{dz_2}{dh_0} + \frac{dL}{dz_1} . \frac{dz_1}{dh_0} \longrightarrow \begin{cases} 1 , & z_0 > 0 \\ 0 , & z_0 \leq 0 \end{cases}$$

$$\underbrace{\qquad}_{\text{calculated}} \quad \downarrow_{w_5} \quad \underbrace{\qquad}_{\text{calculated}} \quad \downarrow_{w_1}$$

$$= \begin{cases} \left[\frac{dL}{dz_2} w_5 + \frac{dL}{dz_1} w_1\right] \times , & z_0 > 0 \\ 0 , & z_0 \leq 0 \end{cases}$$

2) Because of the skip connection, unit 0 receives two gradients when backpropagating. Therefore, when taking the gradient of h_0, we should also calculate the gradients come from unit 2 and unit 1.

3) $z_0 = x_1 \cdot w_0 = 1 \cdot 0.5 = 0.5$

$h_0 = ReLU(z_0) = 0.5$

$z_1 = h_0 \cdot w_1 = 0.5 \cdot 0.5 = 0.25$

$h_1 = ReLU(z_1) = 0.25$

$z_2 = h_1 \cdot w_2 + h_0 \cdot w_5 = 0.25 \cdot 0.5 + 0.5 \cdot 0.5$
$$= 0.125 + 0.250 = 0.375$$

$\hat{y} = g_2(z_2) = 0.375$

$L(\hat{y}, y) = |y - \hat{y}| = |-3 - 0.375| = 3.375$

Using the formulas we derive at Question 1 :

$$\frac{dL}{d\hat{y}} = -1 \qquad \frac{dL}{dz_2} = -1 \qquad \boxed{\frac{dL}{dw_2} = -1 \cdot 1 \cdot h_1 = -0.25}$$

$$\boxed{\frac{dL}{dw_s} = -1 \cdot 1 \cdot h_0 = -0.5}$$

$$\boxed{\frac{dL}{dw_1}} = \begin{cases} \frac{dL}{d\hat{y}} w_2 h_0 & , \; z_1 > 0 \\ 0 & , \; z_1 \leq 0 \end{cases} = \begin{cases} -0.5 \cdot \overset{0.25}{\overset{\wedge}{0.5}} & , \; z_1 > 0 \\ 0 & , \; z_1 \leq 0 \end{cases}$$

$$\boxed{= -0.25 \qquad \text{because} \quad z_1 = 0.25 \; > 0}$$

$$\frac{dL}{dz_1} = \begin{cases} \frac{dL}{d\hat{y}} w_2 & , \; z_1 > 0 \\ 0 & , \; z_1 \leq 0 \end{cases} = \begin{cases} -0.5 & , \; z_1 > 0 \\ 0 & , \; z_1 \leq 0 \end{cases}$$

$$= -0.5 \; \text{because} \; z_1 = 0.25 > 0$$

$$\boxed{\frac{dL}{dw_0}} = \begin{cases} \left[\frac{dL}{dz_2} w_s + \frac{dL}{dz_1} w_1 \right] x & , \; z_0 > 1 \\ 0 & , \; z_0 \leq 0 \end{cases}$$

$$= \begin{cases} \left[-0.5 + (-0.5) \cdot 0.5 \right] \cdot 1 & , \; z_0 > 0 \\ 0 & , \; z_0 \leq 0 \end{cases}$$

$$= \boxed{-0.75 \quad \text{because} \; z_0 = 0.5 \; > 0}$$

Updates and Loss :

$$w_2 = w_2 - \alpha \cdot \frac{dL}{dw_2} = 0.5 + 0.25 = 0.75$$

$$w_1 = 0.5 + 0.25 = \underline{0.75} \qquad w_s = 0.5 + 0.5 = \underline{1}$$

$$w_0 = 0.5 + 0.75 = \underline{1.25}$$

..... / ..... / ....

Loss : $w_0 = 1.25$ $w_1 = 0.75$ $w_2 = 0.75$
          $w_s = 1$

$z_0 = 1 \cdot (1.25) = 1.25$
$h_0 = ReLU(1.25) = 1.25$
$z_1 = h_0 \cdot w_1 = 1.25 \cdot 0.75 = 0.9375$
$h_1 = ReLU(0.9375) = 0.9375$

$z_2 = 0.9375 \cdot (0.75) + (1.25) \cdot 1 = 1.95$

$\hat{y} = g(z_2) = z_2 = 1.95$
$L(\hat{y}, y) = |y - \hat{y}| = |-3 - (1.95)| = \underline{\underline{4.95}}$