Subject: Deep learning HW4
Date

② (3) i_1 → h_1, (5) i_2 → h_2, h_1, h_2 → y'

$h_1 = w_1 i_1 + w_2 i_2 + b_1$

$h_2 = w_3 i_1 + w_4 i_2 + b_2$

$y' = \sigma(w_5 h_1 + w_6 h_2 + b)$

→ epoch : 2 , loss : mse , sigmoid , $y = 1$, Learning rate : 0.1

→ loss = $\frac{1}{n}\sum^{n}(y - y')^2$, $n=1$ → $(y-y')^2$

Initial weights : $\{w_1 : 0.1, w_2 = 0.5, w_3 = 0.2, w_4 = 0.4, w_5 = 0.1, w_6 = 0.3\}$

Initial Bias : $\{b_1 : 0.7, b_2 = 0.3, b = 0.4\}$

Epoch 1 : $h_1 = \underbrace{(0.1)(3)}_{0.3} + \underbrace{(0.5)(5)}_{2.5} + 0.7 = 3.5$

→ forward : $h_2 = \underbrace{(0.2)(3)}_{0.6} + \underbrace{(0.4)(5)}_{2} + 0.3 = 2.9$

→ $y' = \sigma(\underbrace{(0.1)(3.5)}_{0.35} + \underbrace{(0.3)(2.9)}_{0.87} + 0.4)$

$= \sigma(1.62) = 0.83$

→ Backward : $\frac{\partial L}{\partial y'} = -2(y - y') : -2(1 - 0.83) = -0.34$, $\frac{\partial \sigma(z)}{\partial z} = \sigma(z)(1 - \sigma(z)) = 0.14$

→ $(0.14)(-0.34) = -0.047$

$\frac{\partial y'}{\partial w_5} = h_1$, $\frac{\partial L}{\partial w_5} = \frac{\partial L}{\partial y'_{(in)}} \times \frac{\partial y'_{(in)}}{\partial w_5} = (-0.047)(3.5) = -0.16$

$\frac{\partial y'}{\partial w_6} = h_2$ → Same : $(-0.047)(2.9) = -0.13$

$\frac{\partial y'}{\partial h_1} = w_5$ → $\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial y'_{(in)}} \times \frac{\partial y'_{(in)}}{\partial h_1} \times \frac{\partial h_1}{\partial w_1}$ → $\frac{\partial L}{\partial w_1} = (-0.047)(w_5)(i_1) = -0.014$ (0.1, 3)

$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial y'_{(in)}} \times \frac{\partial y'_{(in)}}{\partial h_1} \times \frac{\partial h_1}{\partial w_2}$ → $\frac{\partial L}{\partial w_2} = (-0.047)(w_5)(i_2) = -0.023$ (0.1, 5)

$\frac{\partial y'}{\partial h_2} = w_6$ → $\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial y'_{(in)}} \times \frac{\partial y'_{(in)}}{\partial h_2} \times \frac{\partial h_2}{\partial w_3} = (-0.047)(w_6)(i_1) = -0.042$ (0.3, 3)

$\frac{\partial L}{\partial w_4} = \frac{\partial L}{\partial y'_{(in)}} \times \frac{\partial y'_{(in)}}{\partial h_2} \times \frac{\partial h_2}{\partial w_3} = (-0.047)(w_6)(i_2) = -0.07$ (0.3, 5)

P4PCO

Subject:
Date

$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial y'_{(in)}} \times \frac{\partial y'_{(in)}}{\partial b} = -0.047$ (1)

$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial y'_{(in)}} \times \frac{\partial y'_{(in)}}{\partial h_1} \times \frac{\partial h_1}{\partial b_1} = (-0.047)(w_5) = -0.004$ (0.1, 1)

$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial y'_{(in)}} \times \frac{\partial y'_{(in)}}{\partial h_2} \times \frac{\partial h_2}{\partial b_2} = (-0.047)(w_6) = -0.014$ (0.3)

→ update weights: $\{w_1 = 0.1 - (0.1)(-0.014) = 0.1014, w_2 = 0.5 - (0.1)(-0.023) = 0.5023$

$w_3 = 0.2 - (0.1)(-0.042) = 0.2042, w_4 = 0.4 - (0.1)(-0.07) = 0.407, w_5 = (0.1 - (0.1)(-0.16) = 0.116$

$w_6 = 0.3 - (0.1)(-0.13) = 0.313, b_1 = 0.70009, b_2 = 0.3014, b = 0.4047\}$

→ loss: $(1 - 0.83)^2 = 0.0289$

Epoch2:

→ forward: $\begin{cases} h_1 = (0.1014)(3) + (0.5023)(5) + 0.70009 = 3.5157 \\ h_2 = (0.2042)(3) + (0.407)(5) + 0.3014 = 2.949 \end{cases}$

$y' = \sigma(0.116(3.5157) + 0.313(2.949) + 0.4047) = 0.85$

$\frac{\partial L}{\partial y'} = -2(1 - 0.85) = -0.3$, $\frac{\partial \sigma(z)}{\partial z} = \sigma(z)(1 - \sigma(z)) = 0.85(0.15) = 0.127$

→ $(-0.3)(0.127) = -0.038$

$\frac{\partial L}{\partial w_6} = (-0.038)(2.949) = -0.112$

$\frac{\partial L}{\partial w_5} = (-0.038)(3.5157) = -0.1335$

$\frac{\partial L}{\partial w_1} = (-0.038)(w_5)(3) = -0.013$ $\frac{\partial L}{\partial w_2} = (-0.038)(w_5)(5) = -0.022$

Subject:
Date

$$\begin{cases} \frac{\partial L}{\partial w_3} = (-0.038)(w_6)(3) = -0.035 \\ \frac{\partial L}{\partial w_4} = (0.038)(w_6)(5) = -0.059 \end{cases} \qquad \begin{cases} \frac{\partial L}{\partial b} = -0.038 \\ \frac{\partial L}{\partial b_1} = (-0.038)(w_5) = -0.004 \\ \frac{\partial L}{\partial b_2} = (-0.038)(w_6) = -0.011 \end{cases}$$

→ update weights: $\{ w_1 = 0.1014 - (0.1)(-0.013) = 0.1027$, $w_2 = 0.5023 - (0.1)(-0.022) = 0.5045$

$w_3 = 0.2077$, $w_4 = 0.4129$, $w_5 = 0.1293$, $w_6 = 0.3242$, $b_1 = 0.70044$, $b_2 = 0.3025$,

$b = 0.4085 \}$ 2 epochs done ✓

→ loss: $(1 - 0.85)^2 = 0.0225$

→ predict:

$$\begin{cases} h_1 = 0.1027(3) + 0.5045(5) + 0.70044 = 3.531 \\ h_2 = 0.2077(3) + 0.4129(5) + 0.3025 = 2.99 \end{cases}$$

$y' = \sigma(0.1293(3.531) + 0.3242(2.99) + 0.4085) = \sigma(1.8344) = 0.862$

→ loss $= (1 - 0.862)^2 = 0.019$ loss ↓ , y' ↑

در تعداد 2 epoch آموزش، مقدار loss از 0.0289 به 0.019 رسیده است. این می‌توان فرض کرد

اگر تعداد epoch های بیشتری اجرا شود، هر چند این مقدار تغییر می‌کند و بهتر می‌شود اما با توجه به $0.0289 - 0.019 = 0.0099$

تغییرات بسیار کم تغییرات loss می‌توان گفت مدل همگرا است (جواب ها).