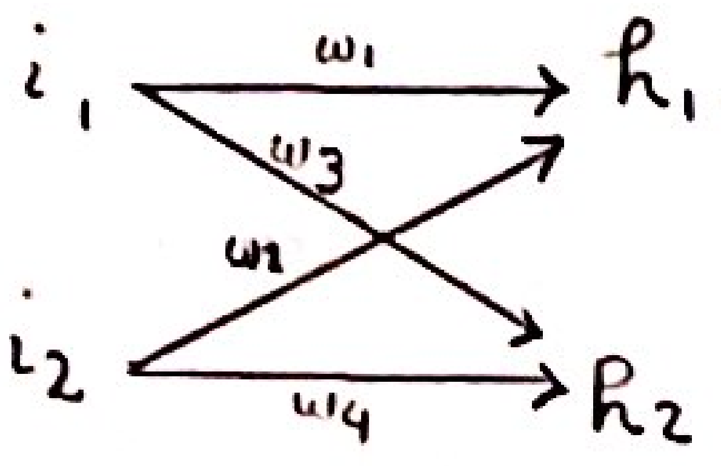
→ hw 7 : - 97521036 - سیکا محمدی کارگیر

- question 2 :



$$\begin{cases} h_1 = \mathrm{Relu}(i_1 w_1 + i_2 w_2 + b_1) \\ h_2 = \mathrm{Relu}(i_1 w_3 + i_2 w_4 + b_2) \\ y' = \mathrm{Relu}(h_1 w_5 + h_2 w_6 + b) \end{cases}$$

- loss : MSE
- epoch : 2
- learning rate : 0.1

L2 Regularization : $\mathrm{Loss} : \frac{1}{n}\sum_{i=1}^{n}(y-y')^2 + \lambda \sum_{1}^{n} w_i^2 \qquad n=2$

Initial weights : $\{ w_1 : 1 \ , w_2 : -1 \ , w_3 : -0.5 , w_4 : 0.5 , w_5 : +2 , w_6 : -2.5 , b_1 = b_2 = b = 0.5 \}$

- Epoch 1 :

- forward : ①
$$\begin{cases} h_1 = R(3(1) + 2(-1) + 0.5) = R(1.5) = 1.5 \\ h_2 = R(\underbrace{3(-0.5)}_{-1.5} + \underbrace{2(0.5)}_{1} + 0.5) = R(0) = 0 \end{cases} \rightarrow y' = R(1.5(+2) + 0(-2.5) + 0.5) = 3.5$$

Consider L_2 : $\nabla_w \tilde{J}(\theta; x,y) = \alpha w + \nabla_w J(\theta; x,y)$

for w [1-4] we have $\partial y' / \partial h$ too.

- Backward : Chain Rule : $\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y'} \cdot \frac{\partial y'}{\partial Z} \cdot \frac{\partial Z}{\partial R} \cdot \frac{\partial R}{\partial w}$, Z , R : Relu $\rightarrow \begin{cases} 0 \\ 1 \end{cases}$ Two for w_1, w_2, w_3, w_4 But for w_5, w_6 we only need one Relu .

②
$$\begin{cases} h_1 = R((15)(1) + 12(-1) + 0.5) = 3.5 \\ h_2 = R(15(-0.5) + 12(0.5) + 0.5) = 0 \end{cases} \rightarrow y' = (3.5(+2) + 0(-2.5) + 0.5) = 7.5$$

$$\rightarrow \begin{cases} ① \ \frac{\partial L}{\partial w_5} = -[\underbrace{(8-\cancel{3.5})}_{4.5} + \underbrace{(20-\cancel{7.5})}_{12.5}](1)(\overset{1.5}{\cancel{h_1}}) = -25.5 \\ \frac{\partial L}{\partial w_5} = -[17](1)(3.5) = -59.5 \end{cases} \quad ⊕ = -85$$

$$\rightarrow \begin{cases} \frac{\partial L}{\partial w_6} = -[17](1) h_2 = 0 \\ \frac{\partial L}{\partial w_6} = -[17](1) h_2 = 0 \end{cases} \quad ⊕ = 0$$

$$\rightarrow \begin{cases} \frac{\partial L}{\partial w_1} = -[17](1)(1)\overset{6}{\cancel{(w_5)(i_1)}} = -102 \\ \frac{\partial L}{\partial w_1} = -[17](1)(1)\underbrace{\cancel{(w_5)(i_1)}}_{2 \times 15} = -510 \end{cases} \quad ⊕ = -612$$

$$\rightarrow \begin{cases} \frac{\partial L}{\partial w_2} = (-17)(1)(1)\overset{4}{\cancel{(w_5)(i_2)}} = -68 \\ \frac{\partial L}{\partial w_2} = (-17)(1)(1)\underbrace{\cancel{(w_5)(i_2)}}_{12} = -408 \end{cases} \quad ⊕ = -476$$

$$\rightarrow \begin{cases} \frac{\partial L}{\partial w_3} = (-17)(1)(0)(w_6)(i_1) = 0 \\ \frac{\partial L}{\partial w_3} = (-17)(1)(0)(w_6)(i_1) = 0 \end{cases}$$ Same for $\frac{\partial L}{\partial w_4} = 0$

→ $\lambda = 0.1$ ✓ , $\begin{cases} \frac{\partial L}{\partial b} = -17 \times 2 \\ \frac{\partial L}{\partial b_1} = (-17)(1)(1)(2) = -34 \times 2 \\ \frac{\partial L}{\partial b_2} = \text{صفر } 0 \end{cases}$, gradient : $\lambda w + \frac{\partial L}{\partial w}$

$\mathrm{Loss} = \frac{1}{2}[(12.5)^2 + (4.5)^2] = 88.25$

→ **Adam :**

$$\begin{cases} fm = \overset{1-\mathrm{Beta1}}{(0.1)}(0.1(\cancel{w_1}) + (-612)) = -61.19 \rightarrow -611.9 \\ Sm = \underset{1-\mathrm{Beta2}}{(0.001)}(-611.9)^2 = 374.42161 \rightarrow 374421.61 \end{cases}$$

$w_1 = 1 + \left(\frac{611.9}{611.9000001}\right)(0.1) = 1.099$

$$\begin{cases} f_m = (0.1)((0.1)(-1) + (-475)) = -47.61 \longrightarrow -476.1 \\ S_m = (0.001)(226671.21) = 226.67121 \longrightarrow 226671.21 \end{cases}$$

$$w_2 = -1 + \left(\frac{476.1}{476.100\ldots}\right)(0.1) = -0.901$$

$$\begin{cases} f_m = (0.1)((0.1)(-0.5)) = -0.005 \longrightarrow -0.05 \\ S_m = (0.001)(25 \times 10^{-4}) = 25 \times 10^{-7} \longrightarrow 25 \times 10^{-4} \end{cases}$$

$$w_3 = -0.5 + \left(\frac{0.05}{0.0500001}\right)(0.1) = -0.401$$

$$\begin{cases} f_m = (0.1)((0.1)(0.5)) = 0.005 \longrightarrow 0.05 \\ S_m = (0.001)(25 \times 10^{-4}) = 25 \times 10^{-7} \longrightarrow 25 \times 10^{-9} \end{cases}$$

$$w_4 = 0.5 - \left(\frac{0.05}{0.0500001}\right)(0.1) = 0.401$$

$$\begin{cases} f_m = (0.1)((0.1)(2) + (-85)) = -8.48 \longrightarrow -84.8 \\ S_m = (0.001)(-84.8)^2 = 7.19104 \longrightarrow 7191.04 \end{cases}$$

$$w_5 = 2 + \left(\frac{84.8}{84.8000001}\right)(0.1) = 2.099$$

learning rate

$$\begin{cases} f_m = (0.1)((0.1)(-2.5) + 0) = -0.025 \longrightarrow -0.25 \\ S_m = (0.001)(625 \times 10^{-4}) = 625 \times 10^{-7} \longrightarrow 625 \times 10^{-4} \end{cases}$$

$$w_6 = -2.5 + \left(\frac{0.25}{0.2500001}\right)(0.1) = -2.401$$

→ $t = 1$ in this iteration, all formulas are from slides.

- **Epoch 2:** $w = \{1.099, -0.901, -0.401, 0.401, 2.099, -2.401\}$

$B = \{0.599, 0.599, 0.401\}$

- forward:

①
$$\begin{cases} h_1 = (3)(1.099) + (2)(-0.901) + 0.599 = 2.094 \\ h_2 = (3)(-0.401) + (2)(0.401) + 0.599 = 0.198 \end{cases} \rightarrow y' = (2.094)(2.099) + (0.198)(-2.401) + 0.401 = 4.32$$

↗ like this

②
$$\begin{cases} h_1 = 6.271 \\ h_2 = 0 \end{cases} \rightarrow y' = 13.563, \quad \text{loss} = \frac{1}{2}\left[(6.437)^2 + (3.68)^2\right] = 27.488$$

All gradients should be calculated like the previous epoch.

$$\begin{cases} \frac{\partial L}{\partial w_5} = -\left[\underset{3.68}{(8 - \cancel{4.32})} + \underset{6.437}{(20 - \cancel{13.563})}\right](1)(\cancel{h_1}) \nearrow 2.094 = -21.184 \\ \frac{\partial L}{\partial w_5} = -(10.117)(1)(6.271) = -63.443 \end{cases} \oplus = -84.627$$

$$\begin{cases} \frac{\partial L}{\partial w_6} = -(10.117)(1)(h_2) = -2.003 \\ \frac{\partial L}{\partial w_6} = -(10.117)(1)(0) = 0 \end{cases}$$

$$\begin{cases} \frac{\partial L}{\partial w_1} = (-10.117)(1)(w_5)(1)(i_1) = -63.706 \\ \frac{\partial L}{\partial w_1} = (-10.117)(1)(1)(w_5)(i_1) = -318.533 \end{cases} \oplus = -382.239$$

$$\begin{cases} \frac{\partial L}{\partial w_2} = (-10.117)(1)(1)(2.099)(2) = -42.471 \\ \frac{\partial L}{\partial w_2} = (-10.117)(1)(1)(2.099)(12) = -254.826 \end{cases} \oplus = -297.297$$

$$\begin{cases} \dfrac{\partial L}{\partial w_3} = (-10.117)(1)(1)(w_6)(3) = +72.872 \\ \dfrac{\partial L}{\partial w_3} = (-10.117)(1)(0)(w_6)(15) = 0 \end{cases} \xrightarrow{\text{Same}} \begin{cases} \dfrac{\partial L}{\partial w_4} = (-10.117)(1)(1)(-2.401)(2) = 48.581 \\ \dfrac{\partial L}{\partial w_4} = 0 \end{cases}$$

$$\begin{cases} \dfrac{\partial L}{\partial b_1} = -10.117 \times 2(\text{input data}) = -20.234 \\ \dfrac{\partial L}{\partial b_1} = (-10.117)(1)(1)(2.099) \times 2 = -42.471 \\ \dfrac{\partial L}{\partial b_2} = (-10.117)(1)(1)(-2.401) = 24.29 \end{cases}$$

→ **Adam:** $t = 2$

w_1

$$\begin{cases} f_m = 0.9 \times (-61.19) + (0.1)((0.1)(1.099) + (-382.239)) = -93.283 \xrightarrow{} f_m/(1-0.81) = -490.963 \\ S_m = (0.999) \times \underbrace{374.421}_{\text{previous}} + (0.001)(-382.129)^2 = 374.046 + 146.022 = 520.068 \xrightarrow{} 260039 \end{cases}$$

$$w_1 = 1.099 - 0.1\left(\frac{-490.963}{509.935}\right) = \boxed{1.195}$$

w_2

$$\begin{cases} f_m = 0.9 \times (-47.61) + (0.1)((0.1)(-0.901) + (-297.297)) = -72.587 \xrightarrow{} = -382.036 \\ S_m = 0.999 \times 226.671 + (0.001)(-297.297)^2 = 314.829 \xrightarrow{} 157414.5 \end{cases}$$

$$w_2 = -0.901 - 0.1\left(\frac{-382.036}{396.754}\right) = \boxed{-0.805}$$

w_3

$$\begin{cases} f_m = 7.238 \xrightarrow{} 38.094 \\ S_m = 5.306 \xrightarrow{} 2653 \end{cases}, \quad w_3 = -0.401 - \left(\frac{7.238}{51.507}\right)(0.1) = \boxed{-0.415}$$

w_4

$$\begin{cases} f_m = 4.817 \xrightarrow{} 25.35 \\ S_m = 2.366 \xrightarrow{} 34.394 \end{cases}, \quad w_4 = 0.401 - 0.073 = 0.328$$

w_5

$$\begin{cases} f_m = -16.073 \xrightarrow{} -84.594 \\ S_m = 14.31 \xrightarrow{} 7155 \end{cases}, \quad w_5 = 2.099 + 0.1 = \boxed{2.199}$$

w_6

$$\begin{cases} f_m = -0.246 \xrightarrow{} -1.294 \\ S_m = 0.005 \xrightarrow{} 2.5 \end{cases}, \quad w_6 = -2.401 + \frac{1.294}{1.581} = \boxed{-2.32}$$

Final weights after 2 epochs ✓

Bias

$$\begin{cases} b \Rightarrow 0.599 + \left(\frac{5.42}{27.83}\right)(0.1) = 0.618 \\ b_1 \Rightarrow 0.599 + \left(\frac{54.47}{56.612}\right)(0.1) = 0.695 \\ b_2 \Rightarrow 0.401 \longrightarrow 0.327 \end{cases}$$

→ **Final Result:**

$$\begin{cases} h_1 = (1.195)(3) + (-0.805)(2) + 0.695 = 2.67 \\ h_2 = (-0.415)(3) + (0.328)(2) + 0.327 = 0 \end{cases} \rightarrow y' = 6.489$$

← از نظر همگرایی نمی‌توان به طور قطع نظر داد چراکه تعداد iteration ها بسیار کم است ولی همانطور که محاسبه شد loss رو به کاهش است و پس از این بار هم احتمالاً کاهش می‌یابد (بنابراین تا حدی می‌توان گفت همگرا است)

تغییرات lossها در epoch اول به دوم بسیار واضح است.

$$\rightarrow \begin{cases} h_1 = (1.195)(15) + (-0.805)(12) + \quad = 8.96 \\ h_2 = (-0.415)(15) + (0.328)(12) + \quad = 0 \end{cases} \rightarrow y' = 20.321$$

$$loss = \frac{1}{2}\left(\underset{0.103}{(0.321)^2} + \underset{2.283}{(1.511)^2} \right) = \underline{1.193} \rightarrow loss$$

به شدت کاهش یافته است.