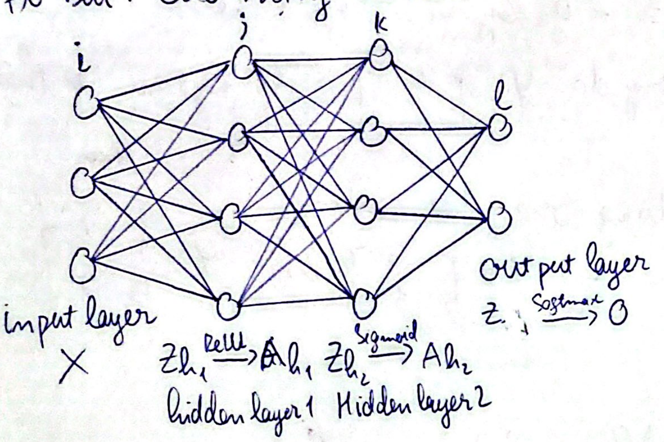
Nguyễn Việt Hoàng, 21129079, STT: 10

Lớp sáng thứ 5, tiết 4 5 6

Đề bài: Cho mạng nơ-ron



Bias tại các lớp:

$$Bh_1 = \begin{bmatrix} bh_{10} \\ bh_{11} \\ bh_{12} \\ bh_{13} \end{bmatrix}_{4\times1} \quad Bh_2 = \begin{bmatrix} bh_{20} \\ bh_{21} \\ bh_{22} \\ bh_{23} \end{bmatrix}_{4\times1} \quad B_O = \begin{bmatrix} bo_0 \\ bo_1 \end{bmatrix}_{2\times1}$$

~~Kí~~ ~~thì~~ kí hiệu của weight:

$Wh_1 = W_{ji}$

$Wh_2 = W_{kj}$

$W_O = W_{lk}$

1. Tính toán lan truyền thẳng

⊕ Tại hidden layer 1:

$$Zh_1 = Wh_1 \cdot X + Bh_1 = \begin{array}{c} \begin{array}{ccc} (i=0) & (i=1) & (i=2) \end{array} \\ \begin{bmatrix} wh_{1(00)} & wh_{1(01)} & wh_{1(02)} \\ wh_{1(10)} & \ddots & \vdots \\ \vdots & & \\ wh_{1(30)} & \cdots & wh_{1(32)} \end{bmatrix}_{4\times3} \end{array} * \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}_{3\times1} + \begin{bmatrix} bh_{10} \\ bh_{11} \\ bh_{12} \\ bh_{13} \end{bmatrix}_{4\times1} = \begin{bmatrix} zh_{10} \\ zh_{20} \\ zh_{30} \\ zh_{40} \end{bmatrix}_{4\times1}$$

~~Tại hidden layer 2.~~ $Ah_1 = ReLU(Zh_1) = \begin{bmatrix} ah_{10} \\ ah_{11} \\ ah_{12} \\ ah_{13} \end{bmatrix}_{4\times1}$

⊕ Tại hidden layer 2:

$$Zh_2 = Wh_2 \cdot Ah_1 + Bh_2 = \begin{array}{c} \begin{array}{cc} (j=0) & (j=3) \end{array} \\ \begin{bmatrix} wh_{2(00)} & \cdots & wh_{2(03)} \\ \vdots & \ddots & \vdots \\ wh_{2(30)} & \cdots & wh_{2(33)} \end{bmatrix}_{4\times4} \end{array} * \begin{bmatrix} ah_{10} \\ ah_{11} \\ ah_{12} \\ ah_{13} \end{bmatrix}_{4\times1} + \begin{bmatrix} bh_{20} \\ bh_{21} \\ bh_{22} \\ bh_{23} \end{bmatrix}_{4\times1} = \begin{bmatrix} zh_{20} \\ zh_{21} \\ zh_{22} \\ zh_{23} \end{bmatrix}_{4\times1}$$

$$Ah_2 = Sigmoid(Zh_2) = \begin{bmatrix} ah_{20} \\ ah_{21} \\ ah_{22} \\ ah_{23} \end{bmatrix}_{4\times1}$$

⊕ Tại lớp ngõ ra

$$Z = W_O \cdot Ah_2 + B_O = \begin{array}{c} \begin{array}{cccc} (k=0) & (k=1) & (k=2) & (k=3) \end{array} \\ \begin{bmatrix} w_{O(0,0)} & w_{O(0,1)} & w_{O(0,2)} & w_{O(0,3)} \\ w_{O(1,0)} & w_{O(1,1)} & w_{O(1,2)} & w_{O(1,3)} \end{bmatrix}_{2\times4} \end{array} * \begin{bmatrix} ah_{20} \\ ah_{21} \\ ah_{22} \\ ah_{23} \end{bmatrix}_{4\times1} + \begin{bmatrix} bo_0 \\ bo_1 \end{bmatrix}_{2\times1} = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix}_{2\times1}$$

$$O = Softmax(Z) = \begin{bmatrix} o_0 \\ o_1 \end{bmatrix}_{2\times1}$$

2. Tính toán lan truyền ngược sử dụng Gradient Descent

⊕ Tại lớp ngõ ra

$$w_{0t} = w_{0(t-1)} - \eta \cdot \frac{\partial L}{\partial W_0}$$

⊕ Tại lớp ẩn thứ 2

$$Wh_{2t} = Wh_{2(t-1)} - \eta \frac{\partial L}{\partial Wh_2}$$

⊕ Tại lớp ẩn thứ 1

$$Wh_{1t} = Wh_{1(t-1)} - \eta \frac{\partial L}{\partial Wh_1}$$

⊕ Xét hàm mất mát cross-entropy loss

$$L = -y_0 \log(O_0) - y_1 \log(O_1)$$

- trong đó $Y = \begin{bmatrix} y_0 \\ y_1 \end{bmatrix}_{2\times 1}$ là ngõ ra sự thật

sử dụng one-hot coding

$\Rightarrow Y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ hoặc $Y = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

- Xét $\frac{\partial L}{\partial W_0} \cancel{= \frac{\partial L}{\partial Z_0} \frac{\partial L}{\partial Z} \cdot \frac{\partial Z}{\partial W_0} = \frac{\partial L}{\partial O} \cdot \frac{\partial O}{\partial Z} \cdot \frac{\partial Z}{\partial W_0}}$

$$= \begin{bmatrix} \frac{\partial L}{\partial W_{0(00)}} & \cdots & \frac{\partial L}{\partial W_{0(03)}} \\ \frac{\partial L}{\partial W_{0(10)}} & \cdots & \frac{\partial L}{\partial W_{0(13)}} \end{bmatrix}_{2\times 4} = \begin{bmatrix} \frac{\partial L}{\partial z_0} \frac{\partial z_0}{\partial W_{0(00)}} & \cdots & \frac{\partial L}{\partial z_0} \cdot \frac{\partial z_0}{\partial W_{0(03)}} \\ \frac{\partial L}{\partial z_1} \frac{\partial z_1}{\partial W_{0(10)}} & \cdots & \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial W_{0(13)}} \end{bmatrix}_{2\times 4}$$

$$= \begin{bmatrix} \frac{\partial L}{\partial z_0} \cdot ah_{10} & \cdots & \frac{\partial L}{\partial z_0} \cdot ah_{13} \\ \frac{\partial L}{\partial z_1} \cdot ah_{10} & \cdots & \frac{\partial L}{\partial z_1} \cdot ah_{13} \end{bmatrix}_{2\times 4} = \begin{bmatrix} \frac{\partial L}{\partial z_0} \\ \frac{\partial L}{\partial z_1} \end{bmatrix}_{2\times 1} * \begin{bmatrix} ah_{20} & ah_{21} & ah_{22} & ah_{23} \end{bmatrix}_{1\times 4}$$

$$= \begin{bmatrix} \frac{\partial L}{\partial z_0} \\ \frac{\partial L}{\partial z_1} \end{bmatrix}_{2\times 1} * \cdot Ah_2^T$$

- Tương tự với lớp ẩn thứ 2 và lớp ẩn thứ 1

$$\frac{\partial L}{\partial Wh_2} = \begin{bmatrix} \frac{\partial L}{\partial zh_{20}} \\ \frac{\partial L}{\partial zh_{21}} \\ \frac{\partial L}{\partial zh_{22}} \\ \frac{\partial L}{\partial zh_{23}} \end{bmatrix}_{4\times 1} * Ah_1^T \qquad \frac{\partial L}{\partial Wh_1} = \begin{bmatrix} \frac{\partial L}{\partial zh_{10}} \\ \frac{\partial L}{\partial zh_{11}} \\ \frac{\partial L}{\partial zh_{12}} \\ \frac{\partial L}{\partial zh_{13}} \end{bmatrix}_{4\times 1} * X^T$$

⊕ xét $\frac{\partial L}{\partial z_0} = \frac{\partial}{\partial z_0}\left[-\sum_{i=0}^{2}(y_i \cdot \log O_i)\right]$

$$= -\sum_{i=0}^{2}\frac{\partial}{\partial z_0}(y_i \log O_i)$$

$$= -\sum_{i=0}^{2}\left(y_i \frac{\partial \log O_i}{\partial O_i}\cdot\frac{\partial O_i}{\partial z_0}\right)$$

$$= -\sum_{i=0}^{2}\left(y_i \cdot \frac{1}{O_i}\cdot\frac{\partial O_i}{\partial z_0}\right)$$

TH_1: $i = 0$, ta có

$$\frac{\partial O_i}{\partial z_0} = \left(\frac{e^{z_0}}{e^{z_0}+e^{z_1}}\right)' = \frac{e^{z_0}(e^{z_0}+e^{z_1}) - e^{z_0}\cdot e^{z_0}}{(e^{z_0}+e^{z_1})^2}$$

$$= \frac{e^{z_0}}{e^{z_0}+e^{z_1}}\cdot\frac{e^{z_1}}{e^{z_0}+e^{z_1}}$$

$$= O_0 . O_1$$

TH_2: $i = 1$, ta có

$$\frac{\partial O_i}{\partial z_0} = \left(\frac{e^{z_1}}{e^{z_0}+e^{z_1}}\right)' = \frac{-e^{z_0}\cdot e^{z_1}}{(e^{z_0}+e^{z_1})^2}$$

$$= -O_0 O_1$$

Từ TH_1 và TH_2, suy ra:

$$\frac{\partial L}{\partial z_0} = -\left[y_0 \frac{1}{O_0}\cdot O_0 O_1 + y_1 \frac{1}{O_1}\cdot(-O_0 O_1)\right]$$

$$= -(y_0 O_1 - y_1 O_0)$$

$$= -[y_0(1-O_0) - y_1 O_0]$$

$$= -[y_0 - O_0(y_0 + y_1)]$$

$$= O_0 - y_0$$ (Tổng các thành phần của véc-tơ one-hot bằng 1)

Tương tự với $\frac{\partial L}{\partial z_1}$, ta có $\frac{\partial L}{\partial z_1} = O_1 - y_1$

$$\Rightarrow \frac{\partial L}{\partial z} = \begin{bmatrix}\frac{\partial L}{\partial z_0}\\ \frac{\partial L}{\partial z_1}\end{bmatrix}_{2\times 1} = \begin{bmatrix}O_0 - y_0\\ O_1 - y_1\end{bmatrix}_{2\times 1} = O - Y = e$$

⊕ xét $\frac{\partial L}{\partial zh_{20}} = \frac{\partial}{\partial zh_{20}}\left[-\sum_{i=0}^{2}(y_i \cdot \log O_i)\right]$

$$= -\sum_{j=0}^{2}\sum_{i=0}^{2}\left(y_i \frac{\partial \log O_i}{\partial O_i}\cdot\frac{\partial O_i}{\partial z_j}\cdot\frac{\partial z_j}{\partial ah_{20}}\cdot\frac{\partial ah_{20}}{\partial zh_{20}}\right)$$

$$= -\sum_{j=0}^{2}\sum_{i=0}^{2}\left(y_i \cdot\frac{1}{O_i}\cdot\frac{\partial O_i}{\partial z_j}\cdot Wh_2(j,0)\cdot ah_{20}(1-ah_{20})\right)$$

$$\left(\text{Do } \frac{\partial ah_{20}}{\partial zh_{20}} = \left(\frac{1}{1+e^{-zh_{20}}}\right)' = \frac{e^{-zh_{20}}}{(1+e^{-zh_{20}})^2} = \frac{\frac{1}{ah_{20}} - 1}{\left(\frac{1}{ah_{20}}\right)^2} = ah_{20}(1-ah_{20})\right)$$

$$= ah_{2_0}(1-ah_{2_0}) \cdot \left[-\sum_{j=0}^{2} \sum_{i=0}^{2} \left(y_i \frac{1}{O_i} \frac{\partial O_i}{\partial z_j} \cdot wh_{2(j;0)} \right) \right]$$

$$= -ah_{2_0}(1-ah_{2_0}) \left[wh_{2(00)} \left(y_0 \frac{1}{O_0} O_0 O_1 - y_1 \frac{1}{O_1} O_0 O_1 \right) + wh_{2(10)} \left(-y_1 \frac{1}{O_1} O_0 O_1 + y_0 \frac{1}{O_0} O_0 O_1 \right) \right]$$

$$= ah_{2_0}(1-ah_{2_0}) \left[wh_{2(00)} (O_0 - y_0) + wh_{2(10)} (O_1 - y_1) \right]$$

$$= \begin{bmatrix} wh_{2(00)} & wh_{2(10)} \end{bmatrix}_{1\times 2} * \begin{bmatrix} O_0 - y_0 \\ O_1 - y_1 \end{bmatrix}_{2\times 1} \cdot ah_{2_0}(1-ah_{2_0})$$

Tương tự với $\frac{\partial L}{\partial zh_{2_1}}$, $\frac{\partial L}{\partial zh_{2_2}}$ và $\frac{\partial L}{\partial zh_{2_3}}$, ta có

$$\begin{bmatrix} \frac{\partial L}{\partial zh_{2_0}} \\ \frac{\partial L}{\partial zh_{2_1}} \\ \frac{\partial L}{\partial zh_{2_2}} \\ \frac{\partial L}{\partial zh_{2_3}} \end{bmatrix} = \begin{bmatrix} wh_{2(00)} & wh_{2(10)} \\ wh_{2(01)} & wh_{2(11)} \\ wh_{2(02)} & wh_{2(12)} \\ wh_{2(03)} & wh_{2(13)} \end{bmatrix}_{4\times 2} * e_{2\times 1} \cdot \begin{bmatrix} ah_{2_0}(1-ah_{2_0}) \\ ah_{2_1}(1-ah_{2_1}) \\ ah_{2_2}(1-ah_{2_2}) \\ ah_{2_3}(1-ah_{2_3}) \end{bmatrix}$$

$$= (W_0^T * e) \cdot Ah_2(1-Ah_2) = eh_2 = \frac{\partial L}{\partial zh_2}$$

~~⊕ Xét~~ $\cancel{\frac{\partial L}{\partial zh_{1_0}} = \frac{\partial}{\partial zh_{1_0}} \left[-\sum_{i=0}^{2} (y_i \log o_i) \right]}$

$\cancel{= -\sum}$

⊕ Xét $\frac{\partial L}{\partial zh_1} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial Ah_2} \cdot \frac{\partial Ah_2}{\partial zh_2} \cdot \frac{\partial zh_2}{\partial Ah_1} \cdot \frac{\partial Ah_1}{\partial zh_1}$.

và $\frac{\partial L}{\partial zh_2} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial Ah_2} \cdot \frac{\partial Ah_2}{\partial zh_2}$, ta thấy:

$$\frac{\partial L}{\partial zh_{(L)}} = \cancel{\frac{\partial L}{\partial z}} \cancel{eh_L} (wh_{(L+1)}^T * eh_{(L+1)}) \cdot Ah_L'(zh_L)$$ (với L là ~~số~~ lớp ẩn tương ứng)

$$\Rightarrow \frac{\partial L}{\partial zh_1} = (wh_2^T * eh_2) \cdot \begin{bmatrix} 1 \ (zh_{1_i} > 0) \\ 0 \ (zh_{1_i} \leq 0) \end{bmatrix} \quad (i = [0;1;2;3])$$

Như vậy, công thức cập nhật trọng số tại các lớp là:

$$e = O - Y$$

$$eh_2 = (W_o^T * e) . Ah_2(1 - Ah_2)$$

$$eh_1 = (Wh_2^T * eh_2) . \begin{cases} 1 \ (zh_{1i} > 0) \\ 0 \ (zh_{1i} \le 0) \end{cases} \quad (i = [0;1;2;3])$$

$$wh_{1t} = wh_{1(t-1)} - \eta . eh_1 * X^T$$

$$wh_{2t} = wh_{2(t-1)} - \eta . eh_2 * Ah_1^T$$

$$wo_t = wo_{(t-1)} - \eta . e * Ah_2^T$$