Họ tên: Nguyễn Hữu Tài.

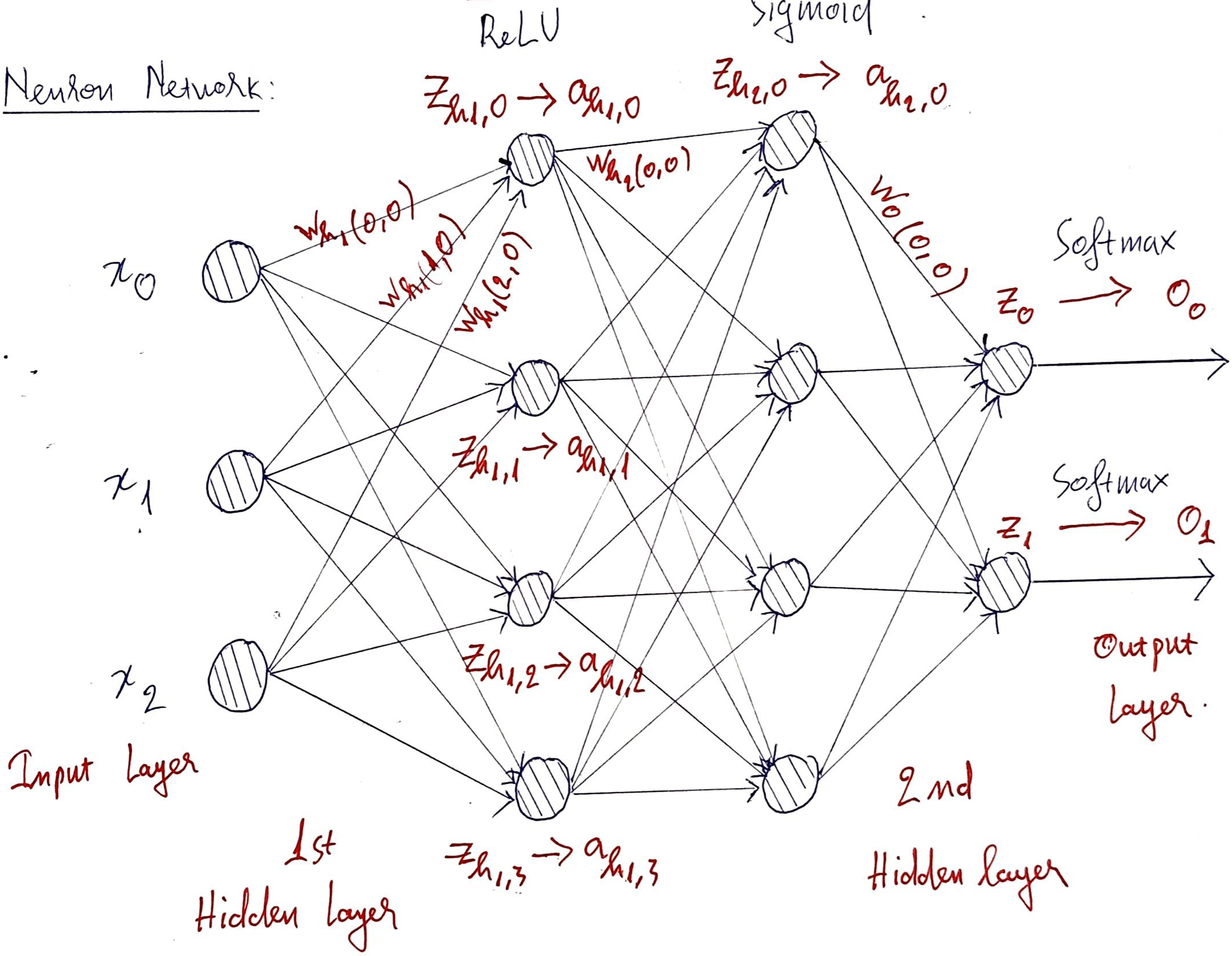
MSSV: 2119375

Lớp thứ 5, tiết 4,5,6.

Số thứ tự: 25

Home WORK

Neuron Network:



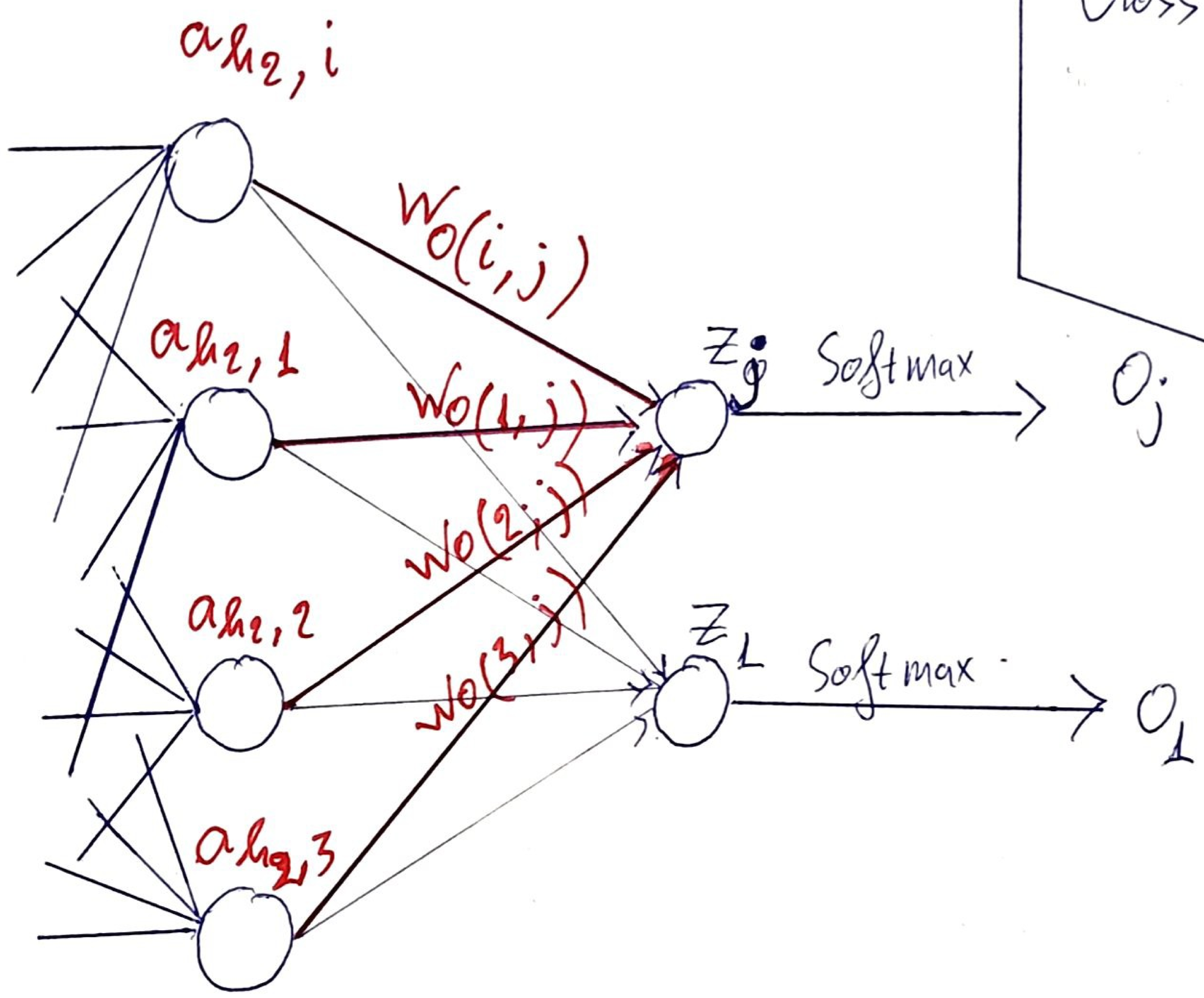
Bias tại 1 st hidden layer: $b_{h_1,0}$, $b_{h_1,1}$, $b_{h_1,2}$, $b_{h_1,3}$

Bias tại 2 nd hidden layer: $b_{h_2,0}$, $b_{h_2,1}$, $b_{h_2,2}$, $b_{h_2,3}$

Bias tại Output layer: b_{z_0}, b_{z_1}

*) Cập nhật trọng số $W_{o(i,j)}$

(Trọng số giữa node i của 2nd Hidden layer và node j của output layer).



Cross Entropy:

$$L = -\sum_{k=0}^{1} y_k \log(o_k)$$

Công thức Gradient Decent:

$$\boxed{W_{o(i,j)} = W_{o(i,j)} - \eta \frac{\partial L}{\partial W_{o(i,j)}}}$$

Chain rule:

$$\frac{\partial L}{\partial W_{o(i,j)}} = \frac{\partial L}{\partial z_j} \cdot \frac{\partial z_j}{\partial W_{o(i,j)}}$$

$\boxed{* \text{ Tìm } \dfrac{\partial L}{\partial z_j}}$

$$\frac{\partial L}{\partial z_j} = \partial\left(-\sum_{k=0}^{1} y_k \cdot \log(o_k)\right) \Big/ \partial z_j$$

$$= -\sum_{k=0}^{1} y_k \cdot \frac{\partial \log(o_k)}{\partial z_j}$$

Ta có: $$\frac{\partial \log(o_k)}{\partial z_j} = \frac{\partial \log(o_k)}{\partial o_k} \cdot \frac{\partial o_k}{\partial z_j} = \frac{1}{o_k} \cdot \frac{\partial o_k}{\partial z_j}$$

Ta có: $$\frac{\partial o_k}{\partial z_j} = \partial\left(\frac{e^{z_k}}{\sum_{l=0}^{1} e^{z_l}}\right) \Big/ \partial z_j$$

* Trường hợp 1: $k = j$

$$\partial\left(\frac{e^{z_k}}{\sum_{l=0}^{1} e^{z_l}}\right) \Big/ \partial z_j = \partial\left(\frac{e^{z_j}}{\sum_{l=0}^{1} e^{z_l}}\right) \Big/ \partial z_j$$

Đạo hàm: $$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

$$\Rightarrow \left(\frac{e^{z_j}}{\sum_{l=0}^{1} e^{z_l}}\right)' = \frac{e^{z_j} \cdot \left(\sum_{l=0}^{1} e^{z_l}\right) - e^{z_j} \cdot \partial\left(\sum_{l=0}^{1} e^{z_l}\right) \Big/ \partial e^{z_j}}{\left(\sum_{l=0}^{1} e^{z_l}\right)^2}$$

$$= \frac{e^{z_j}\left(\sum_{l=0}^{1} e^{z_l}\right) - e^{z_j} \cdot e^{z_j}}{\left(\sum_{l=0}^{1} e^{z_l}\right)^2} = \frac{e^{z_j}\left[\left(\sum_{l=0}^{1} e^{z_l}\right) - e^{z_j}\right]}{\left(\sum_{l=0}^{1} e^{z_l}\right)^2}.$$

$$= \frac{e^{z_j}}{\sum_{l=0}^{1} e^{z_l}} \cdot \left(1 - \frac{e^{z_j}}{\sum_{l=0}^{1} e^{z_l}}\right) = o_j\,(1 - o_j).$$

Vậy khi $k = j$ thì $\partial\left(\frac{e^{z_k}}{\sum_{l=0}^{1} e^{z_l}}\right) / \partial z_j = o_j\,(1 - o_j)$

* <u>Trường hợp 2:</u> $k \neq j$

$$\partial\left(\frac{e^{z_k}}{\sum_{l=0}^{1} e^{z_l}}\right) / \partial z_j = \frac{0.\left(\sum_{l=0}^{1} e^{z_l}\right) - e^{z_k}.\left(\sum_{l=0}^{1} e^{z_l}\right)'}{\left(\sum_{l=0}^{1} e^{z_l}\right)^2}$$

$$= \frac{-e^{z_k} \cdot e^{z_j}}{\left(\sum_{l=0}^{1} e^{z_l}\right)^2} = -o_k \cdot o_j$$

Vậy khi $k \neq j$ thì $\partial\left(\frac{e^{z_k}}{\sum_{l=0}^{1} e^{z_l}}\right) / \partial z_j = -o_k o_j$

Sau 2 trường hợp để tính được $\frac{\partial o_k}{\partial z_j}$, ta có thể tính được $\frac{\partial L}{\partial z_j}$.

$$\frac{\partial L}{\partial z_j} = \cancel{\partial} - \sum_{k=0}^{1} y_k \cdot \frac{1}{o_k} \cdot \frac{\partial o_k}{\partial z_j}$$

$$= -\underbrace{\sum_{k=0, k\neq j}^{1} y_k \cdot \frac{1}{o_k} \cdot \frac{\partial o_k}{\partial z_j}}_{(k \neq j)} - \underbrace{y_j \cdot \frac{1}{o_j} \cdot \frac{\partial o_j}{\partial z_j}}_{(k = j)}$$

Mặt khác, ta có.

$$\sum_{k=0}^{1} y_k = 1.$$

$$\Leftrightarrow \sum_{k=0, k\neq j}^{1} y_k + y_j = 1.$$

$$\Leftrightarrow \sum_{k=0, k\neq j}^{1} y_k = 1 - y_j$$

Nên :

$$\frac{\partial L}{\partial z_j} = -(1-y_j)\cdot\frac{1}{o_k}\cdot\frac{\partial o_k}{\partial z_j} - y_j\cdot\frac{1}{o_j}\cdot\frac{\partial o_j}{\partial z_j}$$

$(k \neq j)$ $(k = j)$

$$= -(1-y_j)\cdot\frac{1}{o_k}\cdot(-o_k\cdot o_j) - y_j\frac{1}{o_j}\cdot o_j(1-o_j)$$

$$= (1-y_j)\cdot o_j - y_j(1-o_j).$$

$$= o_j - y_j$$

Vậy $\boxed{\frac{\partial L}{\partial z_j} = o_j - y_j}$

Vậy :

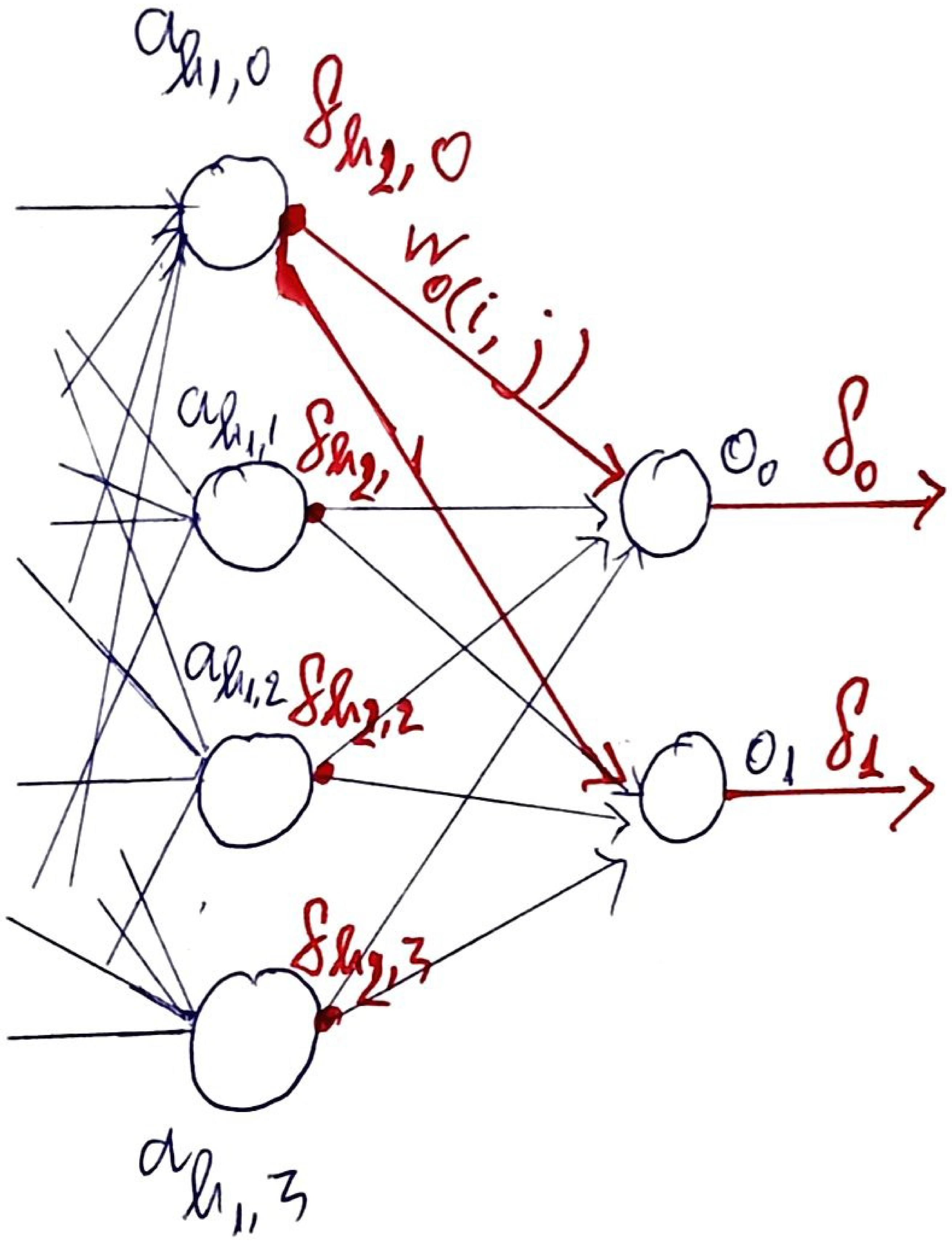
$$w_o(i,j) = w_o(i,j) - \eta\cdot(o_j - y_j)\cdot\frac{\partial z_j}{\partial w_o(i,j)}$$

Mà : $z_j = \sum_{k=0}^{3}\left(a_{h_2,k}\cdot w_o(k,j)\right) + b_{z_j}$

$\Rightarrow \partial z_j / \partial w_o(i,j) = a_{h_2,i}$

Vậy $\boxed{w_o(i,j) = w_o(i,j) - \eta(o_j - y_j)\cdot a_{h_2,i}}$

Gọi $\delta_j = o_j - y_j = \frac{\partial L}{\partial z_j}$ là sai số (error) của node thứ j của output layer.



* Giả sử $\delta_{h_2,i}$ là hàm mất mác theo $a_{h_2,i}$.

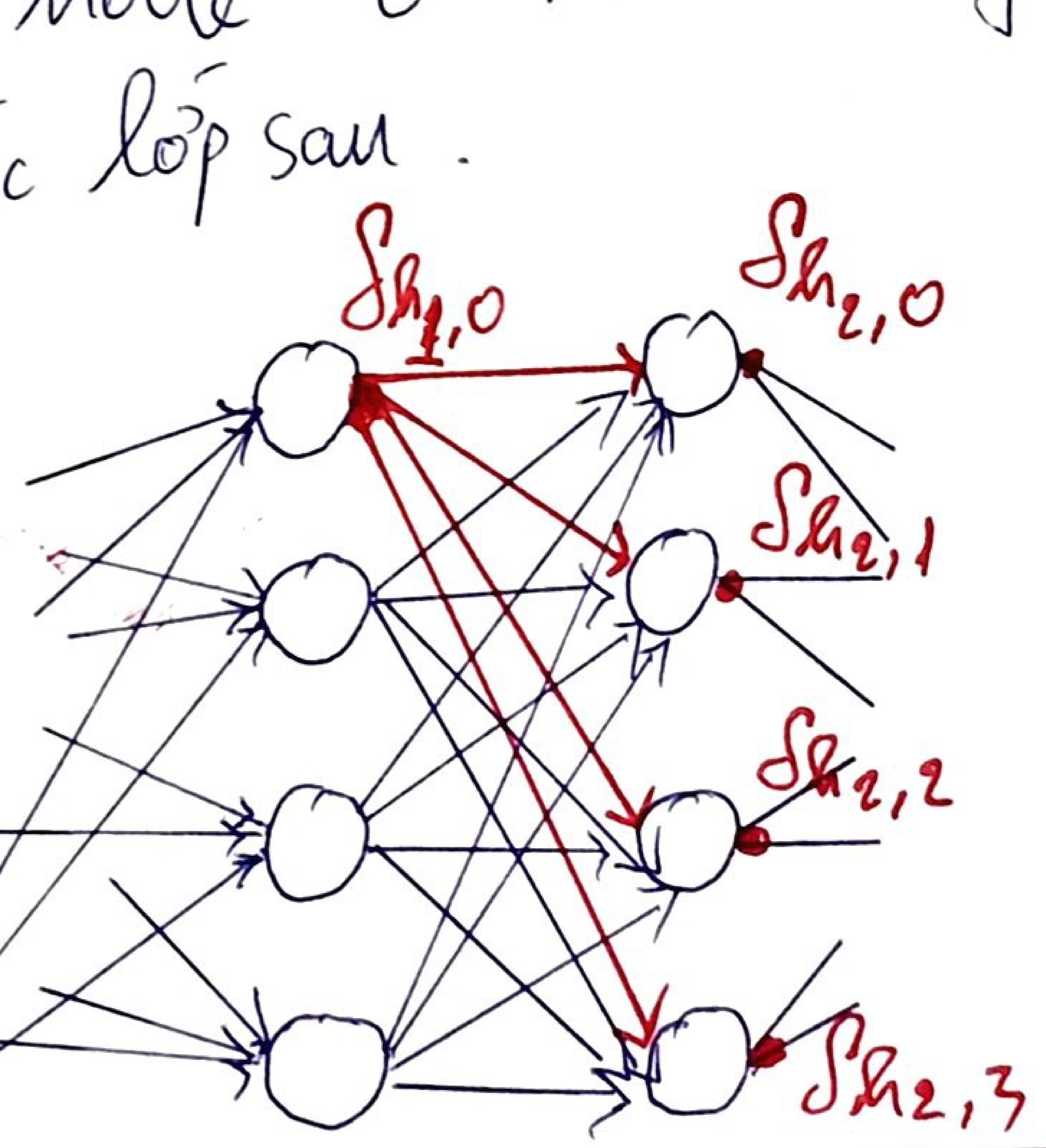
* Ta thấy rằng $\delta_{h_2,i}$ sẽ gây ảnh hưởng đến các trọng số δ_0 và δ_1 ở ngõ ra.

* Vậy để tính được các hàm mất mác $\delta_{h_2,i}$, ta tính theo công thức sau:

$$\boxed{\delta_{h_2,i} = \sum_{j=0}^{1} \delta_j \cdot w_{o(i,j)}}$$

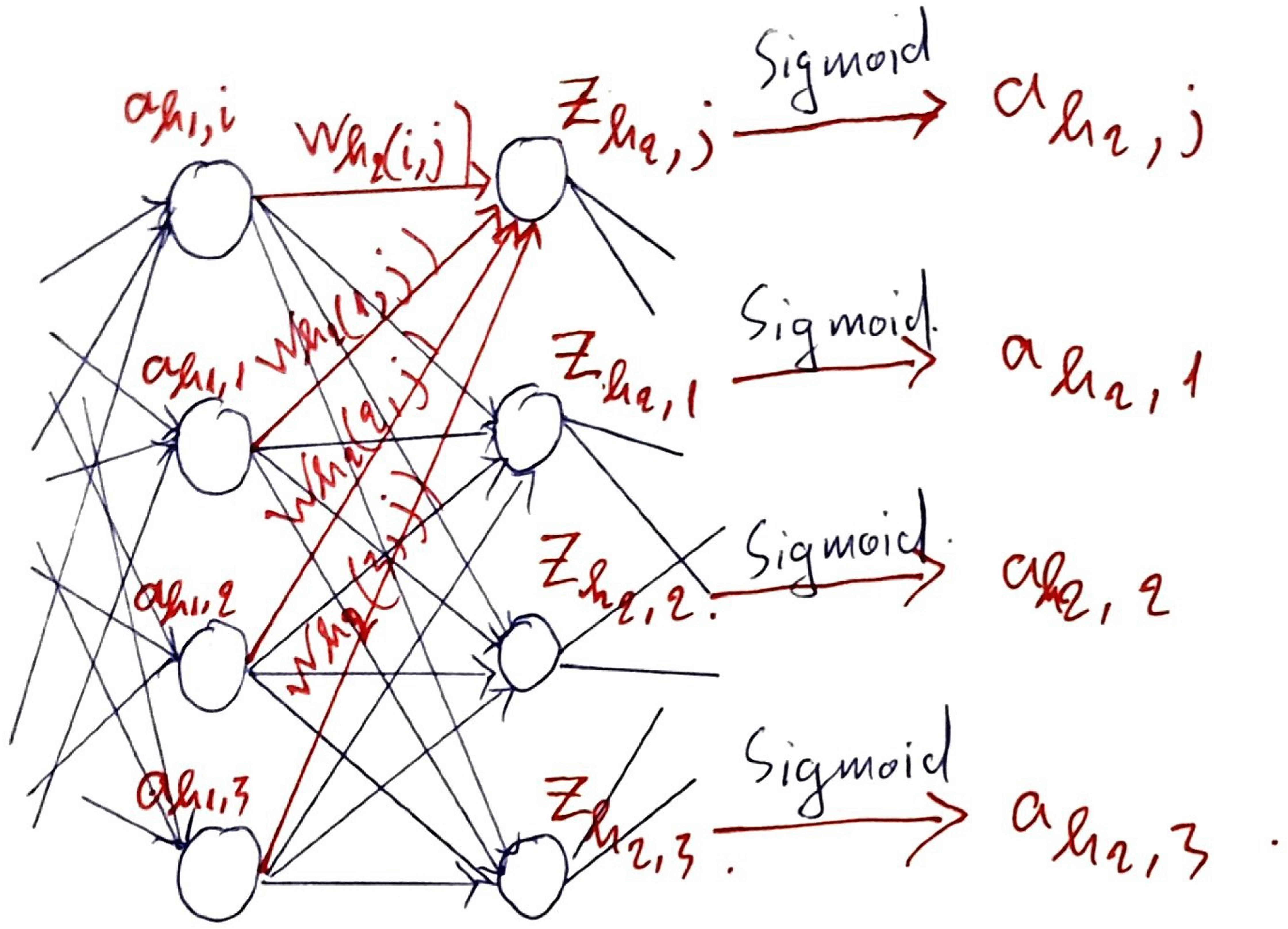
* Tương tự như vậy, error ở các node ở Hidden layer 1 cũng sẽ ảnh hưởng đến các error của các lớp sau.

$$\boxed{\delta_{h_1,i} = \sum_{j=0}^{3} \delta_{h_2,j} \cdot w_{h_2(i,j)}}$$



* Cập nhật trọng số $W_{h_2}(i,j)$

(Trọng số giữa node i của 1 st Hidden Layer và node j của 2 nd Hidden Layer)



Theo công thức Gradient Decent:

$$W_{h_2}(i,j) = W_{h_2}(i,j) - \eta \frac{\partial L}{\partial W_{h_2}(i,j)}$$

$$= W_{h_2}(i,j) - \eta \frac{\partial L}{\partial Z_{h_2,j}} \cdot \frac{\partial Z_{h_2,j}}{\partial W_{h_2}(i,j)}$$

$$= W_{h_2}(i,j) - \eta \frac{\partial L}{\partial a_{h_2,j}} \cdot \frac{\partial a_{h_2,j}}{\partial Z_{h_2,j}} \cdot \frac{\partial Z_{h_2,j}}{\partial W_{h_2}(i,j)}$$

$$= W_{h_2}(i,j) - \eta \cdot \delta_{h_2,j} \cdot \frac{\partial a_{h_2,j}}{\partial Z_{h_2,j}} \cdot \frac{\partial Z_{h_2,j}}{\partial W_{h_2}(i,j)}$$

*) Hàm sigmoid:

$$a_{h_2,j} = \frac{1}{1+e^{-z_{h_2,j}}}$$

$$\Rightarrow \frac{\partial a_{h_2,j}}{\partial z_{h_2,j}} = -\frac{\left(1+e^{-z_{h_2,j}}\right)'}{\left(1+e^{-z_{h_2,j}}\right)^2}$$

$$= \frac{e^{-z_{h_2,j}}}{\left(1+e^{-z_{h_2,j}}\right)^2}$$

$$= \frac{1+e^{-z_{h_2,j}} - 1}{\left(1+e^{-z_{h_2,j}}\right)^2}$$

$$= \frac{1}{1+e^{-z_{h_2,j}}} - \frac{1}{\left(1+e^{-z_{h_2,j}}\right)^2}$$

$$= \frac{1}{1+e^{-z_{h_2,j}}}\left(1 - \frac{1}{1+e^{-z_{h_2,j}}}\right)$$

$$= a_{h_2,j}\left(1 - a_{h_2,j}\right)$$

Vậy $w_{h_2(i,j)} = w_{h_2(i,j)} - \eta \cdot \delta_{h_2,j} \cdot a_{h_2,j} \cdot (1 - a_{h_2,j}) \cdot \dfrac{\partial z_{h_2,j}}{\partial w_{h_2(i,j)}}$

Mà: $z_{h_2,j} = \sum_{k=0}^{3} \left(a_{h_1,k} \cdot w_{h_2(k,j)}\right) + b_{h_2,j}$

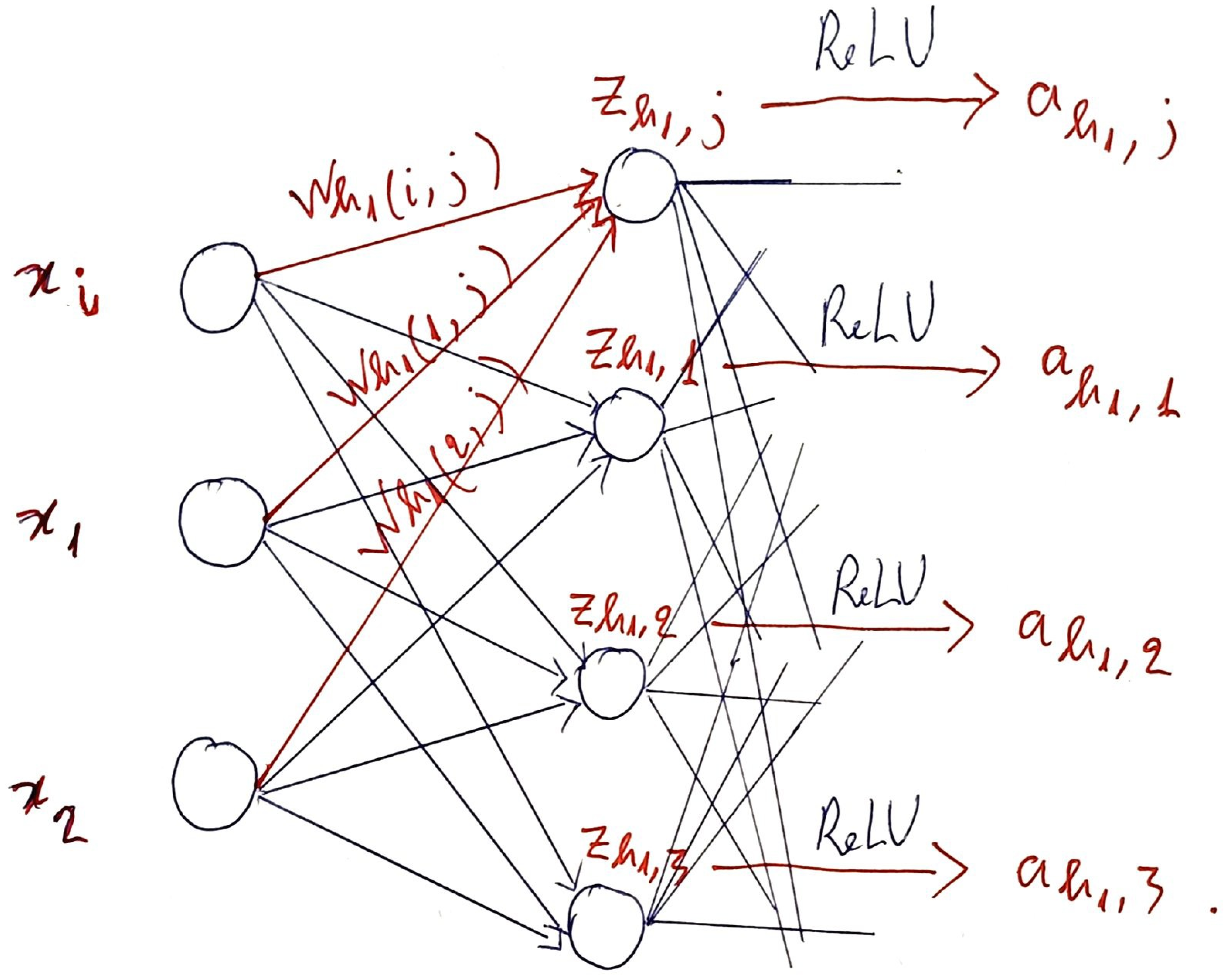
$$\Rightarrow \frac{\partial z_{h_2,j}}{\partial w_{h_2(i,j)}} = a_{h_1,i}$$

Vậy

$$\boxed{w_{h_2(i,j)} = w_{h_2(i,j)} - \eta \cdot \delta_{h_2,j} \cdot a_{h_2,j} \cdot (1 - a_{h_2,j}) \cdot a_{h_1,i}}$$

* Cập nhật trọng số $w_{h_1}(i, j)$

(Trọng số giữa node i của Input Layer và node j của 1st Hidden Layer).



Tương tự, theo Gradient Decent:

$$w_{h_1}(i, j) = w_{h_1}(i, j) - \eta \frac{\partial L}{\partial w_{h_1}(i, j)}$$

$$= w_{h_1}(i, j) - \eta \frac{\partial L}{\partial z_{h_1, j}} \cdot \frac{\partial z_{h_1, j}}{\partial w_{h_1}(i, j)}$$

$$= w_{h_1}(i, j) - \eta \frac{\partial L}{\partial a_{h_1, j}} \cdot \frac{\partial a_{h_1, j}}{\partial z_{h_1, j}} \cdot \frac{\partial z_{h_1, j}}{\partial w_{h_1}(i, j)}$$

$$= w_{h_1(i,j)} - \eta \cdot \delta_{h_1,j} \cdot \frac{\partial a_{h_1,j}}{\partial z_{h_1,j}} \cdot \frac{\partial z_{h_1,j}}{\partial w_{h_1(i,j)}}.$$

*) Hàm ReLU:

$$a_{h_1,j} = \max(0, z_{h_1,j}).$$

$$= \begin{cases} z_{h_1,j} & \text{nếu } z_{h_1,j} > 0. \\ 0 & \text{nếu } z_{h_1,j} \leq 0. \end{cases}$$

$$\Rightarrow \frac{\partial a_{h_1,j}}{\partial z_{h_1,j}} = \begin{cases} 1 & \text{nếu } z_{h_1,j} > 0. \\ 0 & \text{nếu } z_{h_1,j} \leq 0. \end{cases}$$

*) Ta có:

$$z_{h_1,j} = \sum_{k=0}^{2} \left(x_k \cdot w_{h_1(k,j)} \right) + b_{h_1,j}.$$

$$\Rightarrow \frac{\partial z_{h_1,j}}{\partial w_{h_1(i,j)}} = x_i.$$

Vậy

$$\begin{cases} w_{h_1(i,j)} = w_{h_1(i,j)}, \text{ nếu } z_{h_1,j} \leq 0. \\ \\ w_{h_1(i,j)} = w_{h_1(i,j)} - \eta \cdot \delta_{h_1,j} \cdot x_i, \text{ nếu } z_{h_1,j} > 0 \end{cases}$$

KẾT LUẬN

$$\begin{cases} \delta_j = o_j - y_j. \\ \delta_{h_2,i} = \sum_{j=0}^{1} \delta_j \cdot w_o(i,j). \\ \delta_{h_1,i} = \sum_{j=0}^{3} \delta_{h_2,j} \cdot w_{h_2}(i,j). \end{cases}$$

1> $w_{o(i,j)} = w_{o(i,j)} - \eta(o_j - y_j) \cdot a_{h_2,i}$

2> $w_{h_2}(i,j) = w_{h_2}(i,j) - \eta \delta_{h_2,j} \cdot a_{h_2,j}(1 - a_{h_2,j}) \cdot a_{h_1,i}$

3> TH1: $z_{h_1,j} \leq 0.$

$$w_{h_1}(i,j) = w_{h_1}(i,j).$$

TH2: $z_{h_1,j} > 0.$

$$w_{h_1}(i,j) = w_{h_1}(i,j) - \eta \delta_{h_1,j} \cdot x_i$$