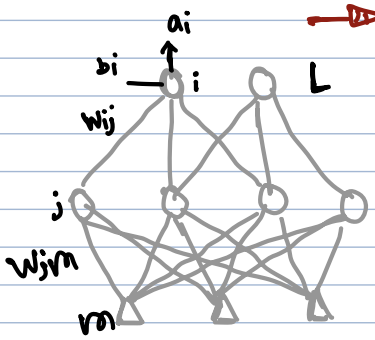
Backpropagation → Activation func in Hidden layer = ReLu
→ Activation func in Last layer = SoftMax
→ Objective func = CrossEntropy



$$a_i = \sigma(z_i)$$
$$z_i = \sum_f^h w_{if} \cdot a_f + b_i$$

$$a_i^L = \frac{e^{z_i}}{\sum_i^c e^{z_i}}$$
$$L = -\sum_i^c y_i \log(a_i^L)$$

(Last Layer) $\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial z_i} \cdot \frac{\partial z_i}{\partial w_{ij}} \Rightarrow$ (Eq 1)

$\frac{\partial L}{\partial z_i} = (a_i - y_i)$, $\frac{\partial z_i}{\partial w_{ij}} = a_j$

$$\frac{\partial L}{\partial b_i} = \frac{\partial L}{\partial z_i} \cdot \frac{\partial z_i}{\partial b_i}$$

$\frac{\partial L}{\partial z_i} = (a_i - y_i)$, $\frac{\partial z_i}{\partial b_i} = 1$

$$L = -\sum_f^c y_f \log(a_f^L)$$
$$\Rightarrow L = -\sum_f^c y_f \log\left(\frac{e^{z_f}}{\sum_j^c e^{z_j}}\right)$$
$$= -\sum_f^c y_f \left[z_f - \sum_j^c e^{z_j}\right] = -\sum_f^c y_f z_f + \sum_f^c y_f \log\left(\sum_j^c e^{z_j}\right)$$
$$\Rightarrow \frac{\partial L}{\partial z_i} = -y_i + \sum_f^c y_f \frac{\partial \left(\log \sum_j^c e^{z_j}\right)}{\partial z_i}$$
$$\Rightarrow \frac{\partial L}{\partial z_i} = -y_i + \sum_f^c y_f \cdot \frac{e^{z_i}}{\sum_j^c e^{z_j}} = -y_i + a_i \sum_f^c y_f$$

(a_i , $\sum_f^c y_f = 1$)

$$\Rightarrow \frac{\partial L}{\partial z_i} = a_i - y_i$$
$$\Rightarrow \frac{\partial z_i}{\partial w_{ij}} = a_j$$

(Hidden Layer)

$$\frac{\partial L}{\partial w_{jm}} = \frac{\partial L}{\partial z_i} \frac{\partial z_i}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{jm}}$$

$\frac{\partial L}{\partial z_i}$ → OK

$\frac{\partial z_i}{\partial z_j} = \frac{\partial z_i}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_j}$

$\frac{\partial z_i}{\partial a_j} = w_{ij}$

$\frac{\partial a_j}{\partial z_j}$ → ReLu → Derivative → 1 or 0

$\frac{\partial z_j}{\partial w_{jm}} = a_m$

(Eq 2)

$$\Rightarrow \frac{\partial L}{\partial w_{jm}} = \left[\frac{\partial L}{\partial z_i} \cdot \left[w_{ij} \cdot \text{DerivativReLu}\right]\right] a_m$$

$\frac{\partial L}{\partial z_i}$ (Upper Layer) → dz

$[w_{ij} \cdot \text{DerivativReLu}]$ → dz

$$\frac{\partial L}{\partial b_j} = \left[\frac{\partial L}{\partial z_i} \frac{\partial z_i}{\partial z_j}\right] \cdot \frac{\partial z_j}{\partial b_j}$$

$\frac{\partial z_j}{\partial b_j} = 1$