Họ và tên: Nguyễn Trung Đức

MSSV: 22161116. Lớp: 22161CL T2B.

Đề: Feed Forward Calculation:

$$\boxed{x} \xrightarrow[wh1_{j,i}]{} \boxed{zh_1} \xrightarrow[ReLU]{} \boxed{a_1} \xrightarrow[wh2_{l,j}]{} \boxed{zh_2} \xrightarrow[sigmoid]{} \boxed{a_2} \xrightarrow[wo_{k,l}]{} \boxed{Z} \xrightarrow[softmax]{} \boxed{O}$$

Bài làm:

Các trọng số:

- $wh1_{j,i}$: trọng số giữa input và hidden1.
- $wh2_{l,j}$: trọng số giữa hidden 1 và hidden 2.
- $wo_{k,l}$: trọng số giữa hidden 2 và output.

Ta có: Mô hình hoá mạng nơron như sau: ⊛ Hidden layer 1:

$$zh_1 = \sum x_i . wh1_{j,i} + b_j \rightarrow a_1 = ReLU(zh_1) = \max(0, zh_{1i}).$$

⊛ Hidden layer 2:

$$zh_2 = \sum a_1 . wh2_{l,j} + b_l \rightarrow a_2 = sigmoid(zh_2) = \frac{1}{1 + e^{-zh_2}}.$$

⊛ Output:

$$z_k = \sum a_2 . wo_{k,l} + b_k \rightarrow o_k = softmax(z_k) = \frac{e^{z_k}}{\sum e^{z_l}}.$$

⊛. Hàm Loss: $L = -\sum_{k=0} y_k . \ln(o_k)$ với: o_k: ngõ ra dự đoán.
(cross-entropy) y_k: ngõ ra sự thật.

⊛ Back propagation (cập nhật trọng số)

$$w = w - \gamma \frac{\partial L}{w}$$

- output: $wo_{k,l} = wo_{k,l} - \gamma \dfrac{\partial L}{\partial wo_{k,l}}$.
- Hidden layer 2: $wh2_{l,j} = wh2_{l,j} - \gamma \dfrac{\partial L}{\partial wh2_{l,j}}$
- Hidden layer 1: $wh1_{j,i} = wh1_{j,i} - \gamma \dfrac{\partial L}{\partial wh1_{j,i}}$

⊛ Output.

$$\frac{\partial L}{\partial wo_{k,l}} = \frac{\partial L}{\partial o_k} . \frac{\partial o_k}{\partial z_k} . \frac{\partial z_k}{\partial wo_{k,l}}$$

$$\frac{\partial L}{\partial o_k} = -\frac{y_k}{o_k} \quad ; \quad \frac{\partial z_k}{\partial wo_{k,l}} = a_2.$$

$\dfrac{\partial o_k}{\partial z_k}$ — TH1: $k = l$. — TH2: $k \neq l$.

TH1: $k = l$.

$$\frac{\partial o_k}{\partial z_k} = \partial\left(\frac{e^{z_k}}{\sum_{l=0} e^{z_l}}\right) \Big/ \partial z_k = \frac{\partial e^{z_k}/\partial z_k . \Sigma - e^{z_k} . \partial \Sigma / \partial z_k}{(\Sigma)^2}$$

$$= \frac{e^{z_k} . \Sigma - e^{z_k} . e^{z_k}}{(\Sigma)^2} = \left(\frac{e^{z_k}}{\Sigma}\right) - \left(\frac{e^{z_k}}{\Sigma}\right)^2 = o_k - o_k^2 = o_k(1 - o_k)$$

TH2) $k \neq l$.

$$\frac{\partial o_k}{\partial z_l} = \partial\left(\frac{e^{z_k}}{\Sigma e^{z_l}}\right) = \frac{0.\Sigma - e^{z_k}.e^{z_l}}{(\Sigma)^2} = -\left(\frac{e^{z_k}}{\Sigma}\right)\cdot\left(\frac{e^{z_l}}{\Sigma}\right) = -o_k.o_l.$$

$$\Rightarrow \frac{\partial L}{\partial z_k} = \frac{\partial L}{\partial o_k}\cdot\frac{\partial o_k}{\partial z_k} = \frac{-y_k}{o_k}.o_k(1-o_k) + \Sigma - \frac{y_l}{o_l}(-o_k.o_l).$$

$$= y_k(o_k - 1) + o_k \Sigma y_l.$$

$$= y_k(o_k-1) + o_k(1-y_k) = o_k - y_k.$$

$$\Rightarrow \frac{\partial L}{\partial w_{ok,l}} = \frac{\partial L}{\partial z_k}.\frac{\partial z_k}{\partial w_{ok,l}} = (o_k - y_k).a_2.$$

$$w_{ok,l} = w_{ok,l} - y\frac{\partial L}{\partial w_{ok,l}}$$

$$\Rightarrow w_{ok,l} = w_{ok,l} - y.(o_k - y_k).a_2.$$

⊛. Hidden layer 2:

$$\frac{\partial L}{\partial w_{h2l,j}} = \frac{\partial L}{\partial z_{h2}}.\frac{\partial z_{h2}}{\partial w_{h2l,j}}.$$

$$\frac{\partial z_{h2}}{\partial w_{h2l,j}} = a_1.$$

$$\frac{\partial L}{\partial z_{h2}} = \frac{\partial L}{\partial a_2}.\frac{\partial a_2}{\partial z_{h2}}$$

$$\frac{\partial L}{\partial a_2} = \Sigma\frac{\partial L}{\partial o_k}.\frac{\partial o_k}{\partial a_2} = \Sigma\frac{\partial L}{\partial o_k}.\frac{\partial o_k}{\partial z_k}.\frac{\partial z_k}{\partial a_2}.$$

$$= \Sigma(o_k - y_k).w_{ok,l}$$

$$\frac{\partial a_2}{\partial z_{h2}} = \frac{(-)(-)e^{-z_{h2}}}{(1+e^{-z_{h2}})^2} = a_2^2\left(\frac{1}{a_2} - 1\right) = a_2 - a_2^2 = a_2(1-a_2)$$

$$\frac{\partial L}{\partial w_{h2l,j}} = \frac{\partial L}{\partial z_{h2}}.\frac{\partial z_{h2}}{\partial w_{h2l,j}}$$

$$= \Sigma(o_k - y_k).w_{ok,l}.a_2(1-a_2).a_1.$$

$$w_{h2l,j} = w_{h2l,j} - y.\frac{\partial L}{\partial w_{h2l,j}}$$

$$\Rightarrow w_{h2l,j} = w_{h2l,j} - y.\Sigma(o_k - y_k)w_{ok,l}.a_2(1-a_2).a_1.$$

⊛ Hidden layer 1:

$$\frac{\partial L}{\partial w_{h1j,i}} = \frac{\partial L}{\partial z_k}.\frac{\partial z_k}{\partial a_2}.\frac{\partial a_2}{\partial z_{h2}}.\frac{\partial z_{h2}}{\partial a_1}.\frac{\partial a_1}{\partial z_{h1}}.\frac{\partial z_{h1}}{\partial w_{h1j,i}}.$$

$$\frac{\partial L}{\partial z_k} = o_k - y_k.$$

$$\frac{\partial z_k}{\partial a_2} = w_{ok,l}.$$

$$\frac{\partial a_2}{\partial z_{h_2}} = a_2(1-a_2).$$

$$\frac{\partial z_{h2}}{\partial a_1} = w_{h2}.$$

$$\frac{\partial a_1}{\partial z_{h1}} = \frac{\partial\left[ReLU(z_{h1})\right]}{\partial z_{h1}} = \frac{\partial\left[\max\{0, z_{h1}\}\right]}{\partial z_{h1}} = \begin{cases} 1 & ; z_{h1} > 0 \\ 0 & ; z_{h1} \leq 0. \end{cases}$$

$$\frac{\partial z_{h1}}{\partial w_{h1j,i}} = x_i.$$

$$\Rightarrow \frac{\partial L}{\partial w_{h1j,i}} = \begin{cases} (o_k - y_k).\, w_{ok,l}.\, a_2(1-a_2).\, w_{h2}.\, x_i & ; \text{ khi } z_{h1} > 0. \\ 0 & ; \text{ khi } z_{h1} \leq 0. \end{cases}$$

$$w_{h1j,i} = w_{h1j,i} - \eta \frac{\partial L}{\partial w_{h1j,i}}$$

$$\Rightarrow \begin{cases} w_{h1ji} = w_{h1ji} - \eta.\,(o_k - y_k).\, w_{ok,l}.\, a_2(1-a_2).\, w_{h2}.\, x_i & ; \text{ khi } z_{h1} > 0. \\ w_{h1ji} = w_{h1ji} & ; \text{ khi } z_{h1} \leq 0. \end{cases}$$