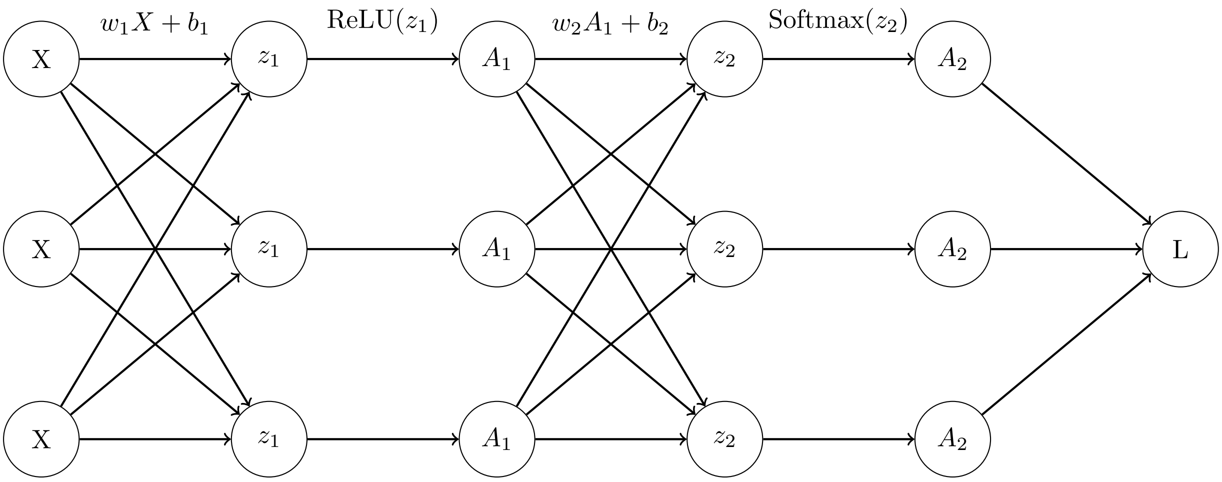
Definition



$$
\begin{aligned}
z_1 &= w_1 X + b_1 \\
A_1 &= \mathrm{ReLU}(z_1) = \begin{cases} z_{1i}, z_{1i} \geq 0 \\ 0, z_{1i} < 0 \end{cases} \\
z_2 &= w_2 A_1 + b_2 \\
A_2 &= \mathrm{Softmax}(z_2) = \frac{e^{z_{2i}}}{\sum_{j=1}^{K} e^{z_{2j}}} \\
L &= -\sum_{i=0}^{K} y_{ti} \log A_{2i}
\end{aligned}
$$

Gradients

$$
\frac{\partial L}{\partial w_2}, \frac{\partial L}{\partial b_2}, \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial b_1} - ?
$$

$$
\begin{aligned}
\frac{\partial L}{\partial w_{2i}} &= \frac{\partial L}{\partial A_{2i}} \cdot \frac{\partial A_{2i}}{\partial w_{2i}} = \frac{\partial L}{\partial A_{2i}} \cdot \frac{\partial A_{2i}}{\partial z_{2i}} \cdot \frac{\partial z_{2i}}{\partial w_{2i}} \\
\frac{\partial z_{2i}}{\partial w_{2i}} &= \frac{\partial}{\partial w_{2i}} (w_{2i} A_{1i} + b_2) = A_{1i}
\end{aligned}
$$

$$
\frac{\partial L}{\partial z_{2i}} = \frac{\partial L}{\partial A_{2i}} \frac{\partial A_{2i}}{\partial z_{2i}} = \frac{\partial}{\partial A_{2i}} \left(-\sum_{i=0}^{K} y_{ti} \log A_{2i} \right) \frac{\partial A_{2i}}{\partial z_{2i}} = -\frac{y_{ti}}{A_{2i}} \frac{\partial}{\partial z_{2i}} \left(\frac{e^{z_{2i}}}{\sum_{j=0}^{K} e^{z_{2j}}} \right) \overset{(1)}{=}
$$

$$
= \begin{cases} -\dfrac{y_{ti}}{A_{2i}} \dfrac{A_{2i} \cdot \sum_{j=0}^{K} e_{z_{2j}} - e_{z_{2i}}}{\sum_{j=0}^{K} e_{z_{2j}}}, i = k \\ -\dfrac{y_{ti}}{A_{2i}} \dfrac{-A_{2i} \cdot e_{z_{2k}}}{\sum_{j=0}^{K} e_{z_{2j}}}, i \neq k \end{cases} = \begin{cases} -y_{ti}(1 - A_{2i}), i = k \\ y_{ti} A_{2k}, i \neq k \end{cases} \overset{(2)}{=}
$$

$$
= \begin{cases} -I(y_{ti} = k)(1 - A_{2i}), i = k \\ -I(y_{ti} = k)(0 - A_{2k}), i \neq k \end{cases} = \begin{cases} A_{2i} \cdot I(y_{ti} = k) - I(y_{ti} = k), i = k \\ A_{2k} \cdot I(y_{ti} = k), i \neq k \end{cases} \Leftrightarrow
$$

$$
\Leftrightarrow A_{2i} I(y_{ti} = k) - I(y_{ti} = k) + A_{2k} I(y_{ti} = k) = A_{2i} - I(y_{ti} = k) = A_{2i} - y_{ti}
$$

$$(1) \quad \frac{\partial}{\partial z_{2i}}\left(\frac{e^{z_{2i}}}{\sum_{j=0}^{K} e^{z_{2j}}}\right) = \begin{cases} \frac{\partial}{\partial z_{2i}}\left(\frac{e^{z_{2i}}}{\sum_{j=0}^{K} e^{z_{2j}}}\right), i = k \\ \frac{\partial}{\partial z_{2i}}\left(\frac{e^{z_{2k}}}{\sum_{j=0}^{K} e^{z_{2j}}}\right), i \neq k \end{cases} = \begin{cases} \frac{e^{z_{2i}}\left(\sum_{j=0}^{K} e^{z_{2j}} - e^{z_{2_i}}\right)}{\left(\sum_{j=0}^{K} e^{z_{2j}}\right)^2}, i = k \\ -\frac{e^{z_{2k}}}{\left(\sum_{j=0}^{K} e^{z_{2j}}\right)^2} \frac{\partial\left(\sum_{j=0}^{K} e^{z_{2j}}\right)}{\partial z_{2i}}, i \neq k \end{cases}$$

$$\frac{\partial\left(\sum_{j=0}^{K} e^{z_{2j}}\right)}{\partial z_{2i}} = e^{z_{2k}}, \frac{e^{z_{2i}}}{\left(\sum_{j=0}^{K} e^{z_{2j}}\right)^2} = A_{2i}$$

$$(2) \qquad I(y_{ti} = k) = y_{ti}$$

$$\boldsymbol{\frac{\partial L}{\partial w_{2i}}} = (A_{2i} - y_{ti}) \cdot A_{1i}$$

$$\frac{\partial z_{2i}}{\partial b_2} = 1$$

$$\boldsymbol{\frac{\partial L}{\partial b_2}} = \frac{\partial L}{\partial z_{2i}} \frac{\partial z_{2i}}{\partial b_2} = \frac{\partial L}{\partial z_{2i}} = A_{2i} - y_{ti}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z_{2i}} \frac{\partial z_{2i}}{\partial w_{1i}} = \frac{\partial L}{\partial z_{2i}} \frac{\partial z_{2i}}{\partial A_{1i}} \frac{\partial A_{1i}}{\partial z_{1i}} \frac{\partial z_{1i}}{\partial w_{1i}}$$

$$\frac{\partial z_{1i}}{\partial w_{1i}} = X$$

$$\frac{\partial L}{\partial A_{1i}} = \frac{\partial L}{\partial z_{2i}} w_1 = (A_{2i} - y_{ti}) \cdot w_1$$

$$\frac{\partial A_{1i}}{\partial z_{1i}} = \begin{cases} 1, z_{1i} \geq 0 \\ 0, z_{1i} < 0 \end{cases}$$

$$\frac{\partial L}{\partial z_{1i}} = \frac{\partial L}{\partial A_{1i}} \frac{\partial A_{1i}}{\partial z_{1i}} = \begin{cases} (A_{2i} - y_{ti}) \cdot w_1, z_{1i} \geq 0 \\ 0, z_{1i} < 0 \end{cases} = (A_{2i} - y_{ti}) \cdot w_1, z_{1i} \geq 0$$

$$\frac{\partial z_{2i}}{\partial A_{1i}} = w_2$$

$$\boldsymbol{\frac{\partial L}{\partial w_{1i}}} = (A_{2i} - y_{ti}) \cdot w_2 X$$

$$\frac{\partial z_{1i}}{\partial b_1} = 1$$

$$\boldsymbol{\frac{\partial L}{\partial b_1}} = \frac{\partial L}{\partial z_{2i}} \frac{\partial z_{2i}}{\partial b_1} = \frac{\partial L}{\partial z_{2i}} \frac{\partial z_{2i}}{\partial A_{1i}} \frac{\partial A_{1i}}{\partial b_1} = \frac{\partial L}{\partial z_{2i}} \frac{\partial z_{2i}}{\partial A_{1i}} \frac{\partial A_{1i}}{\partial z_{1i}} \frac{\partial z_{1i}}{\partial b_1} = \frac{\partial L}{\partial z_{1i}} \frac{\partial z_{1i}}{\partial b_1} = \frac{\partial L}{\partial z_{1i}} \overset{z_{1i} \geq 0}{=} (A_{2i} - y_{ti}) \cdot w_1$$