T1. Compute the forward and backward pass of the following computation. Note that this is a simplified residual connection.

$$x_1 = ReLU(x_0 * w_0 + b_0)$$
$$y_1 = x_1 * w_1 + b_1$$
$$z = ReLU(y_1 + x_0)$$

Let $x_0 = 1.0$, $w_0 = 0.3$, $w_1 = -0.2$, $b_0 = 0.1$, $b_1 = -0.3$. Find the gradient of z with respect to w_0, w_1, b_0, and b_1.

Let

$$u_0 = x_0 * w_0 \qquad u_1 = x_1 * w_1 \qquad z = ReLU(v_1)$$
$$y_0 = u_0 + b_0 \qquad y_1 = u_1 + b_1$$
$$x_1 = ReLU(y_0) \qquad v_1 = y_1 + x_0$$

Forward

$$u_0 = (1.0)(0.3) = 0.3 \qquad u_1 = (0.4)(-0.2) = -0.08 \qquad z = ReLU(0.62) = 0.62$$
$$y_0 = 0.3 + 0.1 = 0.4 \qquad y_1 = -0.08 - 0.3 = -0.38$$
$$x_1 = ReLU(0.4) = 0.4 \qquad v_1 = -0.38 + 1 = 0.62$$

$$\frac{\partial}{\partial x} ReLU(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0 \end{cases}$$

Backward

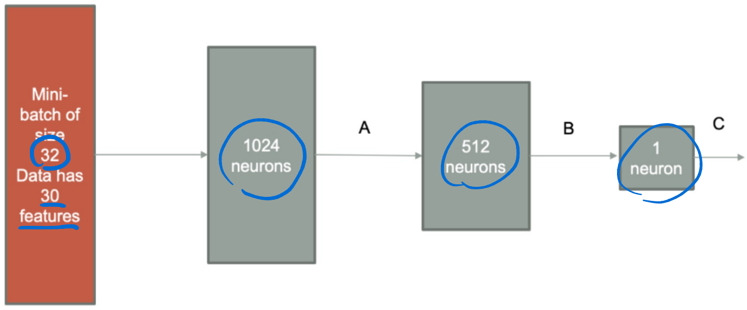
$$\frac{\partial z}{\partial w_0} = \overset{(v_1>0)}{\frac{\partial ReLU(v_1)}{\partial v_1}} \overset{1}{\frac{\partial v_1}{\partial y_1}} \overset{1}{\frac{\partial y_1}{\partial u_1}} \overset{w_1}{\frac{\partial u_1}{\partial x_1}} \overset{(y_1>0)}{\frac{\partial x_1}{\partial y_0}} \overset{1}{\frac{\partial y_0}{\partial u_0}} \overset{x_0}{\frac{\partial u_0}{\partial w_0}}$$
$$= \overset{1}{(\underset{0.62}{v_1}>0)}(1)(1)\overset{-0.2}{(w_1)}(y_1>0)(1)\overset{1.0}{(x_0)} = -0.2$$

$$\frac{\partial z}{\partial w_1} = \overset{(v_1>0)}{\frac{\partial ReLU(v_1)}{\partial v_1}} \overset{1}{\frac{\partial v_1}{\partial y_1}} \overset{1}{\frac{\partial y_1}{\partial u_1}} \overset{x_1}{\frac{\partial u_1}{\partial w_1}}$$
$$= \overset{1}{(\underset{0.62}{v_1}>0)}(1)(1)\overset{0.4}{(x_1)} = 0.4$$

$$\frac{\partial z}{\partial b_0} = \overset{(v_1>0)}{\frac{\partial ReLU(v_1)}{\partial v_1}} \overset{1}{\frac{\partial v_1}{\partial y_1}} \overset{1}{\frac{\partial y_1}{\partial u_1}} \overset{w_1}{\frac{\partial u_1}{\partial x_1}} \overset{(y_0>0)}{\frac{\partial x_1}{\partial y_0}} \overset{1}{\frac{\partial y_0}{\partial b_0}}$$
$$= \overset{1}{(\underset{0.62}{v_1}>0)}(1)(1)\overset{-0.2}{(w_1)}\overset{1}{(\underset{0.4}{y_0}>0)}(1) = -0.2$$

$$\frac{\partial z}{\partial b_1} = \overset{(v_1>0)}{\frac{\partial ReLU(v_1)}{\partial v_1}} \overset{1}{\frac{\partial v_1}{\partial y_1}} \frac{\partial y_1}{\partial b_1}$$
$$= \overset{1}{(\underset{0.62}{v_1}>0)}(1)(1) = 1$$

T2. Given the following network architecture specifications, determine the size of the output A, B, and C.



A : 1024 × 32 = 32,768

B : 512 × 32 = 16,384

C : 1 × 32 = 32

T3. What is the total number of learnable parameters in this network? (Don't forget the bias term)

Hidden layer 1

= #node + #weight

= 1024 + (30 × 1024)

= 31744

Hidden layer 2

= #node + #weight

= 512 + (512 × 1024)

= 524800

Output layer

= #node + #weight

= 1 + (1 × 512)

= 513

Total learnable parameter = 557,057

T4. Prove that the derivative of the loss with respect to h_i is $P(y=i)-y_i$. In other words, find $\frac{\partial L}{\partial h_i}$ for $i \in \{0, ..., N-1\}$ where N is the number of classes. Hint: first find $\frac{\partial P(y=j)}{\partial h_i}$ for the case where $j=i$, and the case where $j \neq i$. Then, use the results with chain rule to find the derivative of the loss.

Recall in class we define the softmax layer as:

$$P(y=j) = \frac{exp(h_j)}{\Sigma_k exp(h_k)}$$

The cross entropy loss is defined as:

$$L = -\Sigma_j y_j log P(y=j)$$

ليكن $o_j = \dfrac{exp(h_j)}{\Sigma_k exp(h_k)}$

لما $i=j$

$$\frac{\partial o_i}{\partial h_i} = \frac{\partial \frac{exp(h_i)}{\Sigma_k exp(h_k)}}{\partial h_i} = o_i(1-o_i)$$

لما $i \neq j$

$$\frac{\partial o_j}{\partial h_i} = \frac{\partial \frac{exp(h_j)}{\Sigma_k exp(h_k)}}{\partial h_i} = -o_i o_j$$

$$\frac{\partial L}{\partial h_i} = \sum_{i \neq j}\left(\frac{\partial L}{\partial o_j} \cdot \frac{\partial o_j}{\partial h_i}\right) + \frac{\partial L}{\partial o_i} \cdot \frac{\partial o_i}{\partial h_i}$$

$$= \sum_{i \neq j}\left(\frac{\partial L}{\partial o_j} \cdot \frac{\partial o_j}{\partial h_i}\right) + \frac{\partial(-\Sigma_j y_j \cdot log(o_j))}{\partial o_i} \frac{\partial o_i}{\partial h_i}$$

$$= \sum_{i \neq j}\left(\frac{\partial L}{\partial o_j} \cdot \frac{\partial o_j}{\partial h_i}\right) + \frac{\partial(y_i \, log(o_i))}{\partial o_i} \frac{\partial o_j}{\partial h_i}$$

$$= \sum_{i \neq j}\left(\frac{\partial L}{\partial o_j} \cdot \frac{\partial o_j}{\partial h_i}\right) + \left(-\frac{y_i}{o_i}\right) \frac{\partial o_i}{\partial h_i}$$

بما $\frac{\partial o_i}{\partial h_i} = o_i(1-o_i)$

$$= \sum_{i \neq j}\left(\frac{\partial L}{\partial o_j} \cdot \frac{\partial o_j}{\partial h_i}\right) + \left(-\frac{y_i}{o_i}\right) o_i(1-o_i)$$

$$= \sum_{i \neq j}\left(\frac{\partial L}{\partial o_j} \cdot \frac{\partial o_j}{\partial h_i}\right) - y_i(1-o_i)$$

$$= \sum_{i \neq j}\left(\frac{\partial(-\Sigma_{j'} y_{j'} log(o_{j'}))}{\partial o_j} \cdot \frac{\partial o_j}{\partial h_i}\right) - y_i(1-o_i)$$

بما $\frac{\partial o_j}{\partial h_i} = -o_i o_j$

$$= \sum_{i \neq j}\left(-\frac{y_j}{o_j} \cdot -o_i o_j\right) - y_i(1-o_i)$$

$$= \sum_{i \neq j} \left(y_j o_i \right) - y_i (1 - o_i)$$

$$= \underline{\sum_{i \neq j} y_j o_i} - y_i + \underline{y_i o_i}$$

$$= \sum_{j} y_j o_i - y_i$$

$$= o_i \sum_{j} y_j - y_i$$

חוכ $\sum_j y_j = 1$

$$\frac{\partial L}{\partial h_i} = o_i - y_i \quad ; \quad o_i = P(y = i)$$