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.

$x_0 = 1.0 \quad w_0 = 0.3 \quad w_1 = -0.2 \quad b_0 = 0.1 \quad b_1 = -0.3$

$$x_1 = Relu(x_0 \cdot w_0 + b_0)$$
$$= Relu(1.0 \cdot 0.3 + 0.1)$$
$$= Relu(0.4)$$
$$= 0.4.$$

$$y_1 = x_1 \cdot w_1 + b_1$$
$$= (0.4)(-0.2) + 0.1$$
$$= -0.7$$

$$z = Relu(y_1 + x_0)$$
$$= Relu(-0.7 + 1.0)$$
$$= Relu(0.3)$$
$$= 0.3.$$

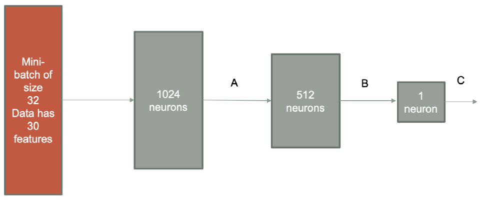
$$\frac{dz}{dw_0} = \frac{d\, Relu(y_1 + x_0)}{d\, y_1 + x_0} \frac{dy_1}{dw_0}$$
$$= Relu'(y_1 + x_0) \cdot \frac{d(x_1 \cdot w_1 + b_1)}{dw_0}$$
$$= Relu'(y_1 + x_0) \cdot w_1 \frac{d}{dw_0} Relu(x_0 w_0 + b_0).$$
$$= Relu'(y_1 + x_0) \cdot w_1\, Relu'(x_0 w_0 + b_0)\, x_0$$
$$= (1)(-0.2)(1)(1.0)$$
$$= -0.2.$$

$$\frac{dz}{db_0} = \frac{d\, Relu(y_1 + x_0)}{d\, y_1 + x_0} \frac{dy_1 + x_0}{db_0}$$
$$= \frac{d\, Relu(y_1 + x_0)}{d\, y_1 + x_0} \frac{d\, x_1 w_1 + b_1}{d\, b_0}$$
$$= Relu'(y_1 + x_0)\, w_1 \frac{d}{db_0} Relu(x_0 w_0 + b_0).$$
$$= Relu'(y_1 + x_0)\, w_1\, Relu'(x_0 w_0 + b_0)$$
$$= (1)(-0.2)(1)$$
$$= -0.2.$$

$$\frac{dz}{dw_1} = \frac{d\, Relu(y_1 + x_0)}{d\, y_1 + x_0} \frac{dy_1}{dw_0}$$
$$= Relu'(y_1 + x_0) \cdot \frac{d(x_1 \cdot w_1 + b_1)}{dw_1}$$
$$= Relu'(y_1 + x_0)\, x_1$$
$$= (1)(0.4)$$
$$= 0.4.$$

$$\frac{dz}{db_1} = \frac{d\, Relu(y_1 + x_0)}{d\, y_1 + x_0} \frac{dy_1 + x_0}{db_0}$$
$$= \frac{d\, Relu(y_1 + x_0)}{d\, y_1 + x_0} \frac{d\, x_1 w_1 + b_1}{d\, b_1}$$
$$= Relu'(y_1 + x_0)$$
$$= 1.$$

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



Minibatch's dimension 30×32

A's dimension 1024×32.

B's dimension 512×32.

C's dimension 1×32.

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

กำหนดให้ M_A, M_B, M_C คือ Matrix ที่ transform input ไปเป็น Output A, B, C ตามลำดับ.

และ $\vec{b}_A, \vec{b}_B, \vec{b}_C$ คือ พจน์ Bias ที่ใช้ในการบวกเพิ่มหลังจาก transform เป็น Output A, B, C ตามลำดับ.

M_A มี Dimension 30×1024 $\quad$ $\vec{b}_A$ มี Dimension 1024×1

M_B มี Dimension 1024×512 $\quad$ $\vec{b}_B$ มี Dimension 512×1

M_C มี Dimension 512×1 $\quad$ $\vec{b}_C$ มี Dimension 1×1

$\therefore$ ภายใน Network ในข้อนี้ มีจำนวน Learnable Parameter ทั้งสิ้น 557,057 ตัว.

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.

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

$$= -\sum_j y_j \log \frac{\exp(h_j)}{\sum_k \exp(h_k)}.$$

$$= -\sum_j \left[y_j \log \exp(h_j) - \log \sum_k \exp(h_k). \right].$$

$$= -\sum_j \left[y_j h_j - \log \sum_k \exp(h_k) \right]$$

$$\frac{\partial L}{\partial h_i} = -\frac{\partial}{\partial h_i} \sum_j \left[y_j h_j - \log \sum_k \exp(h_k). \right].$$

$$= -\left[y_i - \frac{\partial}{\partial h_i} \log \sum_k \exp(h_k) \right].$$

$$= -\left[y_i - \frac{\partial \log \sum_k \exp(h_k)}{\partial \sum_k \exp(h_k)} \frac{\partial \sum_k \exp(h_k)}{\partial h_i} \right]$$

$$= -\left[y_i - \frac{\exp(h_i)}{\sum_k \exp(h_k)} \right].$$

$$= \frac{\exp(h_i)}{\sum_k \exp(h_k)} - y_i$$

$$= P(y=i) - y_i$$