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.

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

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

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

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

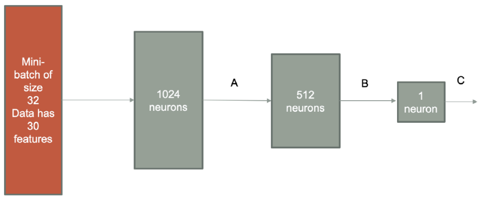
$$\frac{dz}{dw_0} = \frac{d}{d\,y_1+x_0} Relu(y_1+x_0) \frac{dy_1}{dw_0}$$
$$= Relu'(y_1+x_0) \cdot \frac{d}{dw_0}(x_1 \cdot w_1 + b_1)$$
$$= Relu'(y_1+x_0) \cdot w_1 \frac{d}{dw_0} Relu(x_0 w_0 + b_0).$$
$$= Relu'(y_1+x_0) \cdot w_1\, Relu'(x_0 w_0 + b_0)\, x_0$$
$$= (1)(-0.2)(1)(1.0)$$
$$= -0.2.$$

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

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

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

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



Mimibatch'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.

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

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

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

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

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

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

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

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

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

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

Gradient

$$L = -\sum_j \sum_i y_{ji} \log(P_j = i) + \frac{\lambda}{2} \sum_i w_i^2$$

$$y = \begin{array}{cc} & D \\ N & \begin{bmatrix} \quad \end{bmatrix} \end{array}$$

W_1 $\quad$ n = # input $\quad$ m = # neuron

$\frac{\partial L}{\partial W_i}$

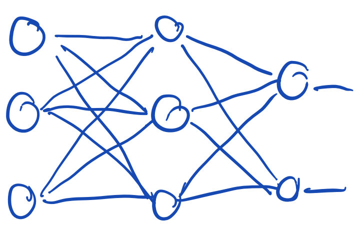
$$\begin{array}{cc} & m \\ n & \begin{bmatrix} \frac{\partial L}{\partial W_i} \end{bmatrix} \end{array}$$

X . $N \times D$

$W_1 = D \times H$

$X_1 = XW_1 + b$ $\quad N \times H$

$X_2 = X_1 W_2 + b$ $\quad N \times C$.



Sigmoid

$$P(X_j = i) = \frac{e^{W_2 X_2 + b_2}}{\sum e^{W_2 X_2 + b_2}}$$

$h = w_2 x_2 + b_2$

$$\frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial W_2}$$

$P(X_j = 1) =$ $\quad$ (H).

$$\frac{dL_1}{dh} = \begin{bmatrix} \frac{\partial L}{\partial h_{i1}} & \frac{\partial L}{\partial h_{i2}} & \frac{\partial L}{\partial h_{i3}} \\ \vdots & \vdots & \vdots \end{bmatrix}_{N*C} = \begin{bmatrix} P(y_{i1}=1)-y_{i1} & P(y_{i2}=2)-y_{i2} & P(y_{i3}=1)-y_{i3} \\ & & \end{bmatrix}$$

$$\frac{dL_2}{dw_1} = \begin{bmatrix} \frac{\lambda}{\cancel{2}} \cdot \cancel{2} w_1 \end{bmatrix}_{D*H} \qquad \frac{dL_2}{dw_2} = \begin{bmatrix} \; \end{bmatrix}_{H*C}.$$

$h = x_1 w_2 + b_2$

$$\frac{dh}{dw_2} = \begin{bmatrix} x_1 & \cdots \\ \vdots & \ddots \end{bmatrix}_{N*H}. \qquad \frac{dh}{db_2} = \begin{bmatrix} 1 & \cdots \\ \vdots & \ddots \end{bmatrix}_{N*1}.$$

$x_1 = \mathrm{Relu}(xw_1 + b_1)$ & $x' = \mathrm{Relu}(xw_1 + b_1)$

$$\frac{dx_1}{dw_1} = \begin{bmatrix} \text{ismorethan}(x_1)w_1 & \cdots \\ \vdots & \ddots \end{bmatrix}_{D*H} \qquad \frac{dh}{dx_1} = \begin{bmatrix} w_2 & \cdots \\ \vdots & \ddots \end{bmatrix}_{H*C}.$$

$$\frac{dx_1}{db_1} = \begin{bmatrix} \text{ismorethan0}(x_1). & \cdots \\ \vdots & \ddots \end{bmatrix}_{N*H}.$$

$$L = -\sum_j \sum_i y_{ji} \log(P_j = i) + \frac{\lambda}{2} \sum_i w_i^2$$

$$= -\sum_j \sum_i y_{ji} (\log \exp(h_{ji}) - \log \sum_i \exp(h_{ji})) + \frac{\lambda}{2} \sum_i w_i^2$$

$$h = x_1 w_2 + b_2$$

$$x_1 = x w_1 + b_1$$

$$\frac{dL}{dw_2} = \frac{\partial L}{\partial h} \frac{\partial h}{\partial w_2}$$

$(H, C)^2$

$$\frac{\partial L}{\partial w_{2_{ij}}} = \frac{\partial L}{\partial h} \frac{\partial h}{\partial w_{2_{ij}}}$$

$$\underset{(1,D)}{\vec{x}} \xrightarrow[(D,H)]{W_1} \oplus \to \underset{(1,H)}{\vec{x}_1} \xrightarrow{\text{Relu}} \bigcirc \xrightarrow[(H,C)]{W_2} \oplus \to \underset{(1,C)}{\vec{x}_2} \xrightarrow{\text{Softmax}} \underset{(1,C)}{\vec{h}} \to L.$$

$\vec{b}_1$ $(1,H)$ $\to \oplus$; $\vec{b}_2$ $(1,C)$ $\to \oplus$

$$\frac{\partial L}{\partial b_{2i}} = \frac{\partial L}{\partial \vec{h}} \frac{\partial \vec{h}}{\partial \vec{x}_2} \frac{\partial \vec{x}_2}{\partial b_{2i}}$$

$$\frac{\partial L}{\partial \vec{h}}$$

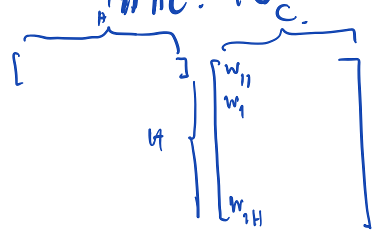
$$L = -\sum_j y_j \log(P_j = i) + \frac{\lambda}{2} \sum_i w_i^2 \qquad h = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

$$= -\sum_i y_j \log h + \frac{\lambda}{2} \sum w_i^2$$

$$\frac{\partial L}{\partial \vec{h}} = \left[-\frac{y_1}{h_j} \quad -\frac{y_2}{h_2} \quad -\frac{y_3}{h_3} \right].$$

$$\frac{\partial L}{\partial w_{2i}} = \frac{\partial L}{\partial \vec{h}} \frac{\partial \vec{h}}{\partial \vec{x}_2} \frac{\partial \vec{x}_2}{\partial w_{2i}} \qquad \vec{x}_2 = \vec{x}_1 W_1 + \vec{b}$$

$1HHC.\ 1C_C.$



$\vec{x}\vec{w}_{1:}, \vec{x}\vec{w}_{2:}, \vec{x}\vec{w}_{3:}, \ldots$

อันที่ m

$$\text{อันที่ } m\left(\frac{\partial \vec{x}_2}{\partial w_{ij}}\right) = [\underbrace{0\ 0\ 0\ x}_{i}\ 0\ 0\ 0\ 0\ 0].$$

$i \to 1H$, $j \to 1C$

$$\frac{\partial L_j}{\partial h} = \begin{bmatrix} P(y=1) - y_1 & P(y=2) - y_2 & P(y=3) - y_3 \end{bmatrix}.$$

$C \times 1$. $\quad \frac{\partial L}{\partial h_1}$

$$\frac{\partial \vec{h}}{\partial \vec{x}_2} = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_c} \\ \vdots & & \vdots \\ \frac{\partial h_c}{\partial x_1} & \cdots & \frac{\partial h_c}{\partial x_c} \end{bmatrix}.$$

$$\frac{\partial \frac{e^{x_1}}{\sum e^{x_i}}}{\partial x_1} = \frac{\sum e^{x_i} e^{x_1} - e^{x_1} e^{x_1}}{(\sum e^{x_i})^2} \qquad \frac{\partial \frac{e^{x_1}}{\sum e^{x_1}}}{\partial x_2} = \frac{\cancel{x} - e^{x_1} e^{x_2}}{(\sum e^{x_i})^2}$$

$$\frac{\partial \vec{x}_2}{\partial b_{2i}} = \begin{bmatrix} 1 & 0 & 0. \end{bmatrix} \qquad x_2 = \vec{x}_1 W_2 + \vec{b}_2$$

$b_{21} \quad b_{22} \quad b_{23}$

$$\frac{\partial L}{\partial h} = \frac{1}{N} \sum_j \frac{\partial L_j}{\partial b_{2i}}$$

$\cancel{h}$ ... $\cancel{h}$

$x_1 \quad \ldots \quad x_c$

$x_1 \quad x_2$

$b_1 \quad b_1$

x_2

b_1

$$\frac{\partial L_j}{\partial x_{2i}} = \frac{\partial \vec{x}_2}{\partial x_{2i}} \cdot \frac{\partial \vec{h}}{\partial \vec{x}_2} \cdot \frac{\partial L_j}{\partial h}$$

$(1 \times C)\,(C \times C)\,(C \times 1).$

x_c

b_1

$$\frac{\partial x_{11}}{\partial x_{21}} \quad \frac{\partial x_{11}}{\partial x_{22}} \quad \vdots \quad \frac{\partial x_{11}}{\partial x_{2c}}$$

$$\frac{\partial L}{\partial x_{1i}} = \frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial \vec{x}_2} \cdot \frac{\partial \vec{x}_2}{\partial x_{1i}}$$

(จำนวน $\times$ dim) $(1 \times C)\,(C \times C)\,(C, 1)$

$$\frac{\partial L}{\partial b_{1i}} = \frac{\partial L}{\partial \vec{x}_1} \cdot \frac{\partial \vec{x}_1}{\partial b_{1i}}$$

$1 \times H \quad \times \quad H \times 1.$

$$L_1 = -\Sigma_j y_j log P(y = j) \qquad (2)$$

where y_j is 1 if y is class j, and 0 otherwise.

Pc

$$\frac{\partial L}{\partial b_2} = \frac{\partial \vec{h}}{\partial b_2} \frac{\partial L}{\partial \vec{h}} \qquad \frac{\partial L}{\partial \vec{h}}$$

$h = [\ -\ -\ -\ -\ -\ ].$

$h \quad N \times C \qquad L \quad 1$

$b \quad 1 \times C \qquad h \quad N \times C$

$$L = -\sum_i \sum_j^c y_i \log P(x_i = j) + \frac{\lambda}{2} \sum_k w_k^2.$$

$$L = L_1 + L_2$$

$$\frac{\partial L_j}{\partial \vec{b}} = \frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial b_{2i}} \qquad 1 \times C$$

$$= \frac{\partial \vec{x}_2}{\partial \vec{b}_{2i}} \frac{\partial \vec{h}}{\partial \vec{x}_2} \cdot \frac{\partial L}{\partial \vec{h}}$$

$N \times [C, 1] \quad C \times C. \quad C \times C \quad C \times 1.$

$\vec{x}_2 = Relu(\vec{y}_1).$

$$\frac{\partial L}{\partial b} = \frac{1}{N} \cdot \sum_j \frac{\partial L_j}{\partial b_{2i}}$$

$\vec{x}_2 = Rule(x_1 w_2 + b_2)$

$(1, H)(H, C) \quad (1, C)$

$[x_1 \quad 1 \times H] \quad [w_1 \quad].$

$$\frac{\partial \vec{x}_2}{\partial \vec{x}_1} = \vec{x}_1 \Bigg\{ \Bigg[\overbrace{\partial x_{21} \qquad\qquad}^{\vec{x}_2} \Bigg].$$

$$\frac{\partial \vec{x}_2}{\partial x_{1i}} = 0 \qquad 0$$

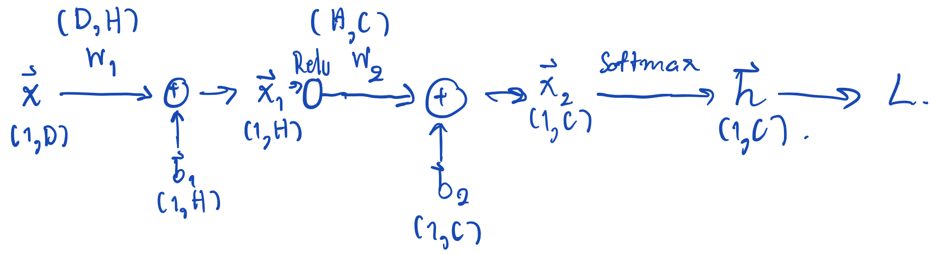
$$\frac{\partial x_{2i}}{\partial x_{1j}} = w_{ji} \cdot sign(x_{2,i})$$

$$\frac{\partial L}{\partial \vec{x}_2} \quad \frac{\partial \vec{X}_2}{\partial \vec{X}_1}$$

$1 \times C \ \times \quad C \times H$

$$\frac{\partial \vec{x}_2}{\partial Rule} \quad \frac{\partial Relu}{\partial \vec{x}_1} = Relu'(x_2)$$

$$\frac{\partial \vec{x}_2}{\partial \vec{x}_1} = \frac{\partial \vec{x}_2}{\partial \vec{y}_1} \frac{\partial \vec{y}_1}{\partial \vec{x}_1}$$



$$\frac{\partial L}{\partial \vec{x}_2} = N, 1, C$$

$$\frac{\partial l}{d b_2} \quad C, N, 1, 1 \qquad \vec{b} \quad (1, C.$$

$$\frac{d l}{d W_2}$$

$$\frac{\partial l}{\partial b_1} = \frac{\partial l}{\partial \vec{x}_1} \cdot \frac{\partial \vec{x}_1}{\partial b_1}$$

$$\frac{\partial l}{\partial W_1} = \frac{\partial l}{\partial x_1} \cdot \frac{\partial x_1}{\partial W_1}$$