Part Two: Back-propogation

姓名：周泽龙
学号：2020213990
课程：深度学习
日期：2021年3月22日

Block One: gradients of some basic layers

(i) BatchNorm

Given a standard BatchNorm layer, the gradients of the output $y_i = BN_{\gamma,\beta}(x_i)$ with respect to the parameters of γ, β are:

$$\frac{\partial y_i}{\partial \gamma} = \hat{x}_i = \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \tag{1}$$

$$\frac{\partial y_i}{\partial \beta} = 1 \tag{2}$$

(ii) Dropout

Given a dropout layer, we have a random mask vector M generated by a random vector r:

$$r = rand() = (r_1, r_2, \ \dots, \ r_n) \tag{3}$$

$$M = (M_1, M_2, \ \dots \ , M_n) \tag{4}$$

$$M_j = \begin{cases} 0, & r_j < p \\ \frac{1}{1-p}, & r_j \geq p \end{cases} \tag{5}$$

The gradient of the i_{th} output y_i with respect to the j_{th} input x_j is:

$$\frac{\partial y_i}{\partial x_j} = \begin{cases} M_j = \begin{cases} 0, & r_j < p \\ \frac{1}{1-p}, & r_j \geq p \end{cases} & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{6}$$

(iii) Softmax Function

Given the i_{th} input z_i and the i_{th} output y_i， we have:

$$y_i = \frac{e^{z_i}}{\sum_{k=1}^{n} e^{z_k}}$$

The gradient of the i_{th} output y_i with respect to the j_{th} input z_j is:

$$\frac{\partial y_i}{\partial z_j} = \begin{cases} \frac{e^{z_i}(\sum_{k=1}^{n} e^{z_k}) - (e^{z_i})^2}{(\sum_{k=1}^{n} e^{z_k})^2} = y_i(1 - y_i) & (i = j) \\ \frac{-e^{z_i} e^{z_j}}{(\sum_{k=1}^{n} e^{z_k})^2} = -y_i y_j & (i \neq j) \end{cases} \tag{7}$$

Block Two: feed-forward and backpropagation of the multi-task network

(i) feed-forward

Task A

FC_{1A}**:**

$$
\begin{aligned}
z_{1A} &= \theta_{1A}x + b_{1A} && (8) \\
a_{1A} &= sin(z_{1A}) = sin(\theta_{1A}x + b_{1A}) && (9)
\end{aligned}
$$

DP**:**

$$
\begin{aligned}
x_{DP} &= a_{1A} \circ M && (10) \\
M &= (M_1, M_2, \ldots, M_n) \\
M_j &= \begin{cases} 0, & r_j < p \\ \frac{1}{1-p}, & r_j \geq p \end{cases}
\end{aligned}
$$

The notation $\circ$ in formula (3) means element-wise product of two vectors.

FC_{2A}**:**

$$
\begin{aligned}
\hat{y}_A &= \theta_{2A}x_{DP} + b_{2A} \\
&= \theta_{2A}((sin(\theta_{1A}x + b_{1A})) \circ M) + b_{2A} && (11)
\end{aligned}
$$

Task B

FC_{1B} **and** BN^***:**

$$
\begin{aligned}
x_{1B} &= \theta_{1B}x && (12) \\
\mu &= \frac{1}{m}\sum_{i=1}^{m} x_{1B}^{i} && (13) \\
x_{BN} &= x_{1B} - \mu + b_{1B} && (14) \\
a_{BN} &= ReLU(x_{BN}) \\
&= max(x_{BN}, 0) && (15)
\end{aligned}
$$

FC_{2B}**:**

$$
\begin{aligned}
z_{2B} &= \theta_{2B}(\hat{y}_A + a_{BN}) + b_{2B} && (16) \\
\hat{y}_B &= a_{2B} \\
&= softmax(z_{2B}) \\
&= \frac{1}{\sum_{j=1}^{k} e^{z_{2B_j}}} e^{z_{2B}} && (17)
\end{aligned}
$$

(ii) backpropagation

Task B

FC_{2B}**:**

$$\frac{\partial L}{\partial \hat{y}_B} = -\frac{1}{m}\sum_{i=1}^{m}\frac{y_{Bi}}{\hat{y}_{Bi}} \tag{18}$$

$$\frac{\partial \hat{y}_{Bi}}{\partial z_{2Bj}} = \begin{cases} y_i(1-y_i) & (i=j) \\ -y_i y_j & (i \neq j) \end{cases} \tag{19}$$

$$\frac{\partial z_{2B}}{\partial \theta_{2B}} = y_A + a_{BN} \tag{20}$$

$$\frac{\partial z_{2B}}{\partial b_{2B}} = 1 \tag{21}$$

综 上，

$$\frac{\partial L}{\partial \theta_{2B}} = \frac{\partial L}{\partial \hat{y}_B}\frac{\partial \hat{y}_B}{\partial z_{2B}}\frac{\partial z_{2B}}{\partial \theta_{2B}} \tag{22}$$

$$\frac{\partial L}{\partial b_{2B}} = \frac{\partial L}{\partial \hat{y}_B}\frac{\partial \hat{y}_B}{\partial z_{2B}}\frac{\partial z_{2B}}{\partial b_{2B}} \tag{23}$$

FC_{1B} **and** BN^***:**

$$\frac{\partial z_{2B}}{\partial a_{BN}} = \theta_{2B} \tag{24}$$

$$\frac{\partial a_{BN}}{\partial x_{BN}} = 1 \tag{25}$$

$$\frac{\partial x_{BN}}{\partial x_{1B}} = 1 \tag{26}$$

$$\frac{\partial x_{1B}}{\partial \theta_{1B}} = x \tag{27}$$

$$\frac{\partial x_{BN}}{\partial b_{1B}} = 1 \tag{28}$$

综 上，

$$\frac{\partial L}{\partial \theta_{1B}} = \frac{\partial L}{\partial \hat{y}_B}\frac{\partial \hat{y}_B}{\partial z_{2B}}\frac{\partial z_{2B}}{\partial a_{BN}}\frac{\partial a_{BN}}{\partial x_{BN}}\frac{\partial x_{BN}}{\partial x_{1B}}\frac{\partial x_{1B}}{\partial \theta_{1B}} \tag{29}$$

$$\frac{\partial L}{\partial b_{1B}} = \frac{\partial L}{\partial \hat{y}_B}\frac{\partial \hat{y}_B}{\partial z_{2B}}\frac{\partial z_{2B}}{\partial a_{BN}}\frac{\partial a_{BN}}{\partial x_{BN}}\frac{\partial x_{BN}}{\partial b_{1B}} \tag{30}$$

Task A

FC_{2A}**:**

$$L_A = \frac{1}{2m}\sum_{i=1}^{m}||\hat{y}_{Ai} - y_{Ai}||_2^2 \tag{31}$$

$$\frac{\partial L_A}{\partial \hat{y}_A} = \frac{1}{m}\sum_{i=1}^{m}\hat{y}_{Ai} - y_{Ai} \tag{32}$$

$$\frac{\partial \hat{y}_A}{\partial \theta_{2A}} = x_{DP} \tag{33}$$

$$\frac{\partial \hat{y}_A}{\partial b_{2A}} = 1 \tag{33}$$

$$\frac{\partial z_{2B}}{\partial \hat{y}_A} = \theta_{2B} \tag{34}$$

综 上，

$$\frac{\partial L}{\partial \theta_{2A}} = \frac{\partial L_A}{\partial \hat{y}_A}\frac{\partial \hat{y}_A}{\partial \theta_{2A}} + \frac{\partial L}{\partial \hat{y}_B}\frac{\partial \hat{y}_B}{\partial z_{2B}}\frac{\partial z_{2B}}{\partial \hat{y}_A}\frac{\partial \hat{y}_A}{\partial \theta_{2A}} \tag{35}$$

$$\frac{\partial L}{\partial b_{2A}} = \frac{\partial L_A}{\partial \hat{y}_A}\frac{\partial \hat{y}_A}{\partial b_{2A}} + \frac{\partial L}{\partial \hat{y}_B}\frac{\partial \hat{y}_B}{\partial z_{2B}}\frac{\partial z_{2B}}{\partial \hat{y}_A}\frac{\partial \hat{y}_A}{\partial b_{2A}} \tag{36}$$

DP:

$$\frac{\partial x_{DPi}}{\partial a_{1Aj}} = \begin{cases} M_j = \begin{cases} 0, & r_j < p \\ \frac{1}{1-p}, & r_j \geq p \end{cases} & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{37}$$

FC_{1A}:

$$\frac{\partial a_{1A}}{\partial z_{1A}} = cos(z_{1A}) \tag{38}$$

$$\frac{\partial z_{1A}}{\partial \theta_{1A}} = x \tag{39}$$

$$\frac{\partial z_{1A}}{\partial b_{1A}} = 1 \tag{40}$$

$$\frac{\partial \hat{y}_A}{\partial x_{DP}} = \theta_{2A} \tag{40}$$

综上，

$$\frac{\partial L}{\partial \theta_{1A}} = \frac{\partial L_A}{\partial \hat{y}_A}\frac{\partial \hat{y}_A}{\partial x_{DP}}\frac{\partial x_{DP}}{\partial a_{1A}}\frac{\partial a_{1A}}{\partial z_{1A}}\frac{\partial z_{1A}}{\partial \theta_{1A}} + \frac{\partial L}{\partial \hat{y}_B}\frac{\partial \hat{y}_B}{\partial z_{2B}}\frac{\partial z_{2B}}{\partial \hat{y}_A}\frac{\partial \hat{y}_A}{\partial x_{DP}}\frac{\partial x_{DP}}{\partial a_{1A}}\frac{\partial a_{1A}}{\partial z_{1A}}\frac{\partial z_{1A}}{\partial \theta_{1A}} \tag{41}$$

$$\frac{\partial L}{\partial b_{1A}} = \frac{\partial L_A}{\partial \hat{y}_A}\frac{\partial \hat{y}_A}{\partial x_{DP}}\frac{\partial x_{DP}}{\partial a_{1A}}\frac{\partial a_{1A}}{\partial z_{1A}}\frac{\partial z_{1A}}{\partial b_{1A}} + \frac{\partial L}{\partial \hat{y}_B}\frac{\partial \hat{y}_B}{\partial z_{2B}}\frac{\partial z_{2B}}{\partial \hat{y}_A}\frac{\partial \hat{y}_A}{\partial x_{DP}}\frac{\partial x_{DP}}{\partial a_{1A}}\frac{\partial a_{1A}}{\partial z_{1A}}\frac{\partial z_{1A}}{\partial b_{1A}} \tag{42}$$