1. BatchNorm 梯度计算

L 为损失函数，求 $\frac{\partial L}{\partial \gamma}$，$\frac{\partial L}{\partial \beta}$

$y = \gamma \hat{x} + \beta$

$$\frac{\partial L}{\partial \gamma} = \sum_i \frac{\partial L}{\partial y_i} \frac{\partial y_i}{\partial \gamma} = \sum_i \frac{\partial L}{\partial y_i} \hat{x}_i = \frac{\partial L}{\partial y} \cdot \hat{x}$$

$$\frac{\partial L}{\partial \beta} = \sum_i \frac{\partial L}{\partial y_i} \frac{\partial y_i}{\partial \beta} = \sum \frac{\partial L}{\partial y_i} = \frac{\partial L}{\partial y}$$

2. Dropout 梯度计算

Input : $x = (x_1, x_2, \dots, x_n)$

$$\hat{x} = \text{dropout}(x) = x \cdot M$$

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

$$\frac{\partial \hat{x}}{\partial x} = M$$

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial \hat{x}} \cdot \frac{\partial \hat{x}}{\partial x} = \frac{\partial L}{\partial \hat{x}} \cdot M$$

3. softmax 梯度计算

输入 z，输出 a

$$a_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}, \quad j \in \{1, 2, \dots, K\}$$

$$\frac{\partial a_j}{\partial z_j} = \frac{e^{z_j} \cdot \left(\sum_{k=1}^{K} e^{z_k} - e^{z_j}\right)}{\left(\sum_{k=1}^{K} e^{z_k}\right)^2} = a_j(1-a_j)$$

$j \neq k$,

$$\frac{\partial a_k}{\partial z_j} = \frac{-e^{z_k} \cdot e^{z_j}}{\left(\sum_{k=1}^{K} e^{z_k}\right)^2} = -a_j a_k$$

$$\frac{\partial a}{\partial z_k} = \begin{bmatrix} \frac{\partial a_1}{\partial z_k} \\ \vdots \\ \frac{\partial a_k}{\partial z_k} \\ \vdots \\ \frac{\partial a_K}{\partial z_k} \end{bmatrix} = \begin{bmatrix} -a_1 \\ \vdots \\ (1-a_k) \\ \vdots \\ -a_K \end{bmatrix} = \left(\begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} - a \right) a_k$$

$$\frac{\partial a_i}{\partial z_j} = \begin{cases} a_i(1-a_i), & i=j \\ -a_i a_j, & i \neq j \end{cases}$$

$$\frac{\partial L}{\partial z_k} = \frac{\partial L}{\partial a} \frac{\partial a}{\partial z_k} = \left(\frac{\partial L}{\partial a_k} - \left[\frac{\partial L}{\partial a}\right]^T a\right) a_k .$$

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \frac{\partial a}{\partial z} = \left(\frac{\partial L}{\partial a} - \left[\frac{\partial L}{\partial a}\right]^T a\right) \odot a$$

$\odot$ 为点乘

Block 2.

经过 FC_{1A}

$$z_{1A} = \sin(\theta_{1A} x + b_{1A})$$

经过 DP

$$z_{DP} = z_{1A} \odot M = \sin(\theta_{1A} x + b_{1A}) \odot M$$

经过 FC_{2A}

$$z_{2A} = \theta_{2A} z_{DP} + b_{2A}$$

$$y_A = z_{2A} = \theta_{2A} [\sin(\theta_{1A} x + b_{1A}) \odot M] + b_{2A}$$

经过 FC_{1B}

$$x_{1B} = \theta_{1B} x$$

$$\mu = \frac{1}{m} \sum_{i=1}^{m} x_{1B}^{i}$$

经过 BN

$$x_{BN} = x_{1B} - \mu + b_{1B} = \theta_{1B} x - \frac{1}{m} \sum_{i=1}^{m} x_{1B}^{i} + b_{1B}$$

$$z_{BN} = ReLU(x_{BN}) \oplus y_A$$, $\oplus$ means concat

$$x_{2B} = \theta_{2B} z_{BN} + b_{2B}$$

$$z_{2B} = Softmax(x_{2B})$$

$$y_B = z_{2B} = softmax(x_{2B})$$

back propagation

$\hat{y}$ means true label

$\hat{y}_B$ means true label of B

$\hat{y}_A$ mean true label of A

$$\frac{\partial L}{\partial x_{2B}} = \frac{\partial L}{\partial \hat{y}_B} \cdot \frac{\partial \hat{y}_B}{\partial x_{2B}}$$

$$= \begin{cases} -\hat{y}_{jB} \cdot (1 - y_{jB}), & i = j \\ \sum_{i \neq j} \hat{y}_{iB} y_{jB}, & i \neq j \end{cases}$$

$$= y_B - \hat{y}_B$$

FC_{2B}:

$$\frac{\partial L}{\partial \theta_{2B}} = \frac{\partial L}{\partial x_{2B}} \frac{\partial x_{2B}}{\partial \theta_{2B}} = (y_B - \hat{y}_B)\, z_{BN}$$

$$\frac{\partial L}{\partial b_{2B}} = \frac{\partial L}{\partial x_{2B}} \frac{\partial x_{2B}}{\partial b_{2B}} = (y_B - \hat{y}_B) \cdot 1 = y_B - \hat{y}_B$$

$$\frac{\partial L}{\partial z_{BN}} = \frac{\partial L}{\partial z_{2B}} \cdot \frac{\partial x_{2B}}{\partial z_{1B}} = (y - \hat{y}_B)\, \theta_{2B}$$

BN + FC_{1B}

$$\frac{\partial L}{\partial x_{BN}} = \frac{\partial L}{\partial z_{BN}} \cdot \frac{\partial z_{BN}}{\partial x_{BN}} = (y-\hat{y}_B)\,\theta_{2B} \cdot (x_{BN} > 0)$$

$$\frac{\partial L}{\partial x_{1B}} = \frac{\partial L}{\partial x_{BN}} \cdot \frac{\partial x_{BN}}{\partial x_{1B}} + \frac{\partial L}{\partial x_{BN}} \cdot \frac{\partial x_{BN}}{\partial \mu} \cdot \frac{\partial \mu}{\partial x_{1B}}$$

$$= \frac{\partial L}{\partial BN}\left(1 - \frac{1}{m}\right)$$

$$\frac{\partial L}{\partial \theta_{1B}} = \frac{\partial L}{\partial x_{1B}} \cdot \frac{\partial x_{1B}}{\partial \theta_{1B}} = \frac{\partial L}{\partial x_{1B}}\, x$$

FC_{2A}

$$\frac{\partial L}{\partial z_{2A}} = \frac{\partial L}{\partial y_A} \cdot \frac{\partial y_A}{\partial z_{2A}} + \frac{\partial L}{\partial z_{1B}} \frac{\partial z_{1B}}{\partial y_A}$$

$$= (y - \hat{y}_A) + (y_B - \hat{y}_B)\,\theta_{2B}$$

$$\frac{\partial L}{\partial \theta_{2A}} = \frac{\partial L}{\partial z_{2A}} \cdot \frac{\partial z_{2A}}{\partial \theta_{2A}} = \frac{\partial L}{\partial z_{2A}} \cdot z_{DP}$$

$$\frac{\partial L}{\partial b_{2,A}} = \frac{\partial L}{\partial z_{2A}} \cdot \frac{\partial z_{2A}}{\partial b_{2A}} = \frac{\partial L}{\partial z_{2A}}$$

$$\frac{\partial L}{\partial z_{DP}} = \frac{\partial L}{\partial z_{2A}} \cdot \frac{\partial z_{2A}}{\partial z_{DP}} = \frac{\partial L}{\partial z_{2A}} \cdot \theta_{2A}$$

DP (with activation function):

$$\frac{\partial L}{\partial x_{1A}} = \frac{\partial L}{\partial z_{DP}} \frac{\partial z_{DP}}{\partial z_{1A}} \frac{\partial z_{1A}}{\partial x_{1A}} = \frac{\partial L}{\partial z_{DP}} \cdot \frac{\partial z_{DP}}{\partial z_{1A}} \cos(x_{1A})$$

$$= \frac{\partial L}{\partial z_{DP}} \cdot M \cos(x_{1A})$$

FC_{1A}

$$\frac{\partial L}{\partial \theta_{1A}} = \frac{\partial L}{\partial x_{1A}} \frac{\partial x_{1A}}{\partial \theta_{1A}} = \frac{\partial L}{\partial x_{1A}} x_{1A}$$

$$\frac{\partial L}{\partial b_{1A}} = \frac{\partial L}{\partial x_{1A}} \frac{\partial x_{1A}}{\partial b_{1A}} = \frac{\partial L}{\partial x_{1A}}$$