1 Backward Gradient of Scaled Sign Function

1.1 Forward scaled sign function

Assume we have the original weights $\mathbf{W}$ as

$$\mathbf{W} = [W_1, W_2, ..., W_n] \tag{1}$$

In the forward pass, we need to binarize $\mathbf{W}$ to $\widetilde{\mathbf{W}}$

$$\widetilde{\mathbf{W}} = [\widetilde{W}_1, \widetilde{W}_2, ..., \widetilde{W}_n] \tag{2}$$

Based on the XNOR-Net paper, the binarized weights $\widetilde{\mathbf{W}}$ is calculated by

$$\widetilde{\mathbf{W}} = \alpha \cdot sign(\mathbf{W}) \tag{3}$$

$$\alpha = \frac{1}{n} \cdot ||\mathbf{W}||_{L1} = \frac{1}{n} \cdot \sum_{i=1}^{n} W_i \tag{4}$$

which means, for each $\widetilde{W}_i$, we have

$$\widetilde{W}_i = \alpha \cdot sign(W_i),\ i \in \{1, 2, ..., n\} \tag{5}$$

this function is named as the scaled sign function.

1.2 Original backward gradient of scaled sign function

In the XNOR-Net paper, assuming C is the cost function, the backward gradient of the scaled sign function is given as

$$\frac{\partial C}{\partial W_i} = \frac{\partial C}{\partial \widetilde{W}_i}(\frac{1}{n} + \frac{\partial sign(W_i)}{\partial W_i} \cdot \alpha) \tag{6}$$

which is calculated from

$$\frac{\partial C}{\partial W_i} = \frac{\partial C}{\partial \widetilde{W}_i} \cdot \frac{\partial \widetilde{W}_i}{\partial W_i} \tag{7}$$

The detailed steps are:

$$\begin{aligned}
\frac{\partial C}{\partial W_i} &= \frac{\partial C}{\partial \widetilde{W}_i} \cdot \frac{\partial \widetilde{W}_i}{\partial W_i} \\
&= \frac{\partial C}{\partial \widetilde{W}_i} \cdot \frac{\partial(\alpha \cdot sign(W_i))}{\partial W_i} \\
&= \frac{\partial C}{\partial \widetilde{W}_i} \cdot [sign(W_i) \cdot \frac{\partial \alpha}{\partial W_i} + \alpha \cdot \frac{\partial sign(W_i)}{\partial W_i}] \\
&= \frac{\partial C}{\partial \widetilde{W}_i}[\frac{1}{n} + \frac{\partial sign(W_i)}{\partial W_i} \cdot \alpha]
\end{aligned} \tag{8}$$

1.3 Correct backward gradient

However, the equation

$$\frac{\partial C}{\partial W_i} = \frac{\partial C}{\partial \widetilde{W_i}} \cdot \frac{\partial \widetilde{W_i}}{\partial W_i} \tag{9}$$

is actually inaccurate. The correct equation should be

$$\frac{\partial C}{\partial W_i} = \sum_{j=1}^{n} (\frac{\partial C}{\partial \widetilde{W_j}} \cdot \frac{\partial \widetilde{W_j}}{\partial W_i}) \tag{10}$$

Therefore, the **correct backward gradient** should be

$$\begin{aligned} \frac{\partial C}{\partial W_i} = \frac{1}{n} \cdot sign(W_i) \cdot \sum_{j=1}^{n} [\frac{\partial C}{\partial \widetilde{W_j}} \cdot sign(W_j)] \\ + \frac{\partial C}{\partial \widetilde{W_i}} \cdot \frac{\partial sign(W_i)}{\partial W_i} \cdot \alpha \end{aligned} \tag{11}$$

The detailed steps to get this gradient are

$$\begin{aligned} \frac{\partial C}{\partial W_i} &= \sum_{j=1}^{n} (\frac{\partial C}{\partial \widetilde{W_j}} \cdot \frac{\partial \widetilde{W_j}}{\partial W_i}) \\ &= \sum_{j=1}^{n} [\frac{\partial C}{\partial \widetilde{W_j}} \cdot \frac{\partial (\alpha \cdot sign(W_j))}{\partial W_i}] \\ &= \sum_{j=1}^{n} [\frac{\partial C}{\partial \widetilde{W_j}} \cdot (sign(W_j) \cdot \frac{\partial \alpha}{\partial W_i} + \frac{\partial sign(W_j)}{\partial W_i} \cdot \alpha)] \\ &= \sum_{j=1}^{n} [\frac{\partial C}{\partial \widetilde{W_j}} \cdot sign(W_j) \cdot \frac{\partial \alpha}{\partial W_i}] + \frac{\partial C}{\partial \widetilde{W_i}} \cdot \frac{\partial sign(W_j)}{\partial W_i} \cdot \alpha \\ &= \frac{\partial \alpha}{\partial W_i} \cdot \sum_{j=1}^{n} [\frac{\partial C}{\partial \widetilde{W_j}} \cdot sign(W_j)] + \frac{\partial C}{\partial \widetilde{W_i}} \cdot \frac{\partial sign(W_j)}{\partial W_i} \cdot \alpha \\ &= \frac{1}{n} \cdot sign(W_i) \cdot \sum_{j=1}^{n} [\frac{\partial C}{\partial \widetilde{W_j}} \cdot sign(W_j)] + \frac{\partial C}{\partial \widetilde{W_i}} \cdot \frac{\partial sign(W_j)}{\partial W_i} \cdot \alpha \end{aligned} \tag{12}$$