

1 Backward Gradient of Scaled Sign Function

1.1 Forward scaled sign function

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

$$\mathbf{W} = [W_1, W_2, \dots, W_n] \quad (1)$$

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

$$\widetilde{\mathbf{W}} = [\widetilde{W}_1, \widetilde{W}_2, \dots, \widetilde{W}_n] \quad (2)$$

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

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

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

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

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

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

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

which is calculated from

$$\frac{\partial C}{\partial W_i} = \frac{\partial C}{\partial \widetilde{W}_i} \cdot \frac{\partial \widetilde{W}_i}{\partial W_i} \quad (7)$$

The detailed steps are:

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

1.3 Correct backward gradient

However, the equation

$$\frac{\partial C}{\partial W_i} = \frac{\partial C}{\partial \widetilde{W}_i} \cdot \frac{\partial \widetilde{W}_i}{\partial W_i} \quad (9)$$

is actually inaccurate. The correct equation should be

$$\frac{\partial C}{\partial W_i} = \sum_{j=1}^n \left(\frac{\partial C}{\partial \widetilde{W}_j} \cdot \frac{\partial \widetilde{W}_j}{\partial W_i} \right) \quad (10)$$

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

$$\begin{aligned} \frac{\partial C}{\partial W_i} &= \frac{1}{n} \cdot \text{sign}(W_i) \cdot \sum_{j=1}^n \left[\frac{\partial C}{\partial \widetilde{W}_j} \cdot \text{sign}(W_j) \right] \\ &\quad + \frac{\partial C}{\partial \widetilde{W}_i} \cdot \frac{\partial \text{sign}(W_i)}{\partial W_i} \cdot \alpha \end{aligned} \quad (11)$$

The detailed steps to get this gradient are

$$\begin{aligned} \frac{\partial C}{\partial W_i} &= \sum_{j=1}^n \left(\frac{\partial C}{\partial \widetilde{W}_j} \cdot \frac{\partial \widetilde{W}_j}{\partial W_i} \right) \\ &= \sum_{j=1}^n \left[\frac{\partial C}{\partial \widetilde{W}_j} \cdot \frac{\partial (\alpha \cdot \text{sign}(W_j))}{\partial W_i} \right] \\ &= \sum_{j=1}^n \left[\frac{\partial C}{\partial \widetilde{W}_j} \cdot \left(\text{sign}(W_j) \cdot \frac{\partial \alpha}{\partial W_i} + \frac{\partial \text{sign}(W_j)}{\partial W_i} \cdot \alpha \right) \right] \\ &= \sum_{j=1}^n \left[\frac{\partial C}{\partial \widetilde{W}_j} \cdot \text{sign}(W_j) \cdot \frac{\partial \alpha}{\partial W_i} \right] + \frac{\partial C}{\partial \widetilde{W}_i} \cdot \frac{\partial \text{sign}(W_i)}{\partial W_i} \cdot \alpha \\ &= \frac{\partial \alpha}{\partial W_i} \cdot \sum_{j=1}^n \left[\frac{\partial C}{\partial \widetilde{W}_j} \cdot \text{sign}(W_j) \right] + \frac{\partial C}{\partial \widetilde{W}_i} \cdot \frac{\partial \text{sign}(W_i)}{\partial W_i} \cdot \alpha \\ &= \frac{1}{n} \cdot \text{sign}(W_i) \cdot \sum_{j=1}^n \left[\frac{\partial C}{\partial \widetilde{W}_j} \cdot \text{sign}(W_j) \right] + \frac{\partial C}{\partial \widetilde{W}_i} \cdot \frac{\partial \text{sign}(W_i)}{\partial W_i} \cdot \alpha \end{aligned} \quad (12)$$