

$$\underset{(m,)}{y} = \underset{(m,n)}{A} \underset{(n)}{x} + \underset{(m)}{b} \quad Z = \text{ReLU}(y)$$

$$\hat{y} \times y_{\text{scale}} = A_q A_{\text{scale}} \hat{x} x_{\text{scale}} + b$$

Define $\hat{y} = y_q - y_{\text{zero}}$, $\hat{x} = x_q - x_{\text{zero}}$
 $\hat{z} = z_q - z_{\text{zero}}$

$$\hat{y} = \frac{A_{\text{scale}} x_{\text{scale}}}{y_{\text{scale}}} \cdot A_q \hat{x} + \frac{b}{y_{\text{scale}}}$$

Given a constant S .

① Approximate: $\frac{A_{\text{scale}} x_{\text{scale}}}{y_{\text{scale}}} \approx u \times 2^{-S}$
 $\frac{b}{y_{\text{scale}}} \approx v \times 2^{-S}$

Define $\tilde{A} = \underset{(m,n)}{A_q} \cdot \underset{(m)}{u}$
 $\underset{(m,)}{\hat{y}} = \left(\underset{(m,n)}{\tilde{A}} \underset{(n)}{\hat{x}} + \underset{(m)}{v} \right) \times \underset{\substack{\uparrow \\ \text{constant}}}{2^{-S}}$

$A_{\text{scale}}: (m,)$
 $x_{\text{scale}}: ()$
 $y_{\text{scale}}: ()$
 $u: (m,)$
 $b: (m,)$
 $v: (m,)$

Because \hat{y} is already subtracted by y_{zero}

$$\hat{z} = \text{ReLU}(\hat{y}) = \text{ReLU} \left[\left(\tilde{A} \hat{x} + v \right) \times 2^{-S} \right]$$