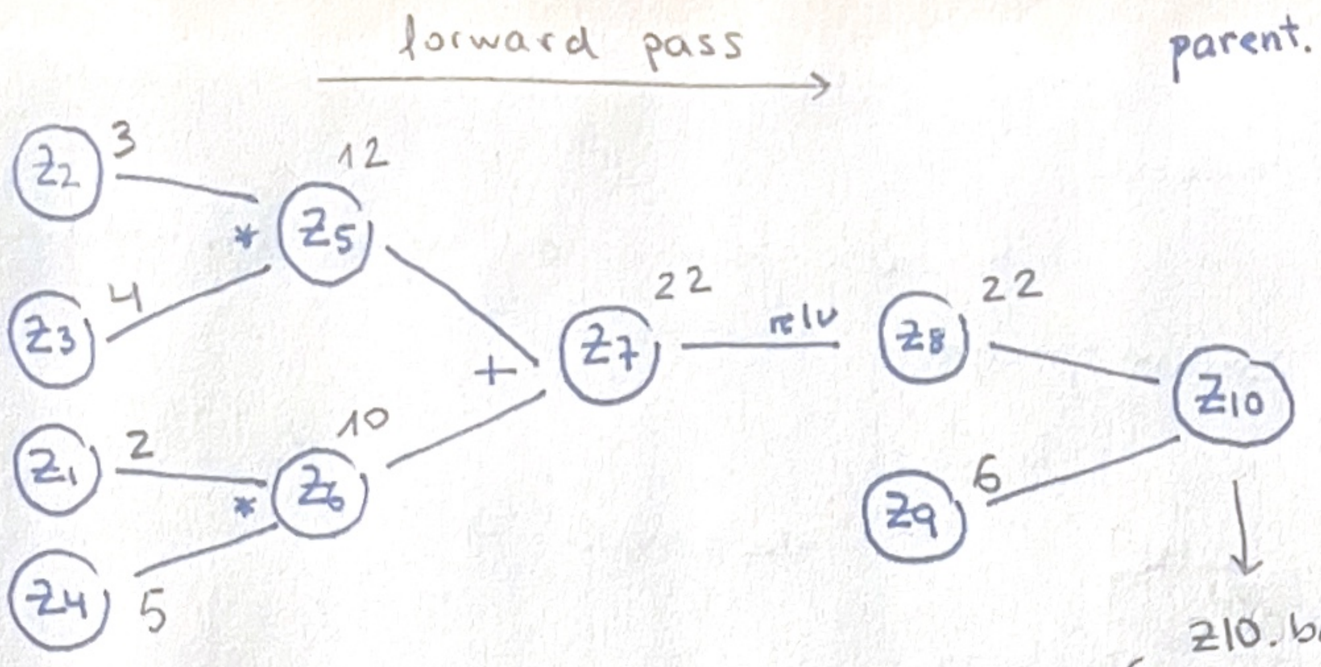

parent.grad += child.grad * local_deriv

z10.backward()

↳ "I want to derivate of z10 with respect to every leaf parameter."

0th Iterations

$z10.grad = 1 \quad \left(\frac{\partial z10}{\partial z10} = 1\right)$

↳ "The output changes at same rate as itself."

$z10 = z8 * z9$

↓ ↳ w0

relu()

$\frac{\partial z10}{\partial z8} = z9 = 6 \qquad \frac{\partial z10}{\partial z9} = z8 = 22$

$z8.grad += z10.grad * \frac{\partial z10}{\partial z8} = 1 \cdot z9 = 6$ // $z9.grad += z10.grad * \frac{\partial z10}{\partial z9}$

$z9.grad = 1 \cdot z8 = 22$

Since z9 is a leaf, the final gradient is $z9.grad = 22$

↓

$w0.grad = 22$

1-st iteration

$z8.grad = 6 \qquad z8 = z7 \cdot relu(z5 + z6)$

$\frac{\partial z8}{\partial z7} = 1$ if $z7 > 0 \qquad z7.grad += z8.grad * \frac{\partial z8}{\partial z7} = 6 \cdot 1 = 6$

2nd iteration

$z7.grad = 6 \qquad z7 = z5 + z6$

$\frac{\partial z7}{\partial z5} = 1 \qquad \frac{\partial z7}{\partial z6} = 1$

$z6.grad += z7.grad * \frac{\partial z7}{\partial z6} = 6 \cdot 1 = 6$

$z5.grad += z7.grad * \frac{\partial z7}{\partial z5} = 6 \cdot 1 = 6$

3ra iteration

$z6.grad = 6 \qquad z6 = z1 * z4$

$\frac{\partial z6}{\partial z4} = z1 \qquad z4.grad += 6 \cdot z1$

$z4.grad = 12$

$\frac{\partial z6}{\partial z1} = z4 \qquad z1.grad = 6 \cdot z4$

$z1.grad = 30$

$w1.grad = 12$

4th iteration

$z5.grad = 6 \qquad z5 = z2 * z3$

$\frac{\partial z5}{\partial z2} = z3 \qquad z2.grad += 6 \cdot 4 = 24$

$\frac{\partial z5}{\partial z3} = z2 \qquad z3.grad += 6 \cdot 3 = 18$

$w2.grad = 18$