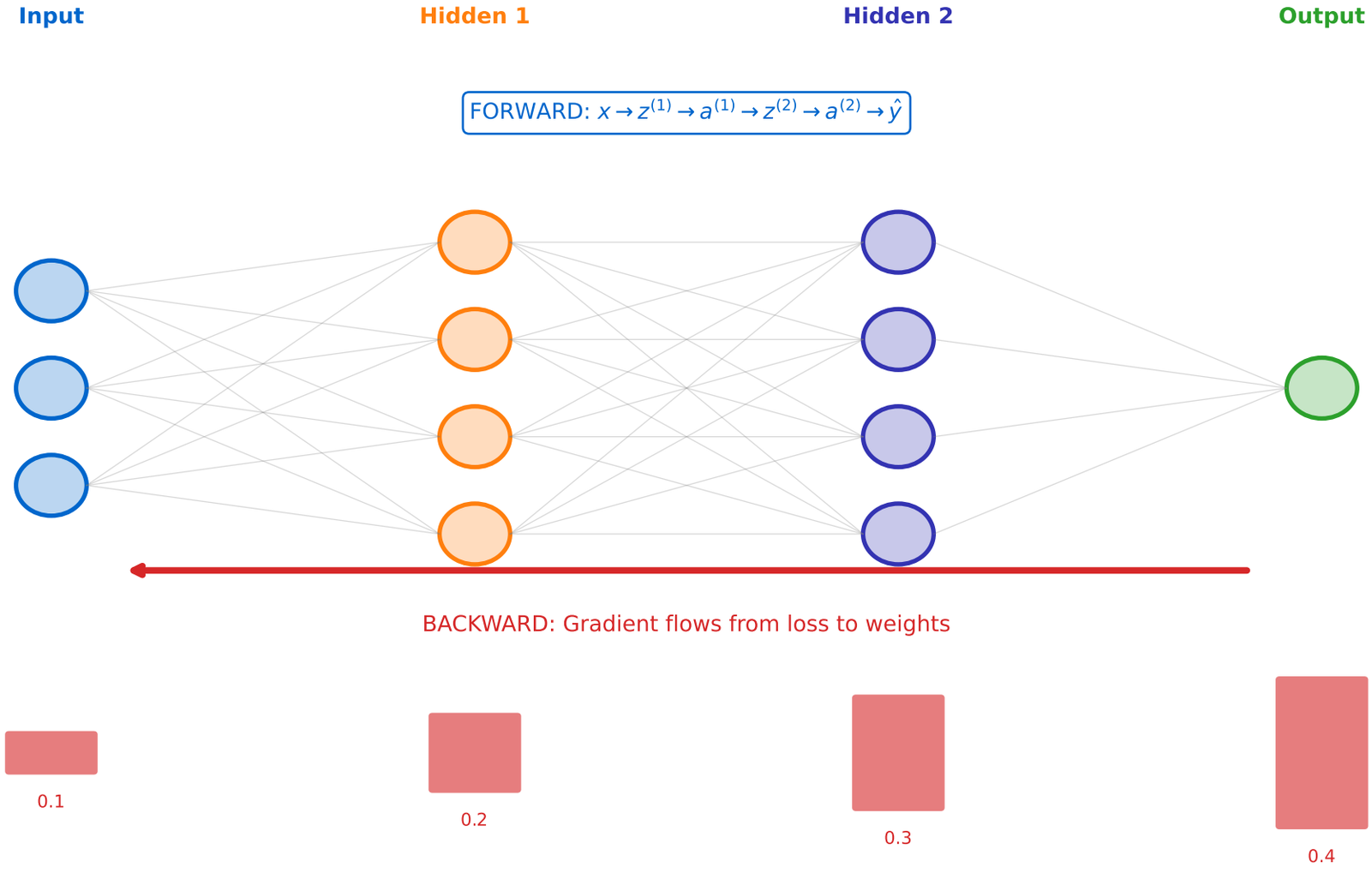
Gradient Flow Through a 3-Layer MLP



Backward Pass Equations:

$\delta^{(L)} = \nabla_a L \odot \sigma'(z^{(L)})$

$\delta^{(l)} = (W^{(l+1)})^T \delta^{(l+1)} \odot \sigma'(z^{(l)})$

$\frac{\partial L}{\partial W^{(l)}} = \delta^{(l)} (a^{(l-1)})^T$

$\frac{\partial L}{\partial b^{(l)}} = \delta^{(l)}$

Gradient magnitude decreases as it flows backward (vanishing gradient problem)