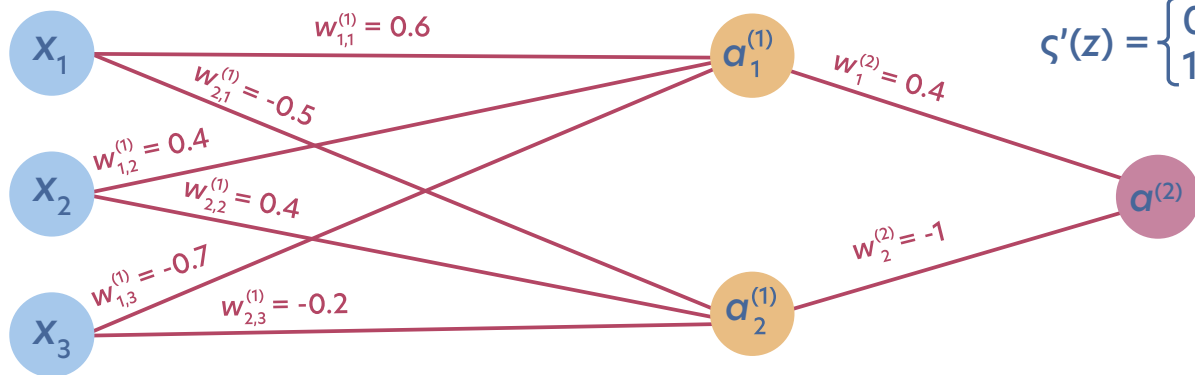


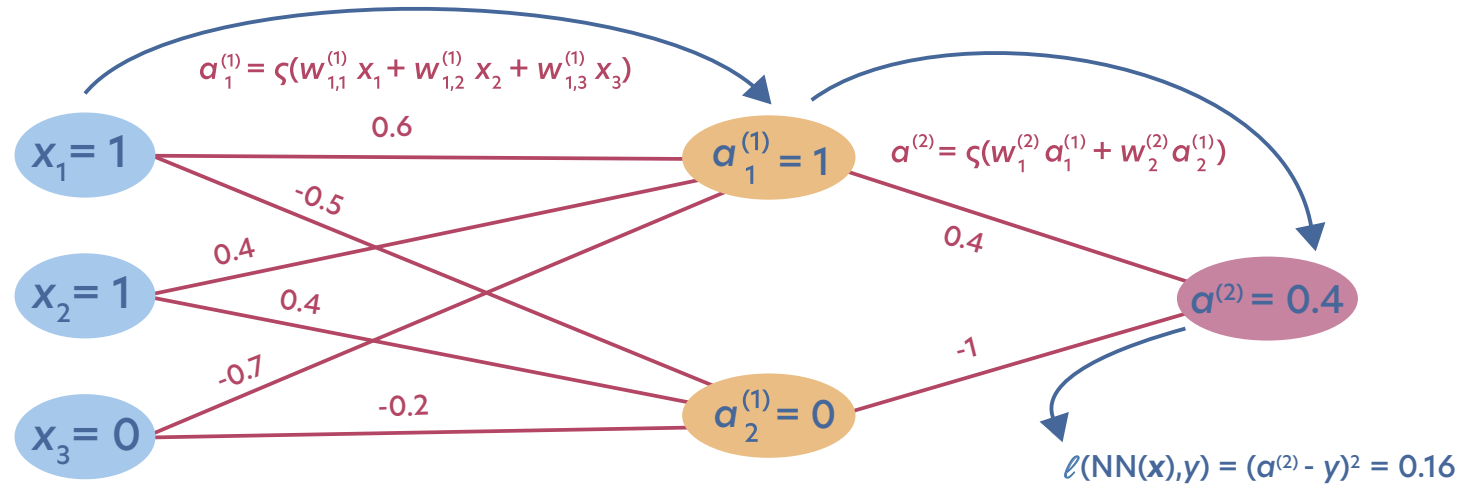
(a) two-layer NN with ReLUs and MSE loss



$$\zeta^{(l)}(z) = \zeta(z) = \text{ReLU}(z)$$

$$\zeta'(z) = \begin{cases} 0, & z \leq 0 \\ 1, & z > 0 \end{cases}$$

(b) forward pass for $(x,y) = ([1,1,0], 0)$ caching intermediate results



(c) backward pass for $(x,y) = ([1,1,0], 0)$ to compute gradients taking advantage of cached results from the forward pass

