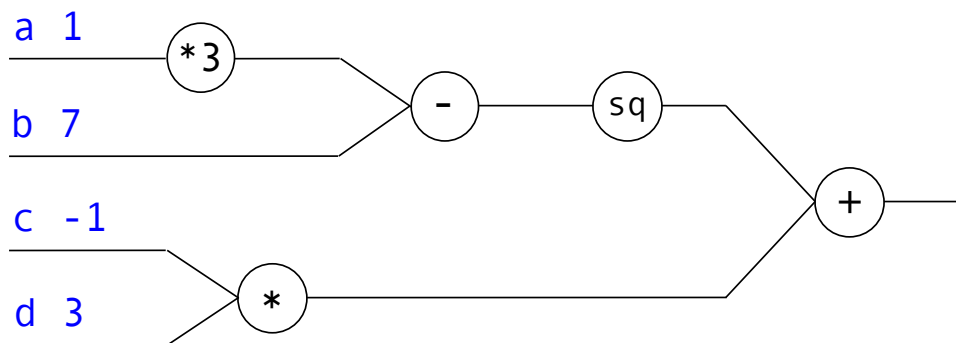


Backpropagation practice

(outside of class)

1. *Example 1.* Let $f(a, b, c, d) = (3a - b)^2 + cd$, a function of four inputs. We could decompose the function into a neural network diagram in many ways, but one way is shown below. Using the given values, first run the “forward pass” to compute the output value of each node (write above the lines). Then use backpropagation to compute the derivatives (write below the lines). If we wanted to *minimize* f with respect to these inputs and we have a learning rate of $\alpha = 0.1$, what values of a, b, c, d would we choose for the next step?



2. *Example 2.* Let $\sigma(z) = \frac{1}{1+e^{-z}}$ be the sigmoid function, and let $h(\vec{x}) = \sigma(w_1x_1 + w_2x_2 + b)$ be our hypothesis about the label (0 or 1) of a given input vector $\vec{x} = (x_1, x_2)$ with two features. Let our loss function be the cross entropy of our prediction relative to the truth, i.e. $\ell_y(h) = -y \log h - (1 - y) \log(1 - h)$. If we are given fixed values $\vec{x} = (1, 3)$ and label $y = 0$, what is the value of the loss function (given the starting weights below)? Again use backpropagation to compute the gradients. Use the fact that $\sigma'(z) = \sigma(z)(1 - \sigma(z))$. Assuming we want to minimize the loss and $\alpha = 0.1$, what are the new values of w_1, w_2, b ?

