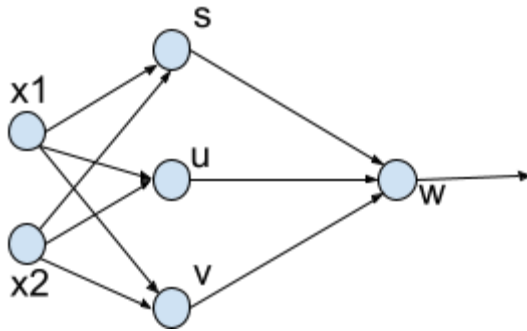


Consider the network below. How do we optimize the model parameters w , s , u , and v ?



We will use the back-propagation method which is to optimize the final layer nodes first and then work your way into the earlier layers of the network. So in the above network we would update the weights of w , then we do s , u , and v , and we repeat until convergence.

Back propagation algorithm for the above network

1. Initialize all weights to random values drawn from $[-.01, .01]$ (or $[0,1]$ is fine for now)
2. Set $prev_obj = infinity$
3. Calculate objective $obj = \sum_{i=0}^{n-1} ((w_1, w_2, w_3)^T (\sigma(s^T x_i), \sigma(u^T x_i), \sigma(v^T x_i)) - y)^2$
4. While $(prev_obj - obj > 0.01)$ do
 - a. $prev_obj = obj$
 - b. Update weights of w : $w = w - \eta df/dw$
 - c. Update weights of s , u , and v : $s = s - \eta df/ds$, $u = u - \eta df/du$,
 $v = v - \eta df/dv$
 - d. Calculate new objective value obj