



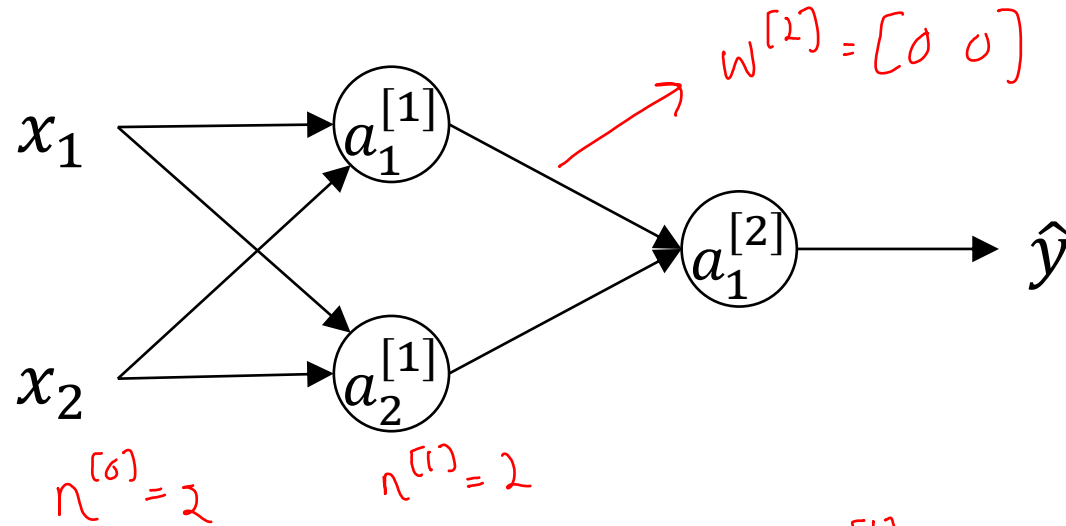
deeplearning.ai

One hidden layer
Neural Network

Random Initialization

What happens if you initialize weights to zero?

Gradient descent won't work!



$$n^{[0]} = 2$$

$$n^{[1]} = 2$$

\mathbb{R}

$$w^{[1]} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad b^{[1]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

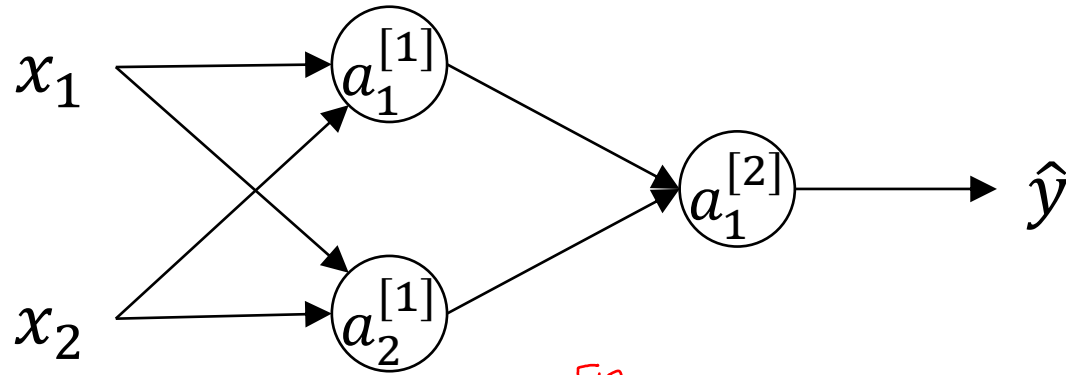
then $a_1^{[1]} = a_2^{[1]}$

$$\Rightarrow z_1^{[1]} = z_2^{[1]}$$

\Rightarrow Hidden units are computing the same values in every iteration

\Rightarrow No point having more than 1 hidden unit

Random initialization

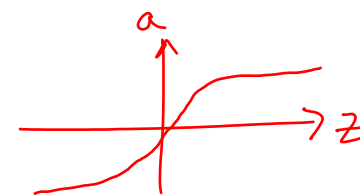


$$W^{[1]} = \text{np.random.randn}(2, 2) * 0.01$$

$$b^{[1]} = \text{np.zeros}(2, 1) \rightarrow \text{this can be zeros, no issues}$$

Same for
 $W^{[2]}, b^{[2]}$

why multiply by 0.01? \rightarrow



$$z^{[1]} = W^{[1]} \cdot x^T + b^{[1]}$$

$$a^{[1]} = g^{[1]}(z^{[1]})$$

If w is large $\Rightarrow z$ is large
 \Rightarrow learning is slow
as gradient is small
 \Rightarrow update to weight
is small