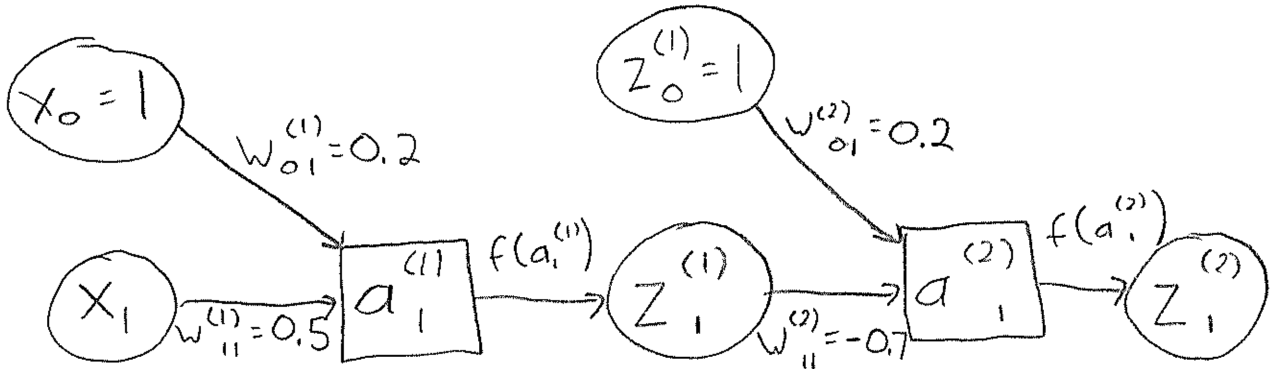
Network with initial weights:



$f(a) = \frac{1}{1+e^{-a}}$ (sigmoid) and $\text{Error} = \frac{1}{n}\sum_{e\in E}\left(Y(e) - Z_1^{(2)}(e)\right)^2$ (mean squared error)

Suppose we want to do backprop after receiving one training example with feature $X_1 = 0.8$ and $Y = 1.0$

Forward Pass:

$$a_1^{(1)} = \sum_i x_i W_{i1} = x_0 W_{01}^{(1)} + x_1 W_{11}^{(1)} = 1(0.2) + 0.8(0.5) = 0.6$$

$$Z_1^{(1)} = f(a_1^{(1)}) = f(0.6) = \frac{1}{1+e^{-0.6}} \approx 0.646$$

$$a_1^{(2)} = \sum_j z_j^{(1)} W_{j1}^{(2)} = z_0^{(1)} W_{01}^{(2)} + z_1^{(1)} W_{11}^{(2)} = 1(0.2) + 0.646(-0.7) \approx -0.252$$

$$\hat{Y} = Z_1^{(2)} = f(a_1^{(2)}) = f(-0.252) = \frac{1}{1+e^{0.252}} \approx 0.437$$

$$\text{Error} = \frac{1}{n}\sum_{e\in E}\left(Y(e) - Z_1^{(2)}(e)\right)^2 = \frac{1}{1}\left(1 - 0.437\right)^2 \approx 0.317$$

^We only have one example, so $n=1$ and E only has one item

Backward Pass:

We want to update all of our weights by:

$$W_{ij}^{(\ell)} \leftarrow W_{ij}^{(\ell)} - \eta \frac{\delta \text{Error}(e)}{\delta W_{ij}^{(\ell)}}$$

We will choose $\eta = 0.1$

For our weights leading to the output layer, we have (from Lecture 9b):

$$\frac{\delta Error}{\delta W_{j1}^{(2)}} = \delta_1^{(2)} z_j^{(1)}, \quad \text{where} \quad \delta_1^{(2)} = \frac{\delta Error}{\delta z_1^{(2)}} f'(a_1^{(2)})$$

For our Error and activation functions, we have derivatives:

$$\frac{\delta Error}{\delta z_1^{(2)}} = -\frac{2}{n}\sum_{e \in E}\left(Y(e) - z_1^{(2)}(e)\right) = -2\left(Y(e) - z_1^{(2)}(e)\right), \quad \text{and}$$

↖Again, we have only one example

$$f'(a) = f(a)(1 - f(a))$$

So,

$$\delta_1^{(2)} = -2(Y - z_1^{(2)}) f(a_1^{(2)})(1 - f(a_1^{(2)})) = -2(1 - 0.437)(0.437)(1 - 0.437)$$

$$\approx -0.277$$

Then,

$$W_{01}^{(2)} \leftarrow W_{01}^{(2)} - \eta \frac{\delta Error}{\delta W_{01}^{(2)}} = W_{01}^{(2)} - \eta\left(\delta_1^{(2)} z_0^{(1)}\right) = 0.2 - 0.1(-0.277)(1) \approx 0.22$$

$$W_{11}^{(2)} \leftarrow W_{11}^{(2)} - \eta \frac{\delta Error}{\delta W_{11}^{(2)}} = W_{11}^{(2)} - \eta\left(\delta_1^{(2)} z_1^{(1)}\right) = -0.7 - 0.1(-0.277)(0.646) \approx -0.68$$

For our weights at the input layer, we have:

$$\frac{\delta Error}{\delta W_{i1}^{(1)}} = \delta_1^{(1)} x_i, \quad \text{where} \quad \delta_1^{(1)} = \delta_1^{(2)} W_{11}^{(2)} f'(a_1^{(1)})$$

↖No summation because we only have one output node

$$= \delta_1^{(2)} W_{11}^{(2)} f(a_1^{(1)})(1 - f(a_1^{(1)}))$$

$$= (-0.277)(-0.7)(0.646)(1 - 0.646) \approx 0.044$$

Then,

$$W_{01}^{(1)} \leftarrow W_{01}^{(1)} - \eta \frac{\delta Error}{\delta W_{01}^{(1)}} = W_{01}^{(1)} - \eta \delta_1^{(1)} x_0 = 0.2 - 0.1(0.044)(1) \approx 0.196$$

$$W_{11}^{(1)} \leftarrow W_{11}^{(1)} - \eta \frac{\delta Error}{\delta W_{11}^{(1)}} = W_{11}^{(1)} - \eta \delta_1^{(1)} x_1 = 0.5 - 0.1(0.044)(0.8) \approx 0.496$$

Forward Pass (with new weights)

$a_1^{(1)} = 1 \times 0.196 + 0.8 \times 0.496 = 0.593$

$z_1^{(1)} = f(a_1^{(1)}) = 0.644$

$a_1^{(2)} = 1 \times .228 + .644 \times (-0.682) = -0.211$

$\hat{y} = 0.447 > 0.437$

$Error = 0.305 < 0.317$

---

try full pass again $\bar{w}$ original weights but $\eta = 0.5$

$W_{01}^{(2)} = 0.2 - 0.5(-0.277) = 0.339$

$W_{11}^{(2)} = -0.7 - 0.5(-0.277)(0.646) = -0.61$

$W_{01}^{(1)} = +0.2 - 0.5(0.044)(1) = 0.178$

$W_{11}^{(1)} = 0.5 - 0.5(0.044)(0.8) = 0.482$

forward pass $\bar{w}$ new weights

$a_1^{(1)} = 1 \times .178 + 0.8 \times 0.482 = 0.564$

$z_1^{(1)} = 0.637$

$a_1^{(2)} = 1 \times .339 + .637 \times (-0.61) = -0.05$

$\hat{y} = 0.488$

$Error = 0.263$