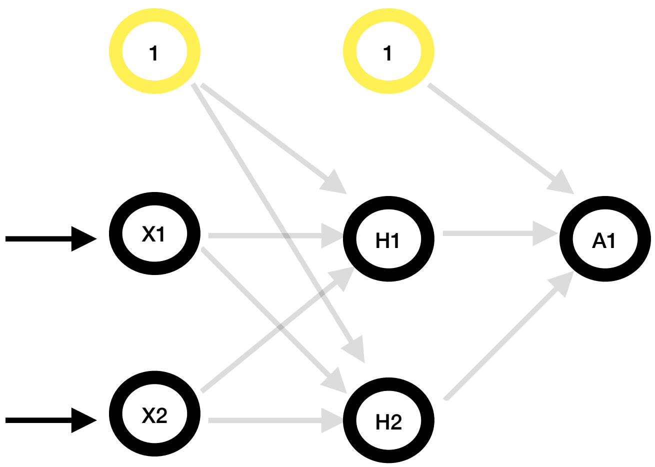
Neural Network

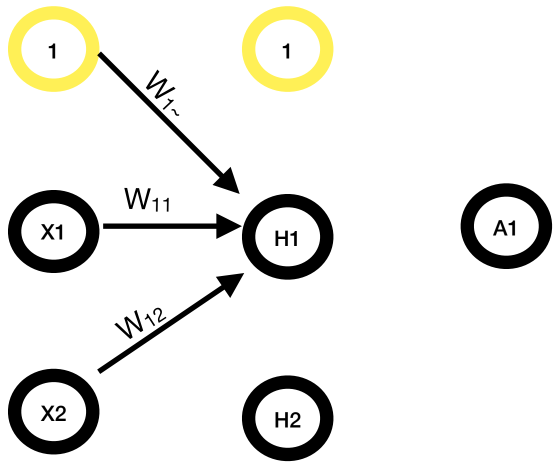



*Simple neural network with an input layer, hidden layer, and output layer*

**Input Layer**:
- Single input or batch
- 1 bias neuron
- 2 TLU(Threshold Logic Unit) neuron
- No weights involved as inputs are raw values fed directly to neuron

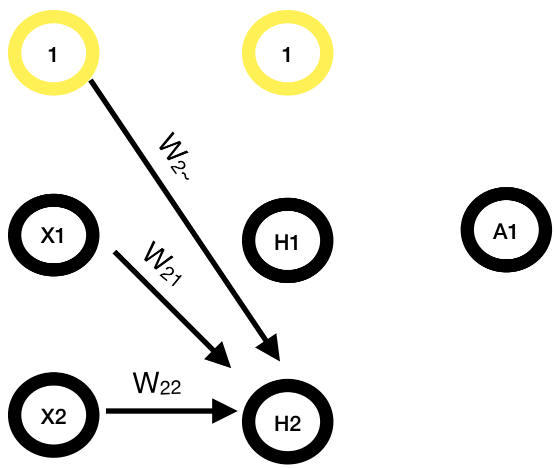
**Hidden Layer**:
- 1 bias neuron
- 2 TLU neurons
- 6 weights ( 3 inputs, 2 neurons)

**Output Layer**:
- 1 TLU neuron
- 3 weights (3 inputs)



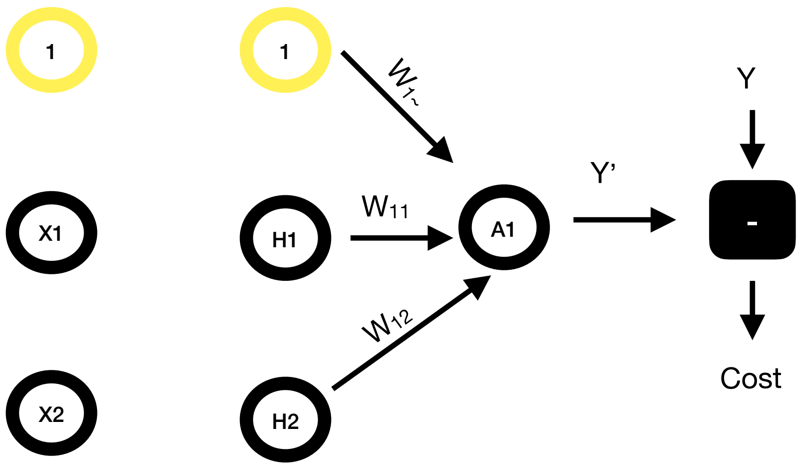

$W_{1b}$ - weight of path to H1 from bias
$W_{11}$ - weight of path to H1 from X1
$W_{12}$ - weight of path to H1 from X2

$$H1 = W_{1\sim_1} * 1 \; + \; X1 * W_{11_1} \; + X2 * W_{12_1}$$




$W_{2\sim}$ - weight of path to H2 from bias
$W_{21}$ - weight of path to H2 from X1
$W_{22}$ - weight of path to H2 from X2
$\sigma$ - neuron activation function

$$H2 = \sigma(W_{2\sim_1} * 1 \; + \; X1 * W_{21_1} \; + X2 * W_{22_1})$$



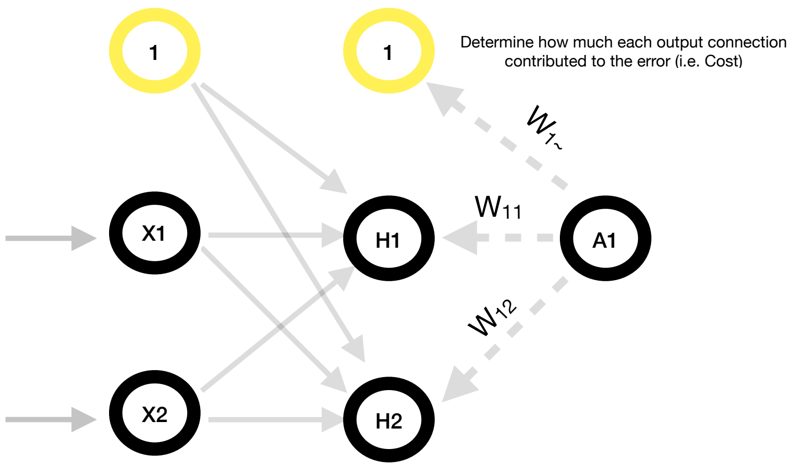

$W_{1\sim}$ - weight of path to A1 from bias
$W_{11}$ - weight of path to A1 from H1
$W_{12}$ - weight of path to A1 from H2
$\sigma$ - neuron activation function

$$A1 = \sigma(W_{1\sim_2} * 1 \;+\; H1 * W_{11_2} \;+ H2 * W_{12_2})$$

$$Cost = (y_i' - y_i)^2$$

$$Avg\ Cost = \frac{1}{N}\sum_{i=1}^{N}(y_i' - y_i)^2 \quad for\ batch$$

Back-propagation-




Partial Derivative of cost with respect to W1~
Contribution of W1~ to error

Chain Rule used to chain function dependent variables: partial derivates are chained

$$\frac{\partial C}{\partial W_{1\sim_2}} = \frac{\partial C}{\partial A1}\frac{\partial A1}{\partial W_{1\sim_2}}$$

Partial Derivative cost with respect to W11
Contribution of W11 to error

$$\frac{\partial C}{\partial W_{11_2}} = \frac{\partial C}{\partial A1}\frac{\partial A1}{\partial W_{11_2}}$$

Partial Derivative cost with respect to W12
Contribution of W12 to error

$$\frac{\partial C}{\partial W_{12_2}} = \frac{\partial C}{\partial A1}\frac{\partial A1}{\partial W_{12_2}}$$

System is fixed after a forward propagation pass. How do we measure how all the weights affect the cost in the network? Small changes in weight (partial derivatives) will affect the cost; we can measure this change and measure error contribution to the cost. The goal is to remove this influence thus reducing the cost.

$$\frac{\partial C}{\partial W_{1\sim_2}} = \frac{\partial \mathbf{C}}{\partial \mathbf{A1}} \times \frac{\partial A1}{\partial W_{1\sim_2}}$$

$$\frac{\partial \mathbf{C}}{\partial \mathbf{A1}} = \frac{\partial C}{\partial u} \times \frac{\partial u}{\partial A1}$$

$$A1 = \sigma(W_{1\sim_2} \times 1 \ + \ H1 \times W_{11_2} \ + H2 \times W_{12_2})$$

$$Cost = (y_i' - y_i)^2$$
$$Cost = (A1 - y_i)^2$$
$$C = (u)^2$$
$$u = A1 - y$$

$$\frac{\partial C}{\partial u} = 2 \times u$$

$$\frac{\partial u}{\partial A1} = 1$$

$$\frac{\partial \mathbf{C}}{\partial \mathbf{A1}} = 2 \times u$$

$$\frac{\partial C}{\partial W_{1\sim 2}} = \frac{\partial C}{\partial A1} \times \frac{\partial \mathbf{A1}}{\partial \mathbf{W_{1\sim 2}}}$$

$$\frac{\partial \mathbf{A1}}{\partial \mathbf{W_{1\sim 2}}} = \frac{\partial A1}{\partial u} \times \frac{\partial u}{\partial W_{1\sim 2}}$$

$$A1 = \sigma(W_{1\sim 2} \times 1 \ + \ H1 \times W_{11_2} \ + H2 \times W_{12_2})$$

$$A1 = \sigma(u)$$

$$u = W_{1\sim 2} \times 1 \ + \ H1 \times W_{11_2} \ + H2 \times W_{12_2}$$

$$\frac{\partial A1}{\partial W_{1\sim 2}} = \frac{\partial A1}{\partial u} \times \frac{\partial u}{\partial W_{1\sim 2}}$$

$$\frac{\partial A1}{\partial u} = \frac{\partial \sigma(u)}{\partial u} = \sigma'(u)$$

$$\frac{\partial A1}{\partial W_{1\sim_2}} = \frac{\partial A1}{\partial u} \times \frac{\partial u}{\partial W_{1\sim_2}}$$

$$\frac{\partial u}{\partial W_{1\sim}} = 1$$

$$\frac{\partial \mathbf{A1}}{\partial \mathbf{W_{1\sim_2}}} = \frac{\partial A1}{\partial u} \times \frac{\partial u}{\partial W_{1\sim_2}} = \sigma'(u) \times 1$$

$$\frac{\partial \mathbf{C}}{\partial \mathbf{A1}} = 2 \times u$$

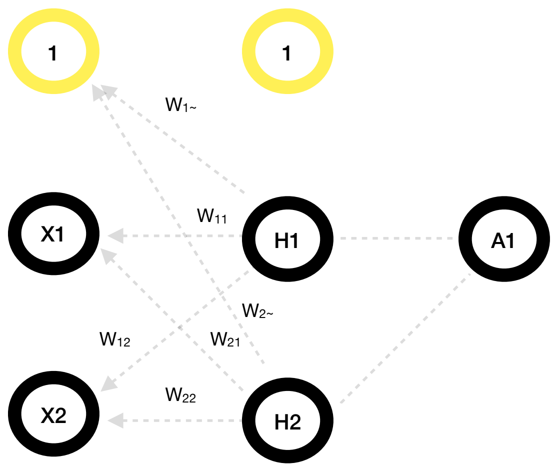
$$\frac{\partial \mathbf{C}}{\partial \mathbf{W_{1\sim_2}}} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial W_{1\sim_2}} = 2 \times u \times \sigma'(u)$$

Similar for other paths connected to output neuron

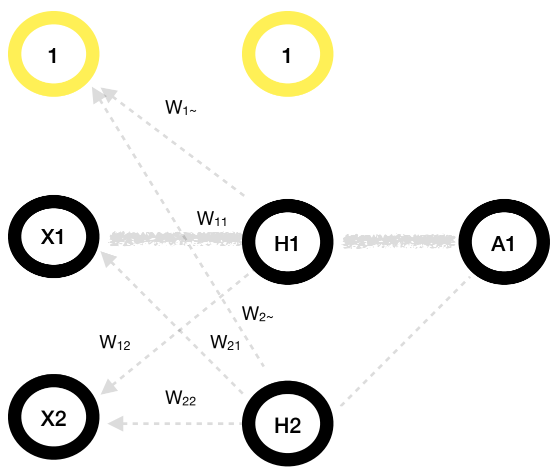
$$\frac{\partial \mathbf{C}}{\partial \mathbf{W_{1\sim_2}}} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial W_{1\sim_2}} = 2 \times u \times \sigma'(u)$$

$$\frac{\partial \mathbf{C}}{\partial \mathbf{W_{11_2}}} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial W_{11_2}} = 2 \times u \times \sigma'(u) \times H1$$

$$\frac{\partial \mathbf{C}}{\partial \mathbf{W_{12_2}}} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial W_{12_2}} = 2 \times u \times \sigma'(u) \times H2$$




Determine how much each path to input neurons from output neuron contributed to the error




Path of interest. A1 neuron to X1.What influence does $W_{11}$(layer 1) have on the cost function. Chain Rule will allow to traverse the path and measure the influence to the cost

$$\frac{\partial C}{\partial W_{11_1}} = \frac{\partial C}{\partial H1} \times \frac{\partial H1}{\partial W_{11_1}}$$

Expand each operand using chain rule

1st operand

$$\frac{\partial C}{\partial H1} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial u_{A1}} \times \frac{\partial u_{A1}}{\partial H1}$$

$$\frac{\partial C}{\partial A1} = 2 \times u_{cost}$$

$$Cost = (y_i' - y_i)^2$$

$$Cost = (A1 - y_i)^2$$

$$C = (u_{cost})^2$$

$$u_{cost} = A1 - y$$

$$\frac{\partial A1}{\partial u_{A1}} = \frac{\partial \sigma(u_{A1})}{\partial u_{A1}} = \sigma'(u_{A1})$$

$$\frac{\partial u_{A1}}{\partial H1} = W_{11_2}$$

$$A1 = \sigma(W_{1\sim_2} \times 1 \ + \ H1 \times W_{11_2} \ + H2 \times W_{12_2})$$

$$A1 = \sigma(u_{A1})$$

$$u_{A1} = W_{1\sim_2} \times 1 \ + \ H1 \times W_{11_2} \ + H2 \times W_{12_2}$$

2nd operand

$$\frac{\partial H1}{\partial W_{11_1}} = \frac{\partial H1}{\partial u_{H1}} \times \frac{\partial u_{H1}}{\partial W_{11_1}}$$

$$\frac{\partial H1}{\partial u_{H1}} = \frac{\partial \sigma(u_{H1})}{\partial u_{H1}} = \sigma'(u_{H1})$$

$$\frac{\partial u_{H1}}{\partial W_{11}} = X1$$

$$H1 = \sigma(W_{1\sim_1} \times 1 \;+\; X1 \times W_{11_1} \;+ X2 \times W_{12_1})$$

$$H1 = \sigma(u_{H1})$$

$$u_{H1} = W_{1\sim_1} \times 1 \;+\; X1 \times W_{11_1} \;+ X2 \times W_{12_1}$$

$$\frac{\partial C}{\partial W_{11_1}} = \frac{\partial C}{\partial H1} \times \frac{\partial H1}{\partial W_{11_1}}$$

$$\frac{\partial C}{\partial H1} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial u_{A1}} \times \frac{\partial u_{A1}}{\partial H1}$$

$$\frac{\partial H1}{\partial W_{11_1}} = \frac{\partial H1}{\partial u_{H1}} \times \frac{\partial u_{H1}}{\partial W_{11_1}}$$

$$\frac{\partial C}{\partial W_{11_1}} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial u_{A1}} \times \frac{\partial u_{A1}}{\partial H1} \times \frac{\partial H1}{\partial u_{H1}} \times \frac{\partial u_{H1}}{\partial W_{11_1}} = 2 \times u_{cost} \times \sigma'(u_{A1}) \times W_{11_2} \times \sigma'(u_{H1}) \times X1$$

Remaining gradients

$$\frac{\partial C}{\partial W_{1\sim_1}} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial u_{A1}} \times \frac{\partial u_{A1}}{\partial H1} \times \frac{\partial H1}{\partial u_{H1}} \times \frac{\partial u_{H1}}{\partial W_{1\sim_1}} = 2 \times u_{cost} \times \sigma'(u_{A1}) \times W_{11_2} \times \sigma'(u_{H1})$$

$$\frac{\partial C}{\partial W_{12_1}} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial u_{A1}} \times \frac{\partial u_{A1}}{\partial H1} \times \frac{\partial H1}{\partial u_{H1}} \times \frac{\partial u_{H1}}{\partial W_{12_1}} = 2 \times u_{cost} \times \sigma'(u_{A1}) \times W_{11_2} \times \sigma'(u_{H1}) \times X2$$

$$H2 = \sigma(W_{2\sim_1} \times 1 \ + \ X1 \times W_{21_1} \ + X2 \times W_{22_1})$$

$$H2 = \sigma(u_{H2})$$

$$u_{H2} = W_{2\sim_1} \times 1 \ + \ X1 \times W_{21_1} \ + X2 \times W_{22_1}$$

$$\frac{\partial C}{\partial W_{2\sim_1}} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial u_{A1}} \times \frac{\partial u_{A1}}{\partial H2} \times \frac{\partial H2}{\partial u_{H2}} \times \frac{\partial u_{H2}}{\partial W_{2\sim_1}} = 2 \times u_{cost} \times \sigma'(u_{A1}) \times W_{12_2} \times \sigma'(u_{H2})$$

$$\frac{\partial C}{\partial W_{21_1}} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial u_{A1}} \times \frac{\partial u_{A1}}{\partial H2} \times \frac{\partial H2}{\partial u_{H2}} \times \frac{\partial u_{H2}}{\partial W_{21_1}} = 2 \times u_{cost} \times \sigma'(u_{A1}) \times W_{12_2} \times \sigma'(u_{H2}) \times X1$$

$$\frac{\partial C}{\partial W_{22_1}} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial u_{A1}} \times \frac{\partial u_{A1}}{\partial H2} \times \frac{\partial H2}{\partial u_{H2}} \times \frac{\partial u_{H2}}{\partial W_{22_1}} = 2 \times u_{cost} \times \sigma'(u_{A1}) \times W_{12_2} \times \sigma'(u_{H2}) \times X2$$

Update weights

Layer 1

$$W_{21_1} = W_{21_1} - \eta \times \frac{\partial C}{\partial W_{21_1}}$$

$$W_{22_1} = W_{22_1} - \eta \times \frac{\partial C}{\partial W_{22_1}}$$

$$W_{2\sim_1} = W_{2\sim_1} - \eta \times \frac{\partial C}{\partial W_{2\sim_1}}$$

$$W_{11_1} = W_{11_1} - \eta \times \frac{\partial C}{\partial W_{11_1}}$$

$$W_{12_1} = W_{12_1} - \eta \times \frac{\partial C}{\partial W_{12_1}}$$

$$W_{1\sim_1} = W_{1\sim_1} - \eta \times \frac{\partial C}{\partial W_{1\sim_1}}$$

Layer 2

$$W_{1\sim_2} = W_{1\sim_2} - \eta \times \frac{\partial C}{\partial W_{1\sim_2}}$$

$$W_{11_2} = W_{11_2} - \eta \times \frac{\partial C}{\partial W_{11_2}}$$

$$W_{12_2} = W_{12_2} - \eta \times \frac{\partial C}{\partial W_{12_2}}$$

Matrix Representation ( Layer 1 analysis)

$$\frac{\partial C}{\partial W_{22_1}} = \frac{\partial C}{\partial A1} \times \frac{\partial A1}{\partial u_{A1}} \times \frac{\partial u_{A1}}{\partial H2} \times \frac{\partial H2}{\partial u_{H2}} \times \frac{\partial u_{H2}}{\partial W_{22_1}} = 2 \times u_{cost} \times \sigma'(u_{A1}) \times W_{12_2} \times \sigma'(u_{H2}) \times X2$$

$$\mathbf{2} \times \mathbf{u_{cost}} \times \sigma'(\mathbf{u_{A1}}) \times W \times \sigma'(u_{H2}) \times X$$

Batch representation

$$\begin{bmatrix} \left[2 \times u_{cost_1} \times \sigma'(u_{A1_1})\right]_1 \\ \left[2 \times u_{cost_2} \times \sigma'(u_{A1_2})\right]_2 \\ \left[2 \times u_{cost_3} \times \sigma'(u_{A1_3})\right]_3 \\ \left[2 \times u_{cost_4} \times \sigma'(u_{A1_4})\right]_4 \\ \vdots \end{bmatrix}$$

$$2 \times u_{cost} \times \sigma'(u_{A1}) \times \mathbf{W} \times \sigma'(u_{H2}) \times X$$

Batch representation

$$\begin{bmatrix} \left[W_{11_2} \quad W_{12_2} \quad W_{1\sim 2}\right]_1 \\ \left[W_{11_2} \quad W_{12_2} \quad W_{1\sim 2}\right]_2 \\ \left[W_{11_2} \quad W_{12_2} \quad W_{1\sim 2}\right]_3 \\ \left[W_{11_2} \quad W_{12_2} \quad W_{1\sim 2}\right]_4 \\ \vdots \end{bmatrix}$$

$$2 \times u_{cost} \times \sigma'(u_{A1}) \times W \times \sigma'(\mathbf{u_H}) \times X$$

Batch representation

$$\begin{bmatrix} [\sigma'(\mathbf{u_{H1}})]_1 & [\sigma'(\mathbf{u_{H2}})]_1 \\ [\sigma'(\mathbf{u_{H1}})]_2 & [\sigma'(\mathbf{u_{H2}})\ ]_2 \\ [\sigma'(\mathbf{u_{H1}})]_3 & [\sigma'(\mathbf{u_{H2}})\ ]_3 \\ [\sigma'(\mathbf{u_{H1}})\ ]_4 & [\sigma'(\mathbf{u_{H2}})\ ]_4 \\ \vdots & \vdots \end{bmatrix}$$

$$2 \times u_{cost} \times \sigma'(u_{A1}) \times W \times \sigma'(u_H) \times \mathbf{X}$$

Batch representation

$$\begin{bmatrix} [X1 \quad X2 \quad 1]_1 \\ [X1 \quad X2 \quad 1]_2 \\ [X1 \quad X2 \quad 1]_3 \\ [X1 \quad X2 \quad 1]_4 \\ \vdots \end{bmatrix}$$

$$\frac{\partial C}{\partial W} = \left[ \left[ \frac{\partial C}{\partial W_{11_1}} \quad \frac{\partial C}{\partial W_{12_1}} \quad \frac{\partial C}{\partial W_{1\sim_1}} \quad \right] \left[ \frac{\partial C}{\partial W_{21_1}} \quad \frac{\partial C}{\partial W_{22_1}} \quad \frac{\partial C}{\partial W_{2\sim_1}} \quad \right] \right]$$

Layer 2 gradient can be solved similarly and in half the time as half the number of layer 1 gradients need to be computed.