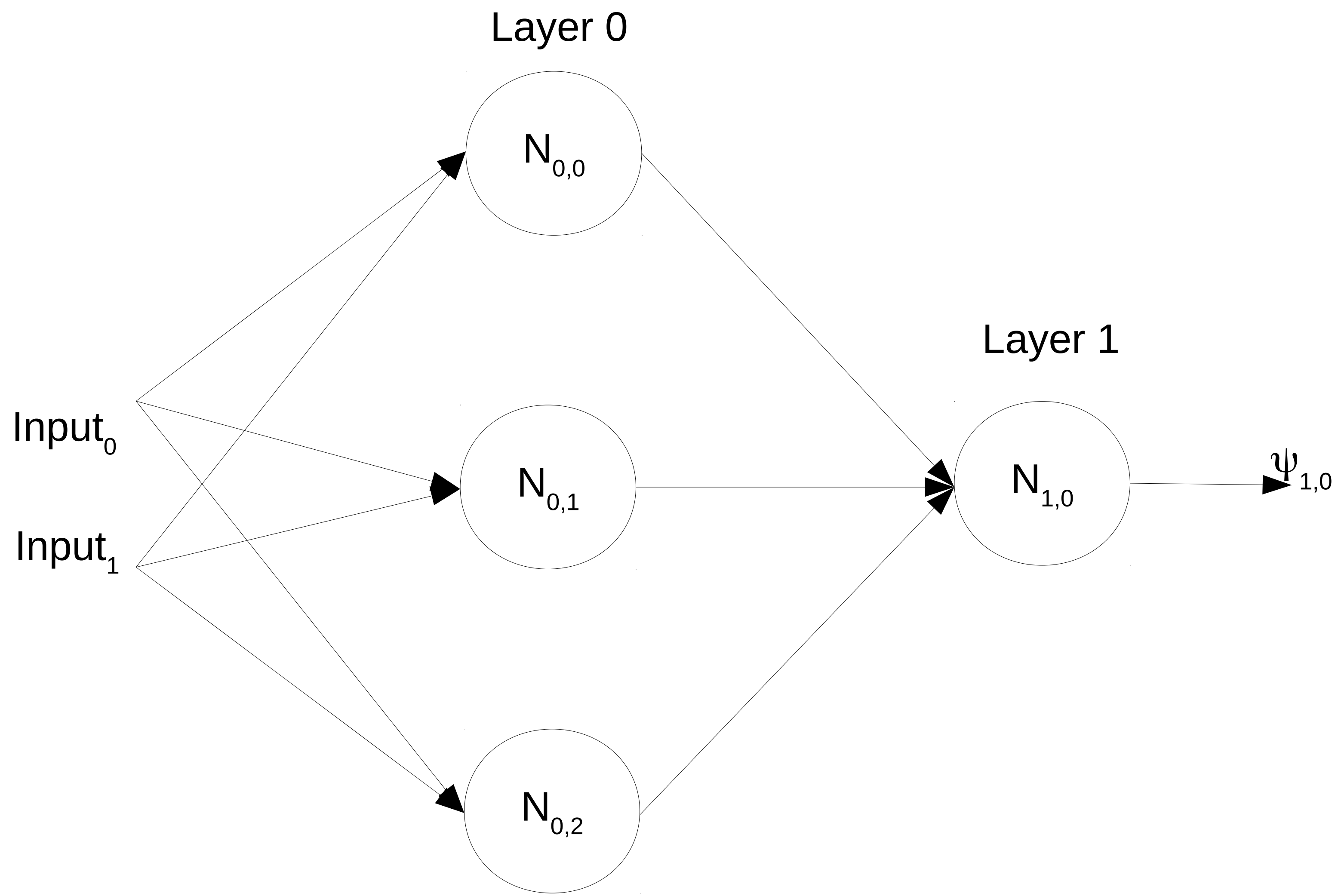
XOR Neural Network used in the Jupyter Notebook



Feed Forward Equations

- $N_{0,0}$
  - $Input_0 * w_{0,0,0} + Input_1 * w_{0,0,1} + w_{0,0,2} = z_{0,0}$
  - $\psi_{0,0} = 1 / (1 + e^{(-z_{0,0})})$
- $N_{0,1}$
  - $Input_0 * w_{0,1,0} + Input_1 * w_{0,1,1} + w_{0,1,2} = z_{0,1}$
  - $\psi_{0,1} = 1 / (1 + e^{(-z_{0,1})})$
- $N_{0,2}$
  - $Input_0 * w_{0,2,0} + Input_1 * w_{0,2,1} + w_{0,2,2} = z_{0,2}$
  - $\psi_{0,2} = 1 / (1 + e^{(-z_{0,2})})$

Feed Forward Equations (cont)

- $N_{1,0}$
  - $\psi_{0,0} * w_{1,0,0} + \psi_{0,1} * w_{1,0,1} + \psi_{0,2} * w_{1,0,2} + w_{1,0,3} = z_{1,0}$
  - $\psi_{1,0} = 1 / (1 + e^{(-z_{1,0})})$

Gradient Equations

- Gradients (with respect to layer 1 weights) for layer 1
- $Error_{1,0} = (t-\psi_{1,0})^2$
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(w_{1,0,0}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * \psi_{0,0}$
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(w_{1,0,1}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * \psi_{0,1}$
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(w_{1,0,2}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * \psi_{0,2}$
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(w_{1,0,3}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0})$

Gradient Equations (cont)

- Gradients (with respect to layer 0 weights, neuron 0) for layer 0
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(\psi_{0,0}) * \delta(\psi_{0,0})/\delta(z_{0,0}) * \delta(z_{0,0})/\delta(w_{0,0,0}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0}) * Input_0$
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(\psi_{0,0}) * \delta(\psi_{0,0})/\delta(z_{0,0}) * \delta(z_{0,0})/\delta(w_{0,0,1}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0}) * Input_1$
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(\psi_{0,0}) * \delta(\psi_{0,0})/\delta(z_{0,0}) * \delta(z_{0,0})/\delta(w_{0,0,2}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0})$
- Gradients (with respect to layer 0 weights, neuron 1) for layer 0
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(\psi_{0,1}) * \delta(\psi_{0,1})/\delta(z_{0,1}) * \delta(z_{0,1})/\delta(w_{0,1,0}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1}) * Input_0$
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(\psi_{0,1}) * \delta(\psi_{0,1})/\delta(z_{0,1}) * \delta(z_{0,1})/\delta(w_{0,1,1}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1}) * Input_1$
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(\psi_{0,1}) * \delta(\psi_{0,1})/\delta(z_{0,1}) * \delta(z_{0,1})/\delta(w_{0,1,2}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1})$

Gradient Equations (cont)

- Gradients (with respect to layer 0 weights, neuron 2) for layer 0
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(\psi_{0,2}) * \delta(\psi_{0,2})/\delta(z_{0,2}) * \delta(z_{0,2})/\delta(w_{0,2,0}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,2} * \psi_{0,2} * (1 - \psi_{0,2}) * Input_0$
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(\psi_{0,2}) * \delta(\psi_{0,2})/\delta(z_{0,2}) * \delta(z_{0,2})/\delta(w_{0,2,1}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,2} * \psi_{0,2} * (1 - \psi_{0,2}) * Input_1$
  - $\delta Error_{1,0}/\delta(\psi_{1,0}) * \delta(\psi_{1,0})/\delta(z_{1,0}) * \delta(z_{1,0})/\delta(\psi_{0,2}) * \delta(\psi_{0,2})/\delta(z_{0,2}) * \delta(z_{0,2})/\delta(w_{0,2,2}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,22} * \psi_{0,2} * (1 - \psi_{0,2})$

Back Propagation

- Layer 1 weight updates using gradients
  - $\delta Error_{1,0}/\delta(w_{1,0,0}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * \psi_{0,0}$
    - The constant 2 is of no consequence since we are going to be using a learning factor, η
    - $\psi_{0,0}$ is an output of the previous layer (from neuron 0 of layer 0)
    - t is the true/expected value – value to be learned
    - So the delta to be subtracted from $w_{1,0,0}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * \psi_{0,0}$
  - $\delta Error_{1,0}/\delta(w_{1,0,1}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * \psi_{0,1}$
    - The delta to be subtracted from $w_{1,0,1}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * \psi_{0,1}$
  - $\delta Error_{1,0}/\delta(w_{1,0,2}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * \psi_{0,2}$
    - The delta to be subtracted from $w_{1,0,2}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * \psi_{0,2}$
  - $\delta Error_{1,0}/\delta(w_{1,0,3}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0})$
    - The delta to be subtracted from $w_{1,0,3}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0})$

- Note that in all of these update equations for layer 1, the term, $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0})$, is the same.

Back Propagation (cont)

- Layer 0, Neuron 0, weight update using gradients
  - $\delta Error_{1,0} / \delta(w_{0,0,0}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0}) * Input_0$
  - $\delta Error_{1,0} / \delta(w_{0,0,1}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0}) * Input_1$
  - $\delta Error_{1,0} / \delta(w_{0,0,2}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0})$
    - The delta to be subtracted from $w_{0,0,0}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0}) * Input_0$
    - The delta to be subtracted from $w_{0,0,1}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0}) * Input_1$
    - The delta to be subtracted from $w_{0,0,2}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0})$
- Layer 0, Neuron 1, weight update using gradients
  - $\delta Error_{1,0} / \delta(w_{0,1,0}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1}) * Input_0$
  - $\delta Error_{1,0} / \delta(w_{0,1,1}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1}) * Input_1$
  - $\delta Error_{1,0} / \delta(w_{0,1,2}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1})$
    - The delta to be subtracted from $w_{0,1,0}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1}) * Input_0$
    - The delta to be subtracted from $w_{0,1,1}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1}) * Input_1$
    - The delta to be subtracted from $w_{0,1,2}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1})$
- Layer 0, Neuron 2, weight update using gradients
  - $\delta Error_{1,0} / \delta(w_{0,2,0}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,2} * \psi_{0,2} * (1 - \psi_{0,2}) * Input_0$
  - $\delta Error_{1,0} / \delta(w_{0,2,1}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,2} * \psi_{0,2} * (1 - \psi_{0,2}) * Input_1$
  - $\delta Error_{1,0} / \delta(w_{0,2,2}) = -2*(t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,2} * \psi_{0,2} * (1 - \psi_{0,2})$
    - The delta to be subtracted from $w_{0,2,0}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,2} * \psi_{0,2} * (1 - \psi_{0,2}) * Input_0$
    - The delta to be subtracted from $w_{0,2,1}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,2} * \psi_{0,2} * (1 - \psi_{0,2}) * Input_1$
    - The delta to be subtracted from $w_{0,2,2}$ is $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0}) * w_{1,0,2} * \psi_{0,2} * (1 - \psi_{0,2})$
- Note, again, that in all of these update equations for layer 0, the term, $-\eta * (t - \psi_{1,0}) * \psi_{1,0} * (1 - \psi_{1,0})$, is the same. Let's call this term δ

Back Propagation (cont)

- Then the delta to be subtracted from $w_{1,0,0}$ is $\delta * \psi_{0,0}$
- Then the delta to be subtracted from $w_{1,0,1}$ is $\delta * \psi_{0,1}$
- Then the delta to be subtracted from $w_{1,0,2}$ is $\delta * \psi_{0,2}$
- Then the delta to be subtracted from $w_{1,0,3}$ is δ
- Then the delta to be subtracted from $w_{0,0,0}$ is $\delta * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0}) * Input_0$
- Then the delta to be subtracted from $w_{0,0,1}$ is $\delta * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0}) * Input_1$
- Then the delta to be subtracted from $w_{0,0,2}$ is $\delta * w_{1,0,0} * \psi_{0,0} * (1 - \psi_{0,0})$
- Then the delta to be subtracted from $w_{0,1,0}$ is $\delta * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1}) * Input_0$
- Then the delta to be subtracted from $w_{0,1,1}$ is $\delta * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1}) * Input_1$
- Then the delta to be subtracted from $w_{0,1,2}$ is $\delta * w_{1,0,1} * \psi_{0,1} * (1 - \psi_{0,1})$
- Then the delta to be subtracted from $w_{0,2,0}$ is $\delta * w_{1,0,2} * \psi_{0,2} * (1 - \psi_{0,2}) * Input_0$
- Then the delta to be subtracted from $w_{0,2,1}$ is $\delta * w_{1,0,2} * \psi_{0,2} * (1 - \psi_{0,2}) * Input_1$
- Then the delta to be subtracted from $w_{0,2,2}$ is $\delta * w_{1,0,2} * \psi_{0,2} * (1 - \psi_{0,2})$