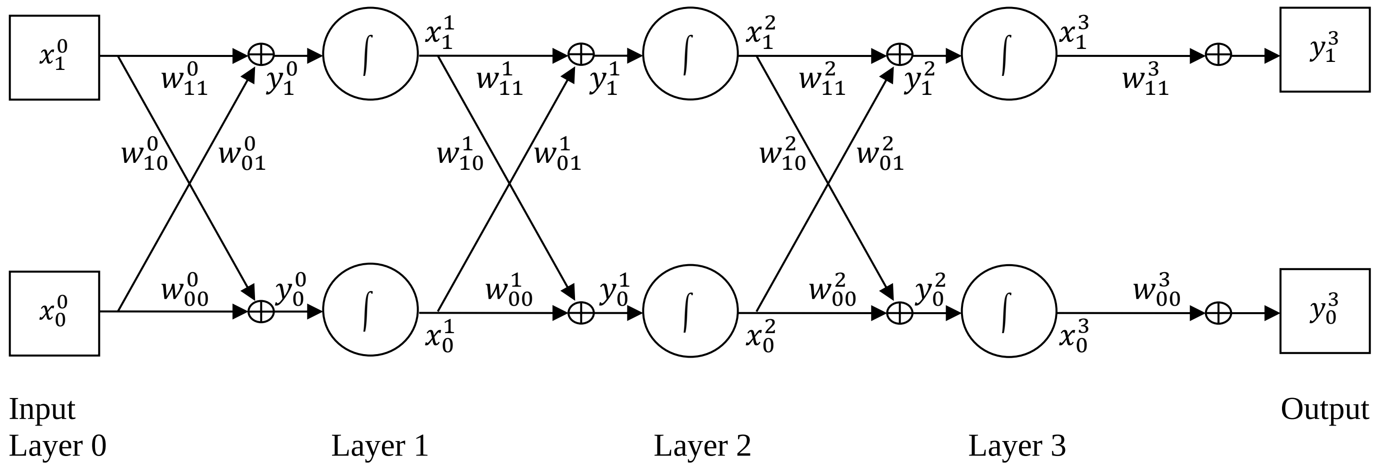
Neural Networks – Basics

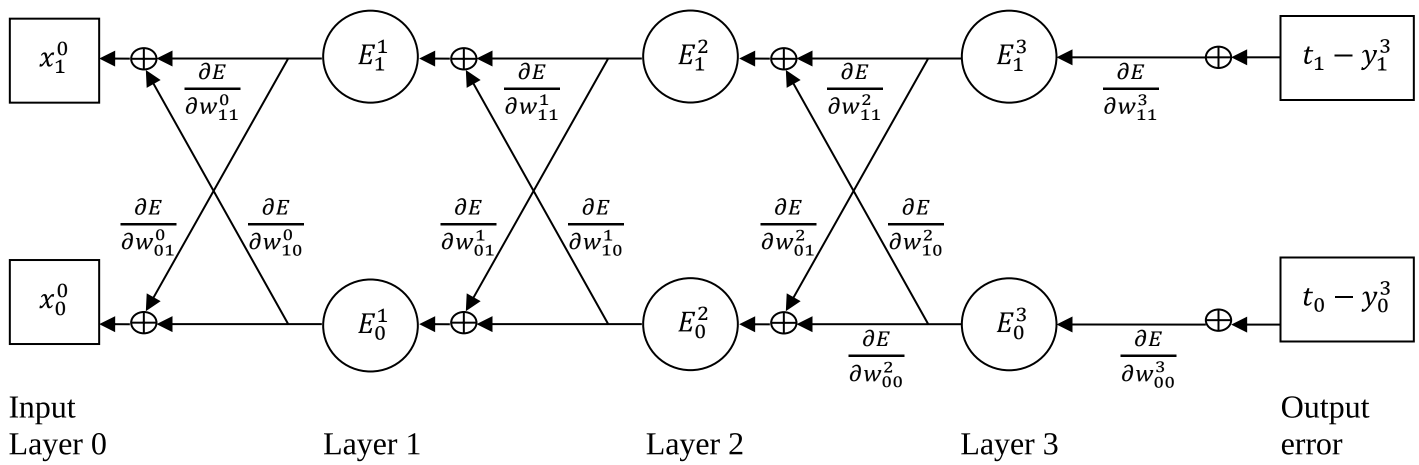


Forward propagation portion of a standard Neural Network

The transfer function in the cells is generally a sigmoid (sometimes called a squashing function). The characteristic of these sigmoids is that the derivatives are straightforward or easy to calculate, especially in terms of the *output* of the function. A typical such function is $f(x) = 1/(1 + e^{-x})$ so that $f'(x) = f(x) \times \big(1 - f(x)\big)$ or, more simply, $f' = f(1 - f)$.

Define the error of the neural network as $E = \frac{1}{2}\sum_i \left(t_i - y_i^{outLayer}\right)^2 = \frac{1}{2}\sum_i \left(t_i - x_i^{final} w_{ii}^{final}\right)^2$.

Neural Networks – Back Propagation



Back propagation portion of a standard Neural Network

We look for the negative of the gradient vector $-\nabla E = -< \frac{\partial E}{\partial w_i}, \cdots >$. This is given recursively by final conditions of $-\frac{\partial E}{\partial w_{ii}^{final}} = \left(t_i - x_i^{final} w_{ii}^{final}\right)\left(x_i^{final}\right)$ and $E_i^{final} = \left(t_i - x_i^{final} w_{ii}^{final}\right) w_{ii}^{final} f'\left(@x_i^{final}\right)$ where the @ in $f'\left(@x_i^{final}\right) = x_i^{final}\left(1 - x_i^{final}\right)$ is a reminder that the argument shown is the output of f and not the input. To be clear, $-\frac{\partial E}{\partial w_{ii}^{final}}$ corresponds to the weights that are between the rightmost circular nodes and the final output rectangular boxes (ie. weights to the right of Layer 3), while E_i^{final} corresponds to the rightmost layer of circular nodes (ie. Layer 3), prior to the final output. In the above, $final = 3$.

The recursive portion (ie. calculating $-\frac{\partial E}{\partial w_{ij}^{layer}}$ and E_i^{layer} for a set of weights and the nodes to the left of those weights) is given by $-\frac{\partial E}{\partial w_{ij}^{layer}} = x_i^{layer} E_j^{layer+1}$ and $E_i^{layer} = \left(\sum_j w_{ij}^{layer} E_j^{layer+1}\right) f'(@x_i^{layer})$.

The weights should then be updated by addition of a scaled multiple (say 0.1) of the negative of the gradient vector.