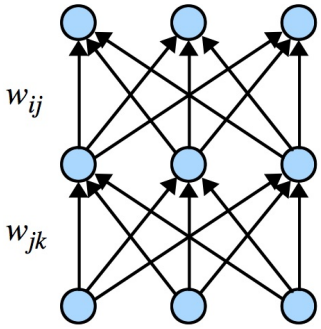
Backpropagation Learning Algorithm for a 3-Layer Feedforward Neural Network



"i-units" a_i = activation of output unit i
b_i = bias of output unit i

"j-units" a_j = activation of hidden unit j
b_j = bias of hidden unit j

"k-units" x_k = activation of input unit k

w_{ij} = connection weight from hidden unit j to output unit i
w_{jk} = connection weight from input unit k to hidden unit j
y_i = target value for output unit i

Forward pass: given input pattern x, compute output activations

1. Sum up all incoming activity received by hidden unit j: $z_j = \left(\sum_k w_{jk}\, x_k \right) + b_j$
2. Compute activation of hidden unit j: $a_j = \sigma(z_j)$ where σ is the sigmoid function $\sigma(z) = \dfrac{1}{1+e^{-z}}$
3. Sum up all incoming activity received by output unit i: $z_i = \left(\sum_j w_{ij}\, a_j \right) + b_i$
4. Compute activation of output unit i: $a_i = \sigma(z_i)$

Backward pass: given output activations, backpropagate error and update weights

5. Compute delta value for each output unit i: $\delta_i = (a_i - y_i)\, a_i\, (1 - a_i)$
6. Compute delta value for each hidden unit j: $\delta_j = \left(\sum_i w_{ij}\, \delta_i \right) a_j\, (1 - a_j)$
7. Compute weight and bias changes for hidden $\rightarrow$ output layer: $\Delta w_{ij} = -\eta\, \delta_i\, a_j \qquad \Delta b_i = -\eta\, \delta_i$
8. Compute weight and bias changes for input $\rightarrow$ hidden layer: $\Delta w_{jk} = -\eta\, \delta_j\, x_k \qquad \Delta b_j = -\eta\, \delta_j$
9. Update hidden $\rightarrow$ output weights and biases: $w_{ij} = w_{ij} + \Delta w_{ij} \qquad b_i = b_i + \Delta b_i$
10. Update input $\rightarrow$ hidden weights and biases: $w_{jk} = w_{jk} + \Delta w_{jk} \qquad b_j = b_j + \Delta b_j$

Momentum

The momentum parameter $0 \leq \alpha \leq 1$ controls how much the previous weight/bias change at time $t-1$ contributes to the current change at time t (these equations replace steps 7 and 8 above):

$$\Delta w_{ij}(t) \;=\; -\eta\,\delta_i\,a_j \;+\; \alpha\,\Delta w_{ij}(t-1) \qquad \text{for the hidden} \rightarrow \text{output weights}$$

$$\Delta w_{jk}(t) \;=\; -\eta\,\delta_j\,x_k \;+\; \alpha\,\Delta w_{jk}(t-1) \qquad \text{for the input} \rightarrow \text{hidden weights}$$

$$\Delta b_i(t) \;=\; -\eta\,\delta_i \;+\; \alpha\,\Delta b_i(t-1) \qquad \text{for the output unit biases}$$

$$\Delta b_j(t) \;=\; -\eta\,\delta_j \;+\; \alpha\,\Delta b_j(t-1) \qquad \text{for the hidden unit biases}$$