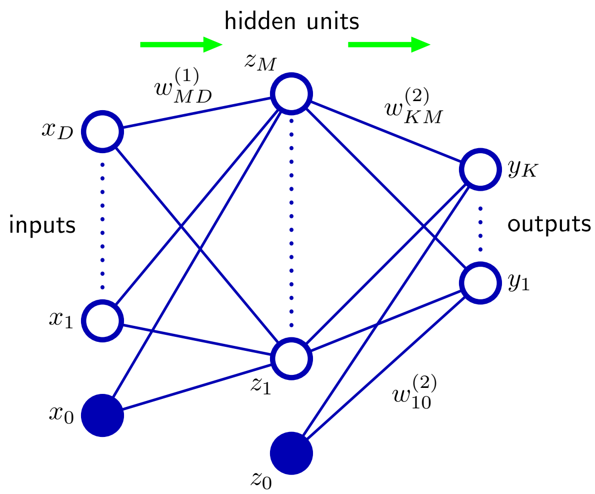
Task 1 - Forward propagation



Consider a two-layer network of the form illustrated above together with a sum-of-squares error, in which the output units have linear activation functions, so that $y_k = a_k$, while the hidden units have sigmoidal activation functions given by

$$h(a) = \tanh(a) \tag{1}$$

$$\text{with} \qquad \tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}} \tag{2}$$

$$\text{and} \qquad h'(a) = 1 - h(a)^2. \tag{3}$$

We denote the inputs to the hidden layer with a_j for $j = 1 \dots M$.
We also consider a standard sum-of-squares error function, so that for datapoint n the error is given by

$$E_n = \frac{1}{2} \sum_{k=1}^{K} (y_k - t_k)^2 \tag{4}$$

where y_k is the activation of output unit k, and t_k is the corresponding target, for a particular input x_n.

Compute forward propagation, i.e. compute the activation of the neurons in the hidden layer and the activation of the output layer neurons.

a) $a_j =$

b) $y_k =$

Task 2 - Backpropagation

To perfom backpropoagation, we need to compute the derivative of the error function with respect to the corresponding weights. Compute the derivative wrt the first and second-layer weights. You therefore need to apply the chain rule.

(a) $\dfrac{\partial E_n}{\partial w^{(2)}_{kj}} =$

(b) $\dfrac{\partial E_n}{\partial w^{(1)}_{ji}} =$

Some solutions

Task 1

(a) $a_j = \sum_{i=0}^{D} w_{ji}^{(1)} x_i$

(b) $y_k = \sum_{j=0}^{M} w_{kj}^{(2)} z_j$ where $z_j = \tanh(a_j)$

Task 2

(a) $\dfrac{\partial E_n}{\partial w_{kj}^{(2)}} = \dfrac{\partial E_n}{\partial y_k} \dfrac{\partial y_k}{\partial w_{kj}^{(2)}} = (y_k - t_k)\tanh(a_j)$

(b) $\dfrac{\partial E_n}{\partial w_{ji}^{(1)}} = \dfrac{\partial E_n}{\partial y} \dfrac{\partial y}{\partial z_j} \dfrac{\partial z_j}{\partial a_j} \dfrac{\partial a_j}{\partial w_{ji}^{(1)}} = \sum_{k=1}^{K} (y_k - t_k) w_{kj}^{(2)} (1 - \tanh(a_j)^2) x_i$

where:

$$\begin{aligned}
\frac{\partial E_n}{\partial y} &= (y_k - t_k)_{(1 \dots K)} \in \mathbb{R}^{1 \times K} \\
\frac{\partial y}{\partial z_j} &= w_{kj}^{(2)}{}_{(1 \dots K)} \in \mathbb{R}^{K \times 1} \\
\frac{\partial z_j}{\partial a_j} &= (1 - \tanh(a_j)^2) \\
\frac{\partial a_j}{\partial w^{(1)}} &= x_i
\end{aligned}$$