Backpropagation

Daniel López Montero

October 2018

1 Definition

w_{ij}^k: weight for node j in layer l_k for incoming node i.
b_i^k: bias for node i in layer l_k.
a_i^k: product sum plus bias (activation) for node i in layer l_k.
o_i^k: output for node i in layer l_k.
n_k: number of nodes i layer l_k.

g: activation function for the hidden layer nodes.
g_o: activation function for the output layer nodes.

2 Feed-forward

$$a_i^k = b_i^k + \sum_{j=1}^{n_{k-1}} w_{ji}^k o_j^{k-1}$$

$$o_i^k = g(a_i^k)$$

3 Cost function

$$E = \frac{1}{2}(y - o_i^k)^2$$

4 Backpropagation

$$\frac{\partial E}{\partial w_{ij}^k} = \frac{\partial E}{\partial o_i^k}\frac{\partial o_i^k}{\partial a_i^k}\frac{\partial a_i^k}{\partial w_{ij}^k}$$

$$\delta_j^k \equiv \frac{\partial E}{\partial a_j^k} = \frac{\partial E}{\partial o_i^k}\frac{\partial o_i^k}{\partial a_i^k}$$

$$\frac{\partial a_i^k}{\partial w_{ij}^k} = \frac{\partial}{w_{ij}^k}(b_i^k + \sum_{j=1}^{n_{k-1}} w_{ji}^k o_j^{k-1}) = o_i^{k-1}$$

$$\frac{\partial E}{\partial w_{ij}^k} = \delta_j^k o_i^{k-1}$$

4.1 Output layer

$$\delta_j^k = \frac{\partial E}{\partial a_j^k} = (g_o(a_j^l) - y)g_o'(a_j^l)$$

$$\frac{\partial E}{\partial w_{ij}^k} = \delta_j^k o_i^{k-1} = (g_o(a_j^l) - y)g_o'(a_j^l)o_i^{l-1}$$

4.2 Hidden layer

where l ranges from 1 to r_{k+1}, number of nodes in the next layer.

$$\delta_j^k = \frac{\partial E}{\partial a_j^k} = \sum_{l=1}^{n_{k+1}} \delta_l^{k+1}\frac{\partial a_l^{k+1}}{\partial a_j^k}$$

$$a_l^{k+1} = \sum_{j=1}^{n_k} w_{jl}^{k+1} g(a_j^k)$$

$$\frac{\partial a_l^{k+1}}{\partial a_j^k} = w_{jl}^{k+1} g'(a_j^k)$$

$$\delta_j^k = \sum_{l=1}^{n_{k+1}} \delta_l^{k+1} w_{jl}^{k+1} g'(a_j^k) = g'(a_j^k)\sum_{l=1}^{n_{k+1}} \delta_l^{k+1} w_{jl}^{k+1}$$

$$\frac{\partial E}{\partial w_{ij}^k} = \delta_j^k o_i^{k-1} = g'(a_j^k) o_i^{k-1} \sum_{l=1}^{n_{k+1}} w_{jl}^{k+1}\delta_k^{k+1}$$