Artificial Intelligence I 2004 – Paper 6 Question 7 (SBH)

For 2004 only, the Question for the Part II Artificial Intelligence course was the same as that for the Part IB course.

Context: the question is from the section of the course dealing with neural networks. It requires the derivation of the backpropagation algorithm in its most general form. For node j define

$$a_j = w_0^{(j)} + \sum_{i=1}^{k} w_i^{(j)} \text{input}(i)$$

$$z_j = g(a_j)$$

and

$$\delta_j = \frac{\partial E}{\partial a_j}$$

The first two quantities are obtained for the ith labelled example by applying it to the inputs of the network and allowing the effects to propagate through. This is the *forward propagation* step. Denote by w_{ji} the weight connecting node i to node j, where weights $w_0^{(\text{node})}$ are assumed connected to an extra input with a constant value of 1.

$$\frac{\partial E}{\partial \mathbf{w}_{ji}} = \frac{\partial E}{\partial a_j} \frac{\partial a_j}{\partial \mathbf{w}_{ji}}$$

where due to the fact that the node performs a weighted summation

$$\frac{\partial a_j}{\partial \mathbf{w}_{ji}} = z_i \tag{1}$$

The determination of the δs splits into two parts, one for nodes that produce outputs and one for other nodes. For nodes that produce outputs δ_j is straightforward to obtain by direct differentiation of the error E and the activation function g.

$$\frac{\partial E}{\partial a_j} = \delta_j = \frac{\partial z_j}{\partial a_j} \frac{\partial E}{\partial z_j} = g'(a_j) \frac{\partial E}{\partial z_j} \tag{2}$$

and here both $g'(a_j)$ and $\frac{\partial E}{\partial z_j}$ can be obtained depending on the specific activation function and E used. For other nodes we have

$$\delta_j = \frac{\partial E}{\partial a_j} = \sum_{\text{nodes } p \text{ at output of } j} \frac{\partial E}{\partial a_p} \frac{\partial a_p}{\partial a_j}$$

As we know the δs for output nodes we can therefore work backwards through the network computing earlier values for δs using the above equation, as

$$\frac{\partial a_p}{\partial a_j} = \frac{\partial a_p}{\partial z_j} \frac{\partial z_j}{\partial a_j}$$

and

$$\frac{\partial z_j}{\partial a_j} = g'(a_j)$$

can again be computed depending on the specific activation function used, while

$$\frac{\partial a_p}{\partial z_j} = w_{pj}$$

and combining these expressions gives

$$\delta_j = \sum_{\text{nodes } p \text{ at output of } j} \delta_p g'(a_j) w_{pj} = g'(a_j) \sum_{\text{nodes } p \text{ at output of } j} \delta_p w_{pj}$$