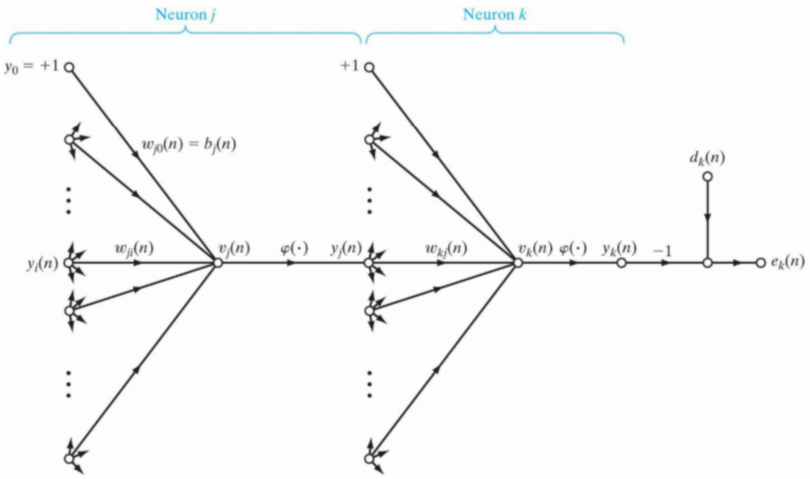
Name: Inchara Raveendra

SCU ID: 00001653600

# Assignment - 10

Consider the following signal-flow graph of a fully-connected neural network that consists of an input layer, one hidden layer and an output layer. $y_i$ is the $i^{th}$ input node in the input layer. Neuron $j$ is the $j^{th}$ neuron in the hidden layer and neuron $k$ is the $k^{th}$ output neuron. Assume the activation function $\varphi(.)$ is sigmoid.




(a). Use back-propagation on the output neuron $k$ to show that the weight correction $\Delta w_{kj}$ for the $n^{th}$ iteration is given by

$$\Delta w_{kj}(n) = \eta.\delta_k(n).y_j(n)$$

Where $\eta$ is the learning rate and the local gradient $\delta_k(n) = [d_k(n) - y_k(n)].[y_k(n)(1 - y_k(n))]$

(b). Use back-propagation on the hidden neuron $j$ to show that the weight correction $\Delta w_{ji}$ for the $n^{th}$ iteration is given by

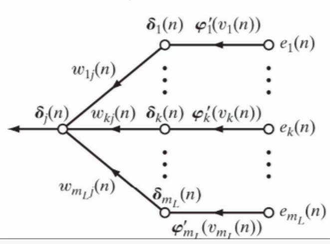
$$\Delta w_{ji}(n) = \eta.\delta_j(n).y_i(n)$$

Where the $\delta_j(n)$ is the overall backpropagated gradient from the layer to the immediate right (i.e., the output layer) given by

$$\delta_j(n) = \sum_k \delta_k(n).w_{kj}(n).y_j(n).\left(1 - y_j(n)\right)$$

and $y_j(n)$ is the output of the hidden neuron $j$.

Note that the effect of all $e_k$'s must be included, hence the summation over $k$.



a) Output Neuron k

least squares error

$$\mathcal{E}(n) = \frac{1}{2} \sum_k e_k^2(n) = \frac{1}{2} \sum_k (d_k(n) - y_k(n))^2$$

$$\frac{\partial \mathcal{E}(n)}{\partial \mathcal{E}_k(n)} = \frac{\partial \mathcal{E}(n)\, \partial \mathcal{E}_k(n)\, \partial y_k(n)\, \partial v_k(n)}{\partial e_k(n)\, \partial y_k(n)\, \partial v_k(n)\, \partial w_{kj}(n)}$$

$$\frac{\partial \mathcal{E}(n)}{\partial e_k(n)} = e_k(n) \qquad \frac{\partial e_k(n)}{\partial y_k(n)} = -1$$

$$y_k(n) = \varphi(v_k(n)) \rightarrow \frac{\partial y_k(n)}{\partial v_k(n)} = \varphi'(v_k(n))$$

$$v_k(n) = \sum_j w_{kj}(n)\, y_j(n) \rightarrow \frac{\partial v_k(n)}{\partial w_{kj}(n)} = y_j(n)$$

$$\therefore \quad \frac{\partial \mathcal{E}(n)}{\partial w_{kj}(n)} = -e_k(n) \cdot \varphi'(v_k(n)) \cdot y_j(n)$$

$$\Delta w_{kj}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial w_{kj}(n)} = \eta \cdot e_k(n) \cdot \varphi'(v_k(n)) \cdot y_j(n)$$

$$\delta_k(n) = e_k(n) \cdot \varphi'(v_k(n))$$

$$\therefore \quad \Delta w_{kj}(n) = \eta \cdot \delta_k(n) \cdot y_j(n)$$

If $\varphi(v_k(n))$ is a sigmoid function, then

$$\varphi'(v_k(n)) = \varphi(v_k(n))\,(1 - \varphi(v_k(n)))$$

But $\varphi(v_k(n)) = y_k(n)$

$$\therefore \delta_k(n) = e_k(n) \cdot \varphi'(v_k(n)) = [d_k(n) - y_k(n)] \cdot [y_k(n)(1 - y_k(n))]$$

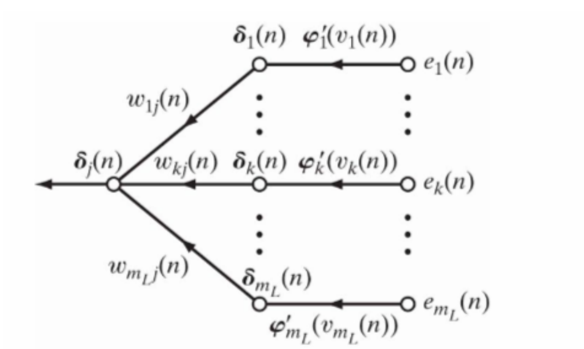
b) Hidden neuron $j$

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = \frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial w_{ji}(n)}$$

where,

$$\mathcal{E}(n) = \frac{1}{2} \sum_k e_k^2(n) = \frac{1}{2} \sum_k (d_k(n) - y_k(n))^2$$

Effect of all $e_k$'s must be included




Hence,

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial y_j(n)}$$

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n) \partial v_k(n)}{\partial v_k(n) \partial y_j(n)}$$

where,

$$e_k(n) = d_k(n) - y_k(n) = d_k(n) - \varphi(v_k(n))$$

$$\rightarrow \frac{\partial e_k(n)}{\partial v_k(n)} = -\varphi'(v_k(n))$$

$$v_k(n) = \sum_j w_{kj}(n) y_j(n) \rightarrow \frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n)$$

Thus,

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n) \partial v_k(n)}{\partial v_k(n) \partial y_j(n)}$$

$$= -\sum_k e_k(n) . \varphi'(v_k(n)) . w_{kj}(n)$$

Next, $$\frac{\partial y_j(n)}{\partial w_{ji}(n)} = \frac{\partial y_j(n)}{\partial v_j(n)} \cdot \frac{\partial v_j(n)}{\partial w_{ji}(n)}$$

where, $$y_j(n) = \varphi[v_j(n)] \rightarrow \frac{\partial y_j(n)}{\partial v_j(n)} = \varphi'(v_j(n))$$

$$v_j(n) = \sum_i w_{ji}(n) y_i(n) \rightarrow \frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n)$$

$$\therefore \frac{\partial y_j(n)}{\partial w_{ji}(n)} = \frac{\partial y_j(n)}{\partial v_j(n)} \cdot \frac{\partial v_j(n)}{\partial w_{ji}(n)} = \varphi'(v_j(n)) . y_i(n)$$

Hence, $\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = \frac{\partial \mathcal{E}(n)\, \partial y_j(n)}{\partial y_j(n)\, \partial w_{ji}(n)}$

$$= -\sum_k e_k(n) \cdot \varphi'(v_k(n)) \cdot w_{kj}(n) \cdot \varphi'(v_j(n)) \cdot y_i(n)$$

$$= -\sum_k \delta_k(n) \cdot w_{kj}(n) \cdot \varphi'(v_j(n)) \cdot y_i(n)$$

From (a) : $\delta_k(n) = e_k(n) \cdot \phi'((v_k(n)))$

Overall local gradient $\delta_j(n)$ which is backpropagated gradient from layer to the immediate right (output layer):

$$\delta_j(n) = \sum_k \delta_k(n) \cdot w_{kj}(n) \cdot \phi'(v_j(n))$$

Thus,

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = -\delta_j(n) \cdot y_i(n)$$

If $\varphi(v_j(n))$ is a sigmoid function, then

$$\varphi'(v_j(n)) = \varphi(v_j(n))\,(1 - \varphi(v_j(n))) = y_j(n)\,(1 - y_j(n))$$

So, $\delta_j(n) = \sum_k \delta_k(n) \cdot w_{kj}(n) \cdot y_j(n) \cdot (1 - y_j(n))$

$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = \eta \cdot \delta_j(n) \cdot y_i(n)$$

Recall $y_j(n)$ is the output of the hidden neuron $j$, $y_i(n)$ is the input node $i$ and $\delta_k(n)$ is the local gradient for the output neuron $k$ described in (a)

$$\therefore \delta_k(n) = [d_k(n) - y_k(n)] \cdot [y_k(n)(1 - y_k(n))]$$