The Backpropagation Update Rules

Network description and notation:

We will assume a network with an input layer, an output layer, and a hidden layer.

Assume the network is fully connected from input to hidden and hidden to output.

Assume all nodes use the same activation function $f(x)$

l = number of nodes in output layer
m = number of nodes in hidden layer
n = number of nodes in input layer

The weight from input node i to hidden node j is w_{ji}

The weight from hidden node j to output node k is v_{kj}

The input to a hidden node is denoted by *h_in*
The output from a hidden node is denoted by *h_out*

The input to an output node is denoted by *o_in*
The output from an output node is denoted by *o_out*

The error in the network is: $E = \frac{1}{2}[\text{desired} - \text{output}]^2$

Since there are multiple output nodes, that becomes:

$$E = \frac{1}{2}\sum_{p=1}^{l}(d[p] - o\_out[p])^2$$

the sum of the errors over all the output nodes.

The update rule for the v weights is: $v_{ki}^{new} = v_{ki}^{old} - \eta \frac{\partial E}{\partial v_{ki}^{old}}$

The update rule for the w weights is: $w_{ij}^{new} = w_{ij}^{old} - \eta \frac{\partial E}{\partial w_{ij}^{old}}$

Now, compute the partials of the error E with respect to the weights.

V update rule :

We want to update a particular weight, say $v[k][i]$. In that case, only output node k is of interest and all terms drop from the sum except for when the index p is equal to k. Giving:

$$\nabla E = \frac{\partial E}{\partial v_{ki}} = \frac{1}{2} \frac{\partial}{\partial v_{ki}} (d[k] - o\_out[k])^2$$

$$= 2 * \frac{1}{2} * (d[k] - o\_out[k]) \frac{\partial}{\partial v_{ki}} (d[k] - f(o\_in[k]))$$

$$= (d[k] - o\_out[k]) \frac{\partial}{\partial v_{ki}} (-f(o\_in[k]))$$

$$= -(d[k] - o\_out[k]) f'(o\_in[k]) \frac{\partial}{\partial v_{ki}} \left[\sum_{q=0}^{m} v_{kq} * h\_out[q] \right]$$

$$= -(d[k] - o\_out[k]) f'(o\_in[k]) * h\_out[i]$$

Again, note that the partial with respect to $v[k][i]$ of the sum is zero for all terms except when $q=i$.

The W update rule:

$$
\begin{aligned}
\frac{\partial E}{\partial w_{ij}} &= \frac{\partial}{\partial w_{ij}} \frac{1}{2} \sum_{p=1}^{l} (d[p] - o\_out[p])^2 \\
&= \frac{1}{2} \sum_{p=1}^{l} \frac{\partial}{\partial w_{ij}} (d[p] - o\_out[p])^2 \\
&= \sum_{p=1}^{l} 2 * \frac{1}{2} * (d[p] - o\_out[p]) \frac{\partial}{\partial w_{ij}} (d[p] - o\_out[p]) \\
&= \sum_{p=1}^{l} (d[p] - o\_out[p]) \frac{\partial}{\partial w_{ij}} (-o\_out[p]) \\
&= -\sum_{p=1}^{l} (d[p] - o\_out[p]) \frac{\partial}{\partial w_{ij}} (f(o\_in[p])) \\
&= -\sum_{p=1}^{l} (d[p] - o\_out[p]) f'(o\_in[p]) \frac{\partial}{\partial w_{ij}} (o\_in[p]) \\
&= -\sum_{p=1}^{l} (d[p] - o\_out[p]) f'(o\_in[p]) \frac{\partial}{\partial w_{ij}} (\sum_{q=0}^{m} v_{pq} * h\_out[q]) \\
&= -\sum_{p=1}^{l} (d[p] - o\_out[p]) f'(o\_in[p]) v_{pi} \frac{\partial}{\partial w_{ij}} h\_out[i] \\
&= -\sum_{p=1}^{l} (d[p] - o\_out[p]) f'(o\_in[p]) v_{pi} \frac{\partial}{\partial w_{ij}} f(h\_in[i]) \\
&= -\sum_{p=1}^{l} (d[p] - o\_out[p]) f'(o\_in[p]) v_{pi} * f'(h\_in[i]) \frac{\partial}{\partial w_{ij}} h\_in[i] \\
&= -\sum_{p=1}^{l} (d[p] - o\_out[p]) f'(o\_in[p]) v_{pi} * f'(h\_in[i]) \frac{\partial}{\partial w_{ij}} \sum_{s=0}^{n} w_{is} * x[s]) \\
&= -\sum_{p=1}^{l} (d[p] - o\_out[p]) f'(o\_in[p]) v_{pi} * f'(h\_in[i]) * x[j] \\
&= -f'(h\_in[i]) * x[j] * \sum_{p=1}^{l} (d[p] - o\_out[p]) f'(o\_in[p]) v_{pi}
\end{aligned}
$$

Notice that the sum only depends on i.

For efficiency, set

$$\delta[i] = \sum_{p=1}^{l} (d[p] - o\_out[p]) f'(o\_in[p]) v_{pi}$$

This gives

$$\frac{\partial E}{\partial w_{ij}} = -f'(h\_in[i]) * x[j] * \delta[i]$$