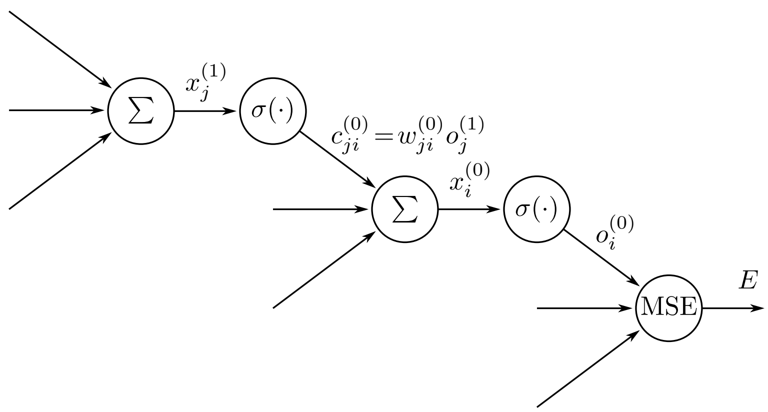

Comments:

L: I think, the notation is still way from optimal since we always need additional parenthesis when your layer index meets an exponent: $(o_i^{(0)})^2$

L: I was talking with Bhiksha this morning. He pointed out, where the recursion actually is. I am not sure, I understood this completely, but we should work on this together tonight. I am as well curious, what your parenthesis down there explain. :P

Name and network definitions:

$$E = \sum_i (o_i^{(0)} - t_i)^2 \qquad o_i^{(0)} = \sigma(x_0^{(i)}) \tag{1}$$

$$x_i^{(0)} = \sum_j w_{ji}^{(0)} o_j^{(1)} \qquad c_{ji}^{(0)} = w_{ji}^{(0)} o_j^{(1)} \tag{2}$$

$$c_{ji}^{(0)} = w_{ji}^{(0)} o_j^{(1)} \qquad c_{ji}^{(0)} = w_{ji}^{(0)} o_j^{(1)} \tag{3}$$

Derivatives with respect to network output:

$$\frac{\partial E}{\partial o_i^{(0)}} = o_i^{(0)} - t_i \qquad \frac{\partial^2 E}{\partial (o_i^{(0)})^2} = 1 \tag{4}$$

First and second derivative of the sigmoid function:

$$\sigma' = \sigma(1-\sigma) \qquad \sigma'' = \sigma'(1-2\sigma) \tag{5}$$

Derivative with respect to last layer output:

$$\frac{\partial E}{\partial x_i^{(0)}} = \frac{\partial E}{\partial o_i^{(0)}} \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} \tag{6}$$

$$= (o_i^{(0)} - t_i)(o_i^{(0)}(1 - o_i^{(0)})) \tag{7}$$

Second derivative with respect to last layer output:

$$\frac{\partial^2 E}{\partial (x_i^{(0)})^2} = \frac{\partial}{\partial x_i^{(0)}} \left(\frac{\partial E}{\partial o_i^{(0)}} \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} \right) \tag{8}$$

$$= \frac{\partial^2 x_i^{(0)}}{\partial (o_i^{(0)})^2} \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} + \frac{\partial E}{\partial o_i^{(0)}} \frac{\partial^2 o_i^{(0)}}{\partial (x_i^{(0)})^2} \tag{9}$$

$$= \frac{\partial^2 x_i^{(0)}}{\partial (o_i^{(0)})^2} o_i^{(0)}(1 - o_i^{(0)}) + (o_i^{(0)} - t_i) \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} \tag{10}$$

$$= \sigma'(x_i^{(0)}) o_i^{(0)}(1 - o_i^{(0)}) + (o_i^{(0)} - t_i)\sigma'(x_i^{(0)})(1 - 2o_i^{(0)}) \tag{11}$$

First derivative with respect to single input contribution:

$$\frac{\partial E}{\partial c_{ij}^{(0)}} = \frac{\partial E}{\partial o_i^{(0)}} \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} \cdot \cdots \frac{\partial x_i^{(0)}}{\partial c_{ji}^{(0)}} \tag{12}$$

$$= \left(o_i^{(0)} - t_i \right) \underbrace{\left(o_i^{(0)} \left(1 - o_i^{(0)} \right) \right)}_{\sigma'} \tag{13}$$

Where we use:

$$\frac{\partial x_i^{(0)}}{\partial c_{ji}^{(0)}} = \frac{\partial}{\partial c_{ji}^{(0)}} \left(\sum w_{ji}^{(0)} o_j^{(1)} \right) = \frac{\partial}{\partial c_{ji}^{(0)}} \left(c_{ji}^{(0)} + k \right) = 1 \tag{14}$$

$$\tag{15}$$

Second derivative with respect to single input contribution:

$$\cdot \frac{\partial x_i^{(0)}}{\partial(((c_{ji}^{(0)}} + \ldots$$

$$\frac{\partial E}{\partial o_i^{(0)}} \cdot \cdots \cdots \cdot \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}}$$

$$\cdot \frac{\partial x_i^{(0)}}{\partial((((((((((c_{ji}^{(0)}))))))))))}$$

Term is just zero, of course. Then, plugging in...

$$\frac{\partial^2 E}{\partial(((((((((((c_{ij}^{(0)})))))))))))))^2} = \frac{\partial}{\partial c_{ij}^{(0)}} \left(o_i^{(0)} - t_i \right) \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} + \left(o_i^{(0)} - t_i \right) \frac{\partial}{\partial c_{ji}^{(0)}} \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} \tag{16}$$

$$= \left(\frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} \right)^2 + \left(o_i^{(0)} - t_i \right) \sigma'' \left(x_i^{(0)} \right) \tag{17}$$

$$= \left(\sigma' \left(x_i^{(0)} \right) \right)^2 + \left(o_i^{(0)} - t_i \right) \sigma'' \left(x_i^{(0)} \right) \tag{18}$$

Moving on to the next layer, we have:

$$\frac{\partial E}{\partial o_j^{(1)}} = \frac{\partial E}{\partial o_i^{(0)}} \cdot \cdots \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} \cdot \cdots \frac{\partial x_i^{(0)}}{\partial c_{ji}^{(0)}} \cdot \frac{\partial c_{ji}^{(0)}}{\partial o_j^{(1)}} = \left(o_i^{(0)} - t_i\right) \sigma'(x_i^{(0)}) w_{ji}^{(0)} \tag{19}$$

$$\begin{aligned}
\frac{\partial^2 E}{\partial((((o_j^{(1)})))))))))))))^2} &= \frac{\partial^2 E}{\partial(((((((((o_j^{(1)} \partial o_i^{(0)}} \cdot \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} \cdot \cdots \frac{\partial x_i^{(0)}}{\partial c_{ji}^{(0)}} \cdot \frac{\partial c_{ji}^{(0)}}{\partial o_j^{(1)}} + \frac{\partial E}{\partial o_i^{(0)}} \cdot \frac{\partial^2 o_i^{(0)}}{\partial o_j^{(1)} \partial x_i^{(0)}} \cdot \cdots \frac{\partial x_i^{(0)}}{\partial c_{ji}^{(0)}} \cdot \frac{\partial c_{ji}^{(0)}}{\partial o_j^{(1)}} \\
&\quad + \frac{\partial E}{\partial o_i^{(0)}} \cdot \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} \cdot \cdots \frac{\partial^2 x_i^{(0)}}{\partial o_j^{(1)} \partial c_{ji}^{(0)}} \cdot \frac{\partial c_{ji}^{(0)}}{\partial o_j^{(1)}} + \frac{\partial E}{\partial o_i^{(0)}} \cdot \frac{\partial o_i^{(0)}}{\partial x_i^{(0)}} \cdot \cdots \frac{\partial x_i^{(0)}}{\partial c_{ji}^{(0)}} \cdot \frac{\partial^2 c_{ji}^{(0)}}{\partial((((((((((o_j^{(1)}))))))))} && (20) \\
&= \left(\sigma'(x_i^{(0)})((((((w_{ji}^{(0)}\right)^2 + \left(o_i^{(0)} - t_i\right) \sigma''(x_i^{(0)})(((((((w_{ji}^{(0)})))))))))))))^2 && (21)
\end{aligned}$$