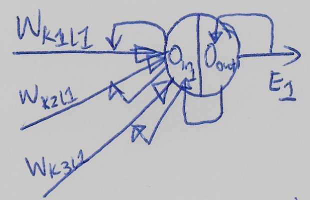
Backpropagating the error from the Output layer:



To propagate ~~$~~ change the weights from first hidden layer we need to calculate $\frac{\partial E_1}{\partial W_{k1l1}}$, we get this from using chain rule as follows:

$$\frac{\partial E_1}{\partial W_{k1l1}} = \frac{\partial E_1}{\partial O_{out1}} \cdot \frac{\partial O_{out1}}{\partial O_{in1}} \cdot \frac{\partial O_{in1}}{\partial W_{k1l1}}$$

, we need these for each weight in the hidden layer.

First, we will propagate error from output layer to the second hidden layer.

Calculate $\frac{\partial E_1}{\partial O_{out1}}$ as follows:

$$\frac{\partial E_1}{\partial O_{out1}} = \frac{\partial\left(-1*(y_1 \log(O_{out1}) + (1-y_1)*\log(1-O_{out1})\right)}{\partial O_{out1}}$$

$$\frac{\partial E_1}{\partial O_{out1}} = -1*y_1(1/O_{out1}) + (1-y_1)*(1/(1-O_{out1}))$$

For the other output nodes we can calculate as follows:

$$\begin{bmatrix} \frac{\partial E_1}{\partial O_{out1}} \\ \frac{\partial E_2}{\partial O_{out2}} \\ \frac{\partial E_3}{\partial O_{out3}} \end{bmatrix} = \begin{bmatrix} -1*(y_1(1/O_{out1}) + (1-y_1)*(1/(1-O_{out1})) \\ -1*(y_2(1/O_{out2}) + (1-y_2)*(1/(1-O_{out2})) \\ -1*(y_3(1/O_{out3}) + (1-y_3)*(1/(1-O_{out3})) \end{bmatrix} = \begin{bmatrix} -1(1(1/0.19858) + \overset{0}{\cancel{(1-1)*(1-0.19858)}}) \\ -1((0)\overset{0}{\cancel{(1/0.2856)}} + (1-0)(1/(1-0.2856))) \\ -1((0)\overset{0}{\cancel{(1/0.5158)}} + (1-0)(1/(1-0.5158))) \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial E_1}{\partial O_{out1}} \\ \frac{\partial E_2}{\partial O_{out2}} \\ \frac{\partial E_3}{\partial O_{out3}} \end{bmatrix} = \begin{bmatrix} -5.036 \\ -1.399 \\ -2.065 \end{bmatrix}$$