# HW 7

Bayard Walsh

May 2023

## 1

As two nodes ($i_1$ and $i_2$) are provided in the network, and the number of nodes equals the feature dimension, we have that feature dimension = **2**

## 2

We will find $o1$ and $o2$ as follows:

We find $net_{h1}$ and $net_{h2}$ below, following the approach outlined in the article:

$$net_{h1} = .15 * .05 + .2 * .1 + .35 * 1 = .3775$$
$$net_{h2} = .25 * .05 + .3 * .1 + .35 * 1 = .3925$$

Note that we can also compute these equations as matrix multiplications, done below

$$i = \begin{matrix} .05 \\ .10 \end{matrix}$$
$$h_1 = \begin{bmatrix} .15 & .2 \end{bmatrix}$$
$$h_2 = \begin{bmatrix} .25 & .3 \end{bmatrix}$$

Therefore we have

$$net_{h1} = h_1 \cdot i + b_1$$
$$net_{h2} = h_2 \cdot i + b_1$$

Now we find $out_{h1}$ and $out_{h2}$ below, following the approach outlined in the article with the logistics function:

$$out_{h1} = \frac{1}{1 + e^{-.3775}} = 0.593269992$$
$$out_{h2} = \frac{1}{1 + e^{-.3925}} = 0.596884378$$

Now we find $net_{o1}$ and $net_{o2}$ below, following the approach outlined in the article:

$$net_{o1} = .4 * 0.593269992 + .45 * 0.596884378 + .6 * 1 = 1.105905967$$

$$net_{o2} = .5 * 0.593269992 + .55 * 0.596884378 + .6 * 1 = 1.2249214039$$

Note that again we can also compute these equations as matrix multiplications, done below

$$h = \begin{matrix} .593269992 \\ .596884378 \end{matrix}$$

$$o_1 = \begin{bmatrix} .4 & .45 \end{bmatrix}$$

$$o_2 = \begin{bmatrix} .5 & .55 \end{bmatrix}$$

Therefore we have

$$net_{o1} = o_1 \cdot h + b_2$$

$$net_{o2} = o_2 \cdot h + b_2$$

Now we find $out_{o1}$ and $out_{o2}$ below, following the approach outlined in the article with the logistics function:

$$out_{o1} = \frac{1}{1 + e^{-1.105905967}} = 0.751365069541$$

$$out_{o2} = \frac{1}{1 + e^{-1.2249214039}} = 0.772928465$$

From these equations, we have our outputs for $o1, o2$, which are $out_{o1}, out_{o2}$

# 3

We can calculate the total squared error of the output of the forward pass as given by the article, using the output values we calculated in **section 2**

$$E_{total} = \Sigma \frac{1}{2}(target - output)^2$$

$$E_{o1} = \frac{1}{2}(.01 - 0.751365069541)^2 = 0.274811083168$$

$$E_{o2} = \frac{1}{2}(.99 - 0.772928465)^2 = 0.0235600256536$$

$$E_{total} = 0.274811083168 + 0.0235600256536 = 0.298371108822$$

Therefore we have

$$E_{total} = 0.298371108822$$

# 4

Following backward propagation, we will first compute $\frac{df}{dw_k}$ for $k = 5 \cdots 8$

From the chain rule, we have that

$$\frac{df}{dw_5} = \frac{dE_{total}}{dOut_{o1}} \cdot \frac{dOut_{o1}}{dNet_{o1}} \cdot \frac{dNet_{o1}}{dw_5}$$

First we find

$$\frac{dE_{total}}{dOut_{o1}}$$

and as we have the original equation of $E_{total}$ as:

$$\frac{1}{2}(target_{o1} - output_{o1})^2 + \frac{1}{2}(target_{o2} - output_{o2})^2$$

we have

$$\frac{dE_{total}}{dOut_{o1}} = (target_{o1} - output_{o1})^1 * -1 = output_{o1} - target_{o1}$$

Next,

$$out_{o1} = \frac{1}{1 + e^{-net_{o1}}}$$

Therefore

$$\frac{dOut_{o1}}{dNet_{o1}} = out_{o1}(1 - out_{o1})$$

Finally, we find

$$\frac{dNet_{o1}}{dw_5}$$

We have

$$Net_{o1} = w_5 \cdot out_{h1} + w_6 \cdot out_{h2} + b_2$$

Giving us:

$$\frac{dNet_{o1}}{dw_5} = out_{h1}$$

Therefore we can calculate

$$\frac{df}{dw_5} = (output_{o1} - target_{o1}) \cdot out_{o1}(1 - out_{o1}) \cdot out_{h1}$$

Now with substitution (based on our previous work) we have:

$$\frac{df}{dw_5} = (0.751365069541 - .01) \cdot 0.751365069541(1 - 0.751365069541) \cdot 0.593269992$$

$$= 0.0821670405506$$

I will now compute $\frac{df}{dw_k}$ in similar fashion for $k = 6$, noting when differences occur:

$$\frac{df}{dw_6} = \frac{dE_{total}}{dOut_{o1}} \cdot \frac{dOut_{o1}}{dNet_{o1}} \cdot \frac{dNet_{o1}}{dw_6}$$

This gives the formula below (note that because of the proximity of $w_5$ and $w_6$ the only difference in the derivative is $out_{h2}$ for the third term, which follows from how that term is computed):

$$\frac{df}{dw_6} = (output_{o1} - target_{o1}) \cdot out_{o1}(1 - out_{o1}) \cdot out_{h2}$$

Now with substitution (based on our previous work) we have:

$$\frac{df}{dw_6} = (0.751365069541 - .01) \cdot 0.751365069541(1 - 0.751365069541) \cdot 0.596884378$$

$$= 0.0826676278128$$

I will now compute $\frac{df}{dw_k}$ in similar fashion for $k = 7$, noting when differences occur. Here we will switch the derivatives for $o2$ instead of $o1$, but otherwise we can follow a similar approach. Note that:

$$\frac{df}{dw_7} = \frac{dE_{total}}{dOut_{o2}} \cdot \frac{dOut_{o2}}{dNet_{o2}} \cdot \frac{dNet_{o2}}{dw_6}$$

So therefore:

$$\frac{df}{dw_7} = (output_{o2} - target_{o2}) \cdot out_{o2}(1 - out_{o2}) \cdot out_{h1}$$

Now with substitution (based on our previous work) we have:

$$\frac{df}{dw_7} = (0.772928465 - .99) \cdot 0.772928465(1 - 0.772928465) \cdot 0.593269992$$

$$= -0.0226025404533$$

I will now compute $\frac{df}{dw_k}$ in similar fashion for $k = 8$, noting when differences occur. In this case, only the final term (like the difference between $k = 5$ and $k = 6$)

$$\frac{df}{dw_8} = (output_{o2} - target_{o2}) \cdot out_{o2}(1 - out_{o2}) \cdot out_{h2}$$

Therefore:

$$\frac{df}{dw_8} = (0.772928465 - .99) \cdot 0.772928465(1 - 0.772928465) \cdot 0.596884378$$

$$= -0.0227402422625$$

For $1 \cdots 4$, we will change our approach as we are computing an inner layer. We start with the following equation:

$$\frac{df}{dw_1} = \frac{dE_{total}}{dOut_{h1}} \cdot \frac{dOut_{h1}}{dNet_{h1}} \cdot \frac{dNet_{h1}}{dw_1}$$

Note that we must consider $h1$ through the inputs of $E_{o1}$ and $E_{o2}$ as well. Therefore we have

$$\frac{dE_{total}}{dOut_{h1}} = \frac{dE_{o1}}{dOut_{h1}} + \frac{dE_{o2}}{dOut_{h1}}$$

Which then gives us:

$$\frac{dE_{total}}{dOut_{h1}} = (\frac{dE_{o1}}{dOut_{o1}} \cdot \frac{dOut_{o1}}{dNet_{o1}} \cdot \frac{dNet_{o1}}{dOut_{h1}}) + (\frac{dE_{o2}}{dOut_{o2}} \cdot \frac{dOut_{o2}}{dNet_{o2}} \cdot \frac{dNet_{o2}}{dOut_{h1}})$$

We now substitute from previous computations:

$$\frac{dE_{total}}{dOut_{h1}} = (.74136507 \cdot 0.751365069541 \cdot (1-0.751365069541) \cdot .4) + (\frac{dE_{o2}}{dOut_{o2}} \cdot \frac{dOut_{o2}}{dNet_{o2}} \cdot .45)$$

$$\frac{dE_{total}}{dOut_{h1}} = 0.0553994246866 + (-.24119 \cdot 0.772928465(1-0.772928465) \cdot .45)$$

$$\frac{dE_{total}}{dOut_{h1}} = 0.0553994246866 + -0.019049119$$

$$\frac{dE_{total}}{dOut_{h1}} = 0.0363503056866$$

As we have computed the other derivatives previously, we have

$$\frac{df}{dw_1} = 0.0363503056866 * .241300709 * .05 = 0.000438567726727$$

Now that we have

$$\frac{dE_{total}}{dOut_{h1}}$$

we can easily calculate

$$\frac{df}{dw_3} = \frac{dE_{total}}{dOut_{h1}} \cdot \frac{dOut_{h1}}{dNet_{h1}} \cdot \frac{dNet_{h1}}{dw_3}$$

We have

$$\frac{df}{dw_3} = 0.0363503056866 * .136918 * .1 = 0.0004977$$

We will follow the same approach as above to calculate for the $w_2$, $w_4$

$$\frac{dE_{total}}{dOut_{h2}} = (\frac{dE_{o1}}{dOut_{o1}} \cdot \frac{dOut_{o1}}{dNet_{o1}} \cdot \frac{dNet_{o1}}{dOut_{h2}}) + (\frac{dE_{o2}}{dOut_{o2}} \cdot \frac{dOut_{o2}}{dNet_{o2}} \cdot \frac{dNet_{o2}}{dOut_{h2}})$$

We can substitute our previous work:

$$\frac{dE_{total}}{dOut_{h2}} = (.74136507 \cdot 0.751365069541 \cdot (1-0.751365069541) \cdot .50$$

$$+(-.15505 \cdot 0.772928465(1-0.772928465) \cdot .55$$

$$\frac{dE_{total}}{dOut_{h2}} = 0.0692492808582 - 0.0149670585441$$

$$\frac{dE_{total}}{dOut_{h2}} = .0567158$$

Now we have

$$\frac{df}{dw_2} = .0567158 \cdot (0.772928465(1 - 0.772928465)) * .05 = .000877714$$

Again, finding $w_4$ is easy given we have $w_2$:

$$\frac{df}{dw_4} = .0567158 \cdot (0.772928465(1 - 0.772928465)) * .1 = 0.00099542$$

Finally we have $df/d_{wk}$ $1 \cdots 8$ :

$$\frac{df}{dw_1} = 0.000438567726727$$

$$\frac{df}{dw_2} = 0.000877714$$

$$\frac{df}{dw_3} = 0.0004977$$

$$\frac{df}{dw_4} = 0.00099542$$

$$\frac{df}{dw_5} = 0.0821670405506$$

$$\frac{df}{dw_6} = 0.0826676278128$$

$$\frac{df}{dw_7} = -0.0226025404533$$

$$\frac{df}{dw_8} = -0.0227402422625$$

# 5

Now we will update each of the given $1 \cdots 8$ to find $w_k^+$ in terms of $\tau$

$$w_1^+ = .15 - \tau * 0.000438567726727$$

$$w_2^+ = .20 - \tau * 0.000877714$$

$$w_3^+ = .25 - \tau * 0.0004977$$

$$w_4^+ = .30 - \tau * 0.00099542$$

$$w_5^+ = .40 - \tau * 0.0821670405506$$

$$w_6^+ = .45 - \tau * 0.0826676278128$$

$$w_7^+ = .50 - \tau * -0.0226025404533$$

$$w_8^+ = .55 - \tau * -0.0227402422625$$