This file accompanies the excel sheet with neural network defined as discussed in the class and as per assignment 2 requirements. The group members are Ritambhra Korpal, Chaitanya Vanapamala, Pralay Ramteke, Pallavi. An effort has been made here to explain the steps while finding the derivatives.

To fine tune the parameters, so as to minimize the total error, the parameters are adjusted by propagating the gradient of error back from output layer towards the input layer, hence the name back propagation. The parameters are adjusted as

$w_i^{new} = w_i^{old} - \eta * \frac{E}{\partial w_i^{old}}$

where E is the total error and η is the learning rate.

To find the gradient in the error with respect to each parameter, it helps to have all the equations first in place. The names used here are as per the names used in the diagram drawn for the network in the excel sheet. So looking at that here are the equations:

$h_1 = w_1 \times i_1 + w_2 \times i_2$ ...(1)

$ah_1 = \sigma(h_1) = \frac{1}{1+e^{-h_1}}$ ....(2)

$h_2 = w_3 \times i_1 + w_4 \times i_2$ ....(3)

$ah_2 = \sigma(h_2) = \frac{1}{1+e^{-h_2}}$ ....(4)

$o_1 = w_5 \times ah_1 + w_6 \times ah_2$ ....(5)

$ao_1 = \sigma(o_1) = \frac{1}{1+e^{-o_1}}$ ....(6)

$o_2 = w_7 \times ah_1 + w_8 \times ah_2$ ....(7)

$ao_2 = \sigma(o_2) = \frac{1}{1+e^{-o_2}}$ ....(8)

$E_1 = \frac{1}{2}(t_1 - ao_1)^2$ ....(9)

$E_2 = \frac{1}{2}(t_2 - ao_2)^2$ ....(10)

E=E$_1 + E_2$

To find the derivative of the total error, with respect to each parameter, it can be seen from these equations, that it makes sense to find the derivative with respect to parameters which are closer to output layer, which can then be used to find the derivatives with respect to parameters farther (towards input layer) from output layer.
So this is how we start:

$\frac{\partial E}{\partial w_5} = \frac{\partial (E_1+E_2)}{\partial w_5}$

Since E_2 does not depend on w_5, $\frac{\partial E_2}{\partial w_5} = 0$

Therefore $\frac{\partial E}{\partial w_5} = \frac{\partial E_1}{\partial w_5}$

Now

$\frac{\partial E_1}{\partial w_5} = \frac{\partial E_1}{\partial ao_1} \times \frac{\partial ao_1}{\partial o_1} \times \frac{\partial o_1}{\partial w_5}$ ....(11)

$\frac{\partial E_1}{\partial ao_1} = -(t_1 - ao_1) = (ao_1 - t_1)$

$\frac{\partial ao_1}{\partial o_1} = \sigma(o_1) \times (1 - \sigma(o_1))$ [derivative of sigmoid function]

Substituting from equation (6) above for $\sigma(o_1)$, we get
$\frac{\partial ao_1}{\partial o_1} = ao_1 \times (1 - ao_1)$

$\frac{\partial o_1}{\partial w_5} = ah_1$

Substituting every thing together in equation(11), we get
$\frac{\partial E_1}{\partial w_5} = (ao_1 - t_1) \times ao_1 \times (1 - ao_1) \times ah_1$ ....(12)

On the similar lines we can say (and generalise) for derivative with respect to parameters w_6, w_7 and w_8 as below:
$\frac{\partial E_1}{\partial w_6} = (ao_1 - t_1) \times ao_1 \times (1 - ao_1) \times ah_2$ ....(13)

$\frac{\partial E_1}{\partial w_7} = (ao_2 - t_2) \times ao_2 \times (1 - ao_2) \times ah_1$ ....(14)

$\frac{\partial E_1}{\partial w_8} = (ao_2 - t_2) \times ao_2 \times (1 - ao_2) \times ah_2$ ....(15)

Now lets have a look at the partial derivatives with respect to parameter w_1:
so
$\frac{\partial E}{\partial w_1} = \frac{\partial(E_1+E_2)}{\partial w_1}$

which can written as
$\frac{\partial E}{\partial w_1} = \frac{\partial E_1}{\partial w_1} + \frac{\partial E_2}{\partial w_1}$ ....(16)

Now
$\frac{\partial E_1}{\partial w_1} = \frac{\partial E_1}{\partial ao_1} \times \frac{\partial ao_1}{\partial o_1} \times \frac{\partial o_1}{\partial ah_1} \times \frac{\partial ah_1}{\partial h_1} \times \frac{\partial h_1}{\partial w_1}$ ....(17)

While $\frac{\partial E_1}{\partial ao_1}$ and $\frac{\partial ao_1}{\partial o_1}$ are same as before, we need to look at $\frac{\partial o_1}{\partial ah_1}$, $\frac{\partial ah_1}{\partial h_1}$ and $\frac{\partial h_1}{\partial w_1}$

So looking at equations (5), (2) and (1), we get

$\frac{\partial o_1}{\partial ah_1} = w_5,$

$\frac{\partial ah_1}{\partial h_1} = ah_1 \times (1 - ah_1)$
and

$\frac{\partial h_1}{\partial w_1} = i_1$

Similarly, we can work out $\frac{\partial E_2}{\partial w_1}$ as
$\frac{\partial E_2}{\partial w_1} = \frac{\partial E_2}{\partial ao_2} \times \frac{\partial ao_2}{\partial o_2} \times \frac{\partial o_2}{\partial ah_1} \times \frac{\partial ah_1}{\partial h_1} \times \frac{\partial h_1}{\partial w_1}$ ....(18)

and
$\frac{\partial o_2}{\partial ah_1} = w_7$

Every thing else remaining the same.
Putting everything back together in equation (16) we get

$\frac{\partial E}{\partial w_1} = ((ao_1 - t_1) \times ao_1 \times (1 - ao_1) \times w_5 + (ao_2 - t_2) \times ao_2 \times (1 - ao_2) \times w_7) \times (ah_1 \times (1 - ah_1) \times i_1)$....(19)

On the similar lines we can write the partial derivatives with respect to w_2, w_3 and w_4 as be-