BMI 6015- Applied Machine Learning

Krithika Iyer
Department of Biomedical Informatics - University of Utah

November 2018

Neural Network Example

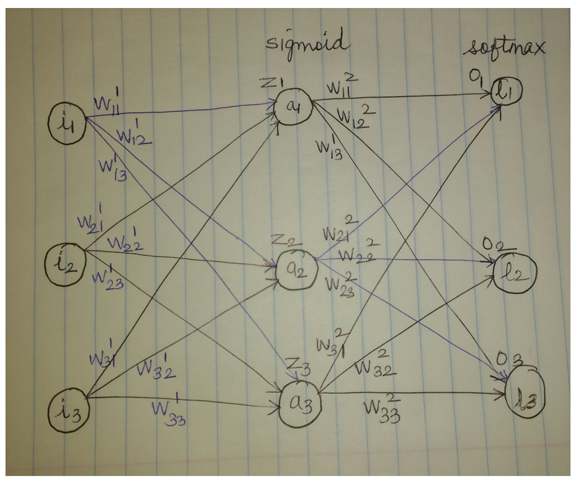

Figure 1: Neural Network

1. This is a classification problem where each set of input can either be class 1/2/3. Hence we have 3 neurons in the final layer. If the input belongs to class 1 then the expected output would be [1,0,0] if class =2 output = 0,1,0] ,if class 3 then [0,0,1].

2. (a) Weights in the neural network are represented as w^l_{jk} where j - is the neuron from which the line starts in layer $l-1$ and k - is the neuron on which the line ends at l layer.

   (b) For example,w^2_{11} tells us that the weight is associated with the line which shows connection from the first neuron in layer 1 and ends at the first neuron in layer 2.

(c) This notation is different from the one seen in the class where all the nodes were named with alphabets and the weight for the link from the first neuron A in layer 1 to first neuron D in layer 2 was written as w_{AD} here it is written as w^1_{11}.

(d) If we were two convert the exisiting notation to one used in class if would be $w_{i_1 a_1}$

3. This neural network has 3 inputs(3 features), one hidden layer and one output layer
4. The hidden layer has 3 neurons and each neuron is connected to all the three inputs and bias term (bias not shown in the image)
5. The activation function used in hidden layer is **SIGMOID**
6. The output layer has **SOFTMAX** activation.$https://en.wikipedia.org/wiki/Softmax_function$
7. The loss function used for this neural network is **CROSS-ENTROPY**
8. The layer-1 consists of $[a_1, a_2, a_3]$
9. The layer-2 consists of $[l_1, l_2, l_3]$

DERIVATIVES

$$Sigmoid = 1/(1 + e^{-x})$$

$$\frac{\partial(1/(1 + e^{-x}))}{\partial x} = 1/(1 + e^{-x}) \times (1 - 1/(1 + e^{-x}))$$

$$\frac{\partial Sigmoid}{\partial x} = Sigmoid \times (1 - Sigmoid)$$

Derivative of Sigmoid

$$Softmax = e^{x_a}/(\sum_{a=1}^{n} e^{x_a}) = e^{x_1}/(e^{x_1} + e^{x_2} + e^{x_3})$$

$$\frac{\partial(Softmax)}{\partial x_1} = (e^{x_1} \times (e^{x_2} + e^{x_3}))/(e^{x_1} + e^{x_2} + e^{x_3})^2$$

Derivative of Softmax

FORWARD PROPAGATION

Inputs : $[i_1 \quad i_2 \quad i_3] = [0.1 \quad 0.2 \quad 0.7]$
Actual Outputs : $[y_1 \quad y_2 \quad y_3] = [1 \quad 0 \quad 0]$
Layer 1: Neuron 1
$z_1 = w_{11}^1 i_1 + w_{21}^1 i_2 + w_{31}^1 i_3 + b_1^1$
$z_1 = 1.5$
$a_1 = sigmoid(z_1)$
$a_1 = \frac{1}{1+\exp(-z_1)}$
$a_1 = 0.8175$

Layer 1: Neuron 2
$z_2 = w_{12}^1 i_1 + w_{22}^1 i_2 + w_{32}^1 i_3 + b_2^1$
$z_2 = 1.41$
$a_2 = sigmoid(z_2)$
$a_2 = \frac{1}{1+\exp(-z_2)}$
$a_2 = 0.8037$

Layer 1: Neuron 3
$z_3 = w_{13}^1 i_1 + w_{23}^1 i_2 + w_{33}^1 i_3 + b_3^1$
$z_3 = 1.75$
$a_3 = sigmoid(z_3)$
$a_3 = \frac{1}{1+\exp(-z_3)}$
$a_3 = 0.8519$

Applying **SOFTMAX ACTIVATION** in the last layer
Layer 2 Neuron 1
$o_1 = w_{11}^2 a_1 + w_{21}^2 a_2 + w_{31}^2 a_3 + b_1^2$
$o_1 = 1.7488$
$l_1 = \frac{exp(o_1)}{\sum_{i=1}^{3} \exp(o_i)}$
$l_1 = 0.2143$

Layer 2: Neuron 2
$o_2 = w_{12}^2 a_1 + w_{22}^2 a_2 + w_{32}^2 a_3 + b_2^2$
$o_2 = 2.0599$
$l_2 = \frac{exp(o_2)}{\sum_{i=1}^{3} \exp(o_i)}$
$l_2 = 0.2926$

Layer 2: Neuron 3
$o_3 = w_{13}^2 a_1 + w_{23}^2 a_2 + w_{33}^2 a_3 + b_3^2$
$o_3 = 2.5814$
$l_3 = \frac{exp(o_3)}{\sum_{i=1}^{3} \exp(o_i)}$
$l_3 = 0.4929$

LOSS FUNCTION

Cross-Entropy Loss E for each output neuron is given by :

$$E = ylog(l_1) + (1-y)log(1-l_1)$$

Total loss is given by:

$$E = [E_1 + E_2 + E_3]$$

$E_1 = y_1log(l_1) + (1-y_1)log(1-l_1)$ = log(0.2143) = -1.540
$E_2 = y_2log(l_2) + (1-y_2)log(1-l_2)$ = log(1-0.2926) = -0.3461
$E_3 = y_3log(l_3) + (1-y_3)log(1-l_3)$ = log(1-0.4929) = -0.6790

$$E = \frac{1}{n}\sum_{i=1}^{3}[y_ilog(l_i) + (1-y_i)log(1-l_i)]$$

$E = (-1.540) + (-0.3461) + (-0.6790) = -2.56514$

BACK PROPAGATION

Layer -2 For weight - w_{11}^2:

$$\frac{\delta E_1}{\delta w_{11}^2} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta w_{11}^2}$$

First Term:

$$\frac{\delta E_1}{\delta l_1} = \frac{y_1}{l_1} - \frac{1-y_1}{1-l_1}$$

$y_1 = 1$ Hence the equation becomes :

$$\frac{\delta E_1}{\delta l_1} = \frac{1}{l_1}$$

Second Term:

$$\frac{\delta l_1}{\delta o_1} = \frac{exp(o_1)[exp(o_2) + exp(o_3)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2}$$

Third Term:

$$\frac{\delta o_1}{\delta w_{11}^2} = \frac{\delta[w_{11}^3a_1 + w_{21}^3a_2 + w_{31}^3a_3 + b_1^3]}{\delta w_{11}^2}$$

Everything is constant with respect to w_{11}^2. Hence the derivative is just the coefficient of w_{11}^2

$$\frac{\delta o_1}{\delta w_{11}^2} = a_1$$

Combining all the results from all 3 terms:

$$\frac{\delta E_1}{\delta w_{11}^2} = \frac{1}{l_1}\frac{exp(o_1)[exp(o_2) + exp(o_3)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2}a_1$$

$$\frac{\delta E_1}{\delta w_{11}^2} = 0.6421$$

The weight w_{11}^2 does not contribute to the loss function of the second and third neuron hence the loss E_2 and E_3 will not be affected by the change in w_{11}^2

$$\frac{\delta E_2}{\delta w_{11}^2} = \frac{\delta E_3}{\delta w_{11}^2} = 0$$

Therefore,

$$\frac{\delta E}{\delta w_{11}^2} = \frac{\delta E_1}{\delta w_{11}^2} + \frac{\delta E_2}{\delta w_{11}^2} + \frac{\delta E_3}{\delta w_{11}^2}$$

$$\frac{\delta E}{\delta w_{11}^2} = \frac{\delta E_1}{\delta w_{11}^2} = 0.6421$$

Following the same steps for the remaining weights,
For weight - w_{21}^2:

$$\frac{\delta E_1}{\delta w_{21}^2} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta w_{21}^2}$$

$$\frac{\delta E_1}{\delta l_1} = \frac{1}{l_1}$$

$$\frac{\delta l_1}{\delta o_1} = \frac{exp(o_1)[exp(o_2) + exp(0_3)]}{[\sum_{i=1}^{3} exp(o_i)]^2}$$

$$\frac{\delta o_1}{\delta w_{21}^2} = a_2$$

$$\frac{\delta E_2}{\delta w_{21}^2} = \frac{\delta E_3}{\delta w_{21}^2} = 0$$

$$\frac{\delta E}{\delta w_{21}^2} = \frac{\delta E_1}{\delta w_{21}^2} = 0.6313$$

For weight - w_{31}^2:

$$\frac{\delta E_1}{\delta w_{31}^2} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta w_{31}^2}$$

$$\frac{\delta E_1}{\delta l_1} = \frac{1}{l_1}$$

$$\frac{\delta l_1}{\delta o_2} = \frac{exp(o_1)[exp(o_2) + exp(0_3)]}{[\sum_{i=1}^{3} exp(o_i)]^2}$$

$$\frac{\delta o_2}{\delta w_{31}^2} = a_3$$

$$\frac{\delta E_2}{\delta w_{31}^2} = \frac{\delta E_3}{\delta w_{31}^2} = 0$$

$$\frac{\delta E}{\delta w_{31}^2} = \frac{\delta E_1}{\delta w_{31}^2} = 0.6691$$

For weight - w_{12}^2:

$$\frac{\delta E_1}{\delta w_{12}^2} = \frac{\delta E_3}{\delta w_{12}^2} = 0$$

$$\frac{\delta E_2}{\delta w_{12}^2} = \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta w_{12}^2}$$

$y_2 = 0$

$$\frac{\delta E_2}{\delta l_2} = \frac{-1}{(1 - l_2)}$$

$$\frac{\delta l_2}{\delta o_2} = \frac{exp(o_2)[exp(o_1) + exp(0_3)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2}$$

$$\frac{\delta o_2}{\delta w_{12}^2} = a_1$$

$$\frac{\delta E}{\delta w_{12}^2} = \frac{-1}{(1 - l_2)}\frac{exp(o_2)[exp(o_1) + exp(o_3)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2}a_1 = -0.2391$$

For weight w_{22}^2

$$\frac{\delta E}{\delta w_{22}^2} = \frac{-1}{(1 - l_2)}\frac{exp(o_2)[exp(o_1) + exp(o_3)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2}a_2 = -0.2350$$

For weight w_{32}^2

$$\frac{\delta E}{\delta w_{32}^2} = \frac{-1}{(1 - l_2)}\frac{exp(o_2)[exp(o_1) + exp(o_3)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2}a_3 = -0.2491$$

For weight w_{13}^2

$$\frac{\delta E}{\delta w_{13}^2} = \frac{-1}{(1 - l_3)}\frac{exp(o_3)[exp(o_1) + exp(o_2)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2}a_1 = -0.4027$$

For weight w_{23}^2

$$\frac{\delta E}{\delta w_{23}^2} = \frac{-1}{(1 - l_3)}\frac{exp(o_3)[exp(o_1) + exp(o_2)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2}a_2 = -0.3959$$

For weight w_{33}^2

$$\frac{\delta E}{\delta w_{33}^2} = \frac{-1}{(1 - l_3)}\frac{exp(o_3)[exp(o_1) + exp(o_2)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2}a_3 = -0.4197$$

Layer 1 :
For weight w_{11}^1

$$\frac{\delta E_1}{\delta w_{11}^1} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta w_{11}^1}$$

The first two terms are same as calculated earlier. For the third term we have :

$$\frac{\delta o_1}{\delta a_1} = \frac{\delta[w_{11}^2 a_1 + w_{21}^2 a_2 + w_{31}^2 a_3 + b_1^2]}{\delta a_1} = w_{11}^2$$

For fourth term we need to calculate the derivative of the sigmoid activation.

$$\frac{\delta a_1}{\delta z_1} = sigmoid(z)(1 - sigmoid(z)) = a_1(1 - a_1)$$

Last term,

$$\frac{\delta z_1}{\delta w_{11}^1} = \frac{\delta[w_{11}^1 i_1 + w_{21}^1 i_2 + w_{31}^1 i_3 + b_1^1]}{\delta w_{11}^1} = i_1$$

Therefore,

$$\frac{\delta E_1}{\delta w_{11}^1} = \frac{1}{l_1}\frac{exp(o_1)[exp(o_2) + exp(o_3)]}{[\sum_{i=1}^3 exp(o_i)]^2} w_{11}^2 a_1(1 - a_1) i_1 = 0.00117$$

Similarly,

$$\frac{\delta E_2}{\delta w_{11}^1} = \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta w_{11}^1}$$

$$\frac{\delta E_2}{\delta w_{11}^1} = \frac{-1}{1 - l_2}\frac{exp(o_2)[exp(o_1) + exp(o_3)]}{[\sum_{i=1}^3 exp(o_i)]^2} w_{12}^2 a_1(1 - a_1) i_1 = -0.00174$$

$$\frac{\delta E_3}{\delta w_{11}^1} = \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta w_{11}^1}$$

$$\frac{\delta E_3}{\delta w_{11}^1} = \frac{-1}{1 - l_3}\frac{exp(o_3)[exp(o_1) + exp(o_2)]}{[\sum_{i=1}^3 exp(o_i)]^2} w_{13}^2 a_1(1 - a_1) i_1 = -0.005880$$

Adding the three derivatives of E_1, E_2 and E_3

$$\frac{\delta E}{\delta w_{11}^1} = -0.00645$$

Similarly this can extended for other weights in layer 1.
For weight w_{12}^1

$$\frac{\delta E_1}{\delta w_{12}^1} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta w_{12}^1} = \frac{1}{l_1}\frac{exp(o_1)[exp(o_2) + exp(o_3)]}{[\sum_{i=1}^3 exp(o_i)]^2} w_{21}^2 a_2(1 - a_2) i_1 = 0.003717$$

$$\frac{\delta E_2}{\delta w_{12}^1} = \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta w_{12}^1} = \frac{-1}{1 - l_2}\frac{exp(o_2)[exp(o_1) + exp(o_3)]}{[\sum_{i=1}^3 exp(o_i)]^2} w_{22}^2 a_2(1 - a_2) i_1 = -0.00323$$

$$\frac{\delta E_3}{\delta w^1_{13}} = \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta w^1_{12}} = \frac{-1}{1-l_3}\frac{exp(o_3)[exp(o_1)+exp(o_2)]}{\left[\sum_{i=1}^3 exp(o_i)\right]^2}w^2_{23}a_2(1-a_2)i_1 = -0.00155$$

Adding the three derivatives of E_1, E_2 and E_3

$$\frac{\delta E}{\delta w^1_{12}} = -0.001063$$

For weight w^1_{13},

$$\frac{\delta E_1}{\delta w^1_{13}} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta w^1_{13}} = \frac{1}{l_1}\frac{exp(o_1)[exp(o_2)+exp(o_3)]}{\left[\sum_{i=1}^3 exp(o_i)\right]^2}w^2_{31}a_3(1-a_3)i_1 = 0.004955$$

$$\frac{\delta E_2}{\delta w^1_{13}} = \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta w^1_{13}} = \frac{-1}{1-l_2}\frac{exp(o_2)[exp(o_1)+exp(o_3)]}{\left[\sum_{i=1}^3 exp(o_i)\right]^2}w^2_{32}a_3(1-a_3)i_1 = -0.0007381$$

$$\frac{\delta E_3}{\delta w^1_{13}} = \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta w^1_{13}} = \frac{-1}{1-l_3}\frac{exp(o_3)[exp(o_1)+exp(o_2)]}{\left[\sum_{i=1}^3 exp(o_i)\right]^2}w^2_{33}a_3(1-a_3)i_1 = -0.005534$$

Adding the three derivatives of E_1, E_2 and E_3

$$\frac{\delta E}{\delta w^1_{13}} = -0.0013171$$

For weight w^1_{21},

$$\frac{\delta E_1}{\delta w^1_{21}} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta w^1_{21}} = \frac{1}{l_1}\frac{exp(o_1)[exp(o_2)+exp(o_3)]}{\left[\sum_{i=1}^3 exp(o_i)\right]^2}w^2_{11}a_1(1-a_1)i_2 = 0.002343$$

$$\frac{\delta E_2}{\delta w^1_{21}} = \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta w^1_{21}} = \frac{-1}{1-l_2}\frac{exp(o_2)[exp(o_1)+exp(o_3)]}{\left[\sum_{i=1}^3 exp(o_i)\right]^2}w^2_{12}a_1(1-a_1)i_2 = -0.003491$$

$$\frac{\delta E_3}{\delta w^1_{21}} = \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta w^1_{21}} = \frac{-1}{1-l_3}\frac{exp(o_3)[exp(o_1)+exp(o_2)]}{\left[\sum_{i=1}^3 exp(o_i)\right]^2}w^2_{13}a_1(1-a_1)i_2 = -0.01176$$

Adding the three derivatives of E_1, E_2 and E_3

$$\frac{\delta E}{\delta w^1_{21}} = -0.0129$$

For weight w^1_{22},

$$\frac{\delta E_1}{\delta w^1_{22}} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta w^1_{22}} = \frac{1}{l_1}\frac{exp(o_1)[exp(o_2)+exp(o_3)]}{\left[\sum_{i=1}^3 exp(o_i)\right]^2}w^2_{21}a_2(1-a_2)i_2 = 0.007435$$

$$\frac{\delta E_2}{\delta w^1_{22}} = \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta w^1_{22}} = \frac{-1}{1-l_2}\frac{exp(o_2)[exp(o_1)+exp(o_3)]}{\left[\sum_{i=1}^3 exp(o_i)\right]^2}w^2_{22}a_2(1-a_2)i_2 = -0.00646$$

$$\frac{\delta E_3}{\delta w^1_{22}} = \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta w^1_{22}} = \frac{-1}{1-l_3}\frac{exp(o_3)[exp(o_1)+exp(o_2)]}{\left[\sum_{i=1}^3 exp(o_i)\right]^2}w^2_{23}a_2(1-a_2)i_2 = -0.003109$$

Adding the three derivatives of E_1, E_2 and E_3

$$\frac{\delta E}{\delta w^1_{22}} = -0.002134$$

For weight w^1_{23},

$$\frac{\delta E_1}{\delta w^1_{13}} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta w^1_{23}} = \frac{1}{l_1}\frac{exp(o_1)[exp(o_2)+exp(o_3)]}{[\sum_{i=1}^{3} exp(o_i)]^2} w^2_{31} a_3(1-a_3) i_2 = 0.002123$$

$$\frac{\delta E_2}{\delta w^1_{23}} = \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta w^1_{23}} = \frac{-1}{1-l_2}\frac{exp(o_2)[exp(o_1)+exp(o_3)]}{[\sum_{i=1}^{3} exp(o_i)]^2} w^2_{32} a_3(1-a_3) i_2 = -0.001476$$

$$\frac{\delta E_3}{\delta w^1_{23}} = \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta w^1_{23}} = \frac{-1}{1-l_3}\frac{exp(o_3)[exp(o_1)+exp(o_2)]}{[\sum_{i=1}^{3} exp(o_i)]^2} w^2_{33} a_3(1-a_3) i_2 = -0.01118$$

Adding the three derivatives of E_1, E_2 and E_3

$$\frac{\delta E}{\delta w^1_{23}} = -0.010533$$

For weight w^1_{31},

$$\frac{\delta E_1}{\delta w^1_{31}} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta w^1_{31}} = \frac{1}{l_1}\frac{exp(o_1)[exp(o_2)+exp(o_3)]}{[\sum_{i=1}^{3} exp(o_i)]^2} w^2_{11} a_1(1-a_1) i_3 = 0.0082065$$

$$\frac{\delta E_2}{\delta w^1_{31}} = \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta w^1_{31}} = \frac{-1}{1-l_2}\frac{exp(o_2)[exp(o_1)+exp(o_3)]}{[\sum_{i=1}^{3} exp(o_i)]^2} w^2_{12} a_1(1-a_1) i_3 = -0.01221$$

$$\frac{\delta E_3}{\delta w^1_{31}} = \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta w^1_{31}} = \frac{-1}{1-l_3}\frac{exp(o_3)[exp(o_1)+exp(o_2)]}{[\sum_{i=1}^{3} exp(o_i)]^2} w^2_{13} a_1(1-a_1) i_3 = -0.04116$$

Adding the three derivatives of E_1, E_2 and E_3

$$\frac{\delta E}{\delta w^1_{31}} = -0.04516$$

For weight w^1_{32},

$$\frac{\delta E_1}{\delta w^1_{32}} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta w^1_{32}} = \frac{1}{l_1}\frac{exp(o_1)[exp(o_2)+exp(o_3)]}{[\sum_{i=1}^{3} exp(o_i)]^2} w^2_{21} a_2(1-a_2) i_3 = 0.02602$$

$$\frac{\delta E_2}{\delta w^1_{32}} = \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta w^1_{32}} = \frac{-1}{1-l_2}\frac{exp(o_2)[exp(o_1)+exp(o_3)]}{[\sum_{i=1}^{3} exp(o_i)]^2} w^2_{22} a_2(1-a_2) i_3 = -0.02261$$

$$\frac{\delta E_3}{\delta w^1_{32}} = \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta w^1_{32}} = \frac{-1}{1-l_3}\frac{exp(o_3)[exp(o_1)+exp(o_2)]}{[\sum_{i=1}^{3} exp(o_i)]^2} w^2_{23} a_2(1-a_2) i_3 = -0.01080$$

Adding the three derivatives of E_1, E_2 and E_3

$$\frac{\delta E}{\delta w^1_{32}} = -0.00738$$

For weight w^1_{33},

$$\frac{\delta E_1}{\delta w^1_{33}} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta w^1_{33}} = \frac{1}{l_1}\frac{exp(o_1)[exp(o_2)+exp(o_3)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2} w^2_{31} a_3(1-a_3) i_3 = 0.03468$$

$$\frac{\delta E_2}{\delta w^1_{33}} = \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta w^1_{33}} = \frac{-1}{1-l_2}\frac{exp(o_2)[exp(o_1)+exp(o_3)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2} w^2_{32} a_3(1-a_3) i_3 = -0.0051668$$

$$\frac{\delta E_3}{\delta w^1_{33}} = \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta w^1_{33}} = \frac{-1}{1-l_3}\frac{exp(o_3)[exp(o_1)+exp(o_2)]}{\left[\sum_{i=1}^{3} exp(o_i)\right]^2} w^2_{33} a_3(1-a_3) i_3 = -0.03916$$

Adding the three derivatives of E_1, E_2 and E_3

$$\frac{\delta E}{\delta w^1_{33}} = -0.009646$$

Weights	Updates	Weights	Updates
w^1_{11}	$w^1_{11} - \mu\frac{\delta E}{\delta w^1_{11}} = 0.20645$	w^2_{11}	$w^2_{11} - \mu\frac{\delta E}{\delta w^2_{11}} = -0.5421$
w^1_{12}	$w^1_{12} - \mu\frac{\delta E}{\delta w^1_{12}} = 0.301063$	w^2_{12}	$w^2_{12} - \mu\frac{\delta E}{\delta w^2_{12}} = 0.6391$
w^1_{13}	$w^1_{13} - \mu\frac{\delta E}{\delta w^1_{13}} = 0.501317$	w^2_{13}	$w^2_{13} - \mu\frac{\delta E}{\delta w^2_{13}} = 1.2027$
w^1_{21}	$w^1_{21} - \mu\frac{\delta E}{\delta w^1_{21}} = 0.30129$	w^2_{21}	$w^2_{21} - \mu\frac{\delta E}{\delta w^2_{21}} = -0.3313$
w^1_{22}	$w^1_{22} - \mu\frac{\delta E}{\delta w^1_{22}} = 0.502134$	w^2_{22}	$w^2_{22} - \mu\frac{\delta E}{\delta w^2_{22}} = 0.935$
w^1_{23}	$w^1_{23} - \mu\frac{\delta E}{\delta w^1_{23}} = 0.7010553$	w^2_{23}	$w^2_{23} - \mu\frac{\delta E}{\delta w^2_{23}} = 0.5959$
w^1_{31}	$w^1_{31} - \mu\frac{\delta E}{\delta w^1_{31}} = 0.604516$	w^2_{31}	$w^2_{31} - \mu\frac{\delta E}{\delta w^2_{31}} = -0.1691$
w^1_{32}	$w^1_{32} - \mu\frac{\delta E}{\delta w^1_{32}} = 0.40738$	w^2_{32}	$w^2_{32} - \mu\frac{\delta E}{\delta w^2_{32}} = 0.4491$
w^1_{33}	$w^1_{33} - \mu\frac{\delta E}{\delta w^1_{33}} = 0.809646$	w^2_{33}	$w^2_{33} - \mu\frac{\delta E}{\delta w^2_{33}} = 1.3197$

For bias terms Layer 2 :

b^2_1

$$\frac{\delta E}{\delta b^2_1} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta b^2_1}$$

First two terms are same as calculated in the first step of back propagation while calculating $\frac{\delta E}{\delta w^2_{11}}$ For the last term:

$$\frac{\delta o_1}{\delta b^2_1} = \frac{\delta[w^2_{11}a_1 + w^2_{21}a_2 + w^2_{31}a_3 + b^2_1]}{\delta b^2_1} = 1$$

All terms are constant with respect to b_1^2. And E_2, E_3 are not affected by the change is b_1^2 hence the derivatives are zero.

$$\frac{\delta E}{\delta b_1^2} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1} = \frac{1}{l_1}\frac{exp(o_1)[exp(o_2)+exp(o_3)]}{\sum_{i=1}^{3} exp(o_i)} = 0.7855$$

Similarly for b_2^2 and b_3^2 we get

$$\frac{\delta E}{\delta b_2^2} = \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2} = \frac{-1}{1-l_2}\frac{exp(o_2)[exp(o_1)+exp(o_3)]}{\sum_{i=1}^{3} exp(o_i)} = -0.2925$$

$$\frac{\delta E}{\delta b_3^2} = \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3} = \frac{-1}{1-l_3}\frac{exp(o_3)[exp(o_1)+exp(o_2)]}{\sum_{i=1}^{3} exp(o_i)} = -0.4927$$

For bias terms Layer 1 :
b_1^1

$$\frac{\delta E}{\delta b_1^1} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta b_1^1} + \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta b_1^1} + \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_1}\frac{\delta a_1}{\delta z_1}\frac{\delta z_1}{\delta b_1^1}$$

All the terms in each of the derivatives for E_1, E_2 and E_3 except the last term have already been calculated.
For the last term:

$$\frac{\delta z_1}{\delta b_1^1} = 1$$

Therefore now we have ,

$$\frac{\delta E}{\delta b_1^1} = ++$$

For b_2^1,

$$\frac{\delta E}{\delta b_2^1} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta b_2^1} + \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta b_2^1} + \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_2}\frac{\delta a_2}{\delta z_2}\frac{\delta z_2}{\delta b_2^1}$$

$$\frac{\delta z_2}{\delta b_2^1} = 1$$

Therefore,

$$\frac{\delta E}{\delta b_2^1} =$$

For b_3^1,

$$\frac{\delta E}{\delta b_3^1} = \frac{\delta E_1}{\delta l_1}\frac{\delta l_1}{\delta o_1}\frac{\delta o_1}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta b_3^1} + \frac{\delta E_2}{\delta l_2}\frac{\delta l_2}{\delta o_2}\frac{\delta o_2}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta b_3^1} + \frac{\delta E_3}{\delta l_3}\frac{\delta l_3}{\delta o_3}\frac{\delta o_3}{\delta a_3}\frac{\delta a_3}{\delta z_3}\frac{\delta z_3}{\delta b_3^1}$$

$$\frac{\delta z_3}{\delta b_3^1} = 1$$

Therefore,

$$\frac{\delta E}{\delta b_3^1} =$$

Bias	Updates
b_1^1	$b_1^1 - \mu\frac{\delta E}{\delta b_1^1}$
b_2^1	$b_2^1 - \mu\frac{\delta E}{\delta b_2^1}$
b_3^1	$b_3^1 - \mu\frac{\delta E}{\delta b_2^1}$

Bias	Updates
b_1^2	$b_1^2 - \mu\frac{\delta E}{\delta b_1^2} = 0.2145$
b_2^2	$b_2^2 - \mu\frac{\delta E}{\delta b_2^2} = 1.2925$
b_3^2	$b_3^2 - \mu\frac{\delta E}{\delta b_2^2} = 1.49827$

REFERENCES

1. https://medium.com/@14prakash/back-propagation-is-very-simple-who-made-it-complicated-97b794c97e5c

2. https://www.edureka.co/blog/backpropagation/

NOTE

1. The notations in this example are different from that followed in the class,if any doubts please email me and Prof.Sameer.

2. If you find any calculation mistake in the solution here please let me know.

3. All logarithms are to the base **'e'**