Introduction to Machine Learning

Homework 1

Nuray Akar 2380004

November 2022

1 Part 1

1.1 Regression

We know that the generic weight update rule is as follows:

$$\omega = \omega - \alpha \frac{\partial E(x)}{\partial \omega}$$

The update of weights for the regression is like:

$$a_{k0}^1 = a_{k0}^1 - \alpha \frac{\partial SEE(y, O_0^2)}{\partial a_{k0}^1}$$

For k=0;

$$a_{00}^1 = a_{00}^1 - \alpha \frac{\partial SEE(y, O_0^2)}{\partial a_{00}^1}$$

$$\frac{\partial SEE(y, O_0^2)}{\partial a_{00}^1} = \frac{\partial SEE(y, O_0^2)}{\partial O_0^2} . \frac{\partial O_0^2}{\partial a_{00}^1} = -2(y - O_0^2).O_0^1$$

Since O_0^1 is given as 1;

$$\frac{\partial SEE(y, O_0^2)}{\partial a_{00}^1} = -2(y - O_0^2)$$

As a result the weight update equation is equal to:

$$a_{00}^1 = a_{00}^1 + \alpha \cdot 2(y - O_0^2)$$

For k=1;

$$a_{10}^1 = a_{10}^1 - \alpha \frac{\partial SEE(y, O_0^2)}{\partial a_{10}^1}$$

$$\frac{\partial SEE(y, O_0^2)}{\partial a_{10}^1} = \frac{\partial SEE(y, O_0^2)}{\partial O_0^2} . \frac{\partial O_0^2}{\partial a_{10}^1} = -2(y - O_0^2).O_1^1$$

We know that:

$$O_1^1 = \sigma(\sum_{i=0} O_i^0 * a_{i1}^0) = \frac{1}{1+e^{-(a_{01}^0 + O_1^0 a_{11}^0 + O_2^0 a_{21}^0)}}$$

As a result the weight update equation is equal to:

$$a_{10}^1 = a_{10}^1 + \alpha \cdot 2(y - O_0^2)\frac{1}{1+e^{-(a_{01}^0 + O_1^0 a_{11}^0 + O_2^0 a_{21}^0)}}$$

For k=2;

$$a_{20}^1 = a_{20}^1 - \alpha\frac{\partial SEE(y, O_0^2)}{\partial a_{20}^1}$$

$$\frac{\partial SEE(y, O_0^2)}{\partial a_{20}^1} = \frac{\partial SEE(y, O_0^2)}{\partial O_0^2}.\frac{\partial O_0^2}{\partial a_{20}^1} = -2(y - O_0^2).O_2^1$$

We know that:

$$O_2^1 = \frac{1}{1+e^{-(a_{02}^0 + O_1^0 a_{12}^0 + O_2^0 a_{22}^0)}}$$

As a result the weight update equation is equal to:

$$a_{20}^1 = a_{20}^1 + \alpha \cdot 2(y - O_0^2)\frac{1}{1+e^{-(a_{02}^0 + O_1^0 a_{12}^0 + O_2^0 a_{22}^0)}}$$

For k=3;

$$a_{30}^1 = a_{30}^1 - \alpha\frac{\partial SEE(y, O_0^2)}{\partial a_{30}^1}$$

$$\frac{\partial SEE(y, O_0^2)}{\partial a_{30}^1} = \frac{\partial SEE(y, O_0^2)}{\partial O_0^2}.\frac{\partial O_0^2}{\partial a_{30}^1} = -2(y - O_0^2).O_3^1$$

We know that:

$$O_3^1 = \frac{1}{1+e^{-(a_{03}^0 + O_1^0 a_{13}^0 + O_2^0 a_{23}^0)}}$$

As a result the weight update equation is equal to:

$$a_{30}^1 = a_{30}^1 + \alpha \cdot 2(y - O_0^2)\frac{1}{1+e^{-(a_{03}^0 + O_1^0 a_{13}^0 + O_2^0 a_{23}^0)}}$$

For the first layer the update of weights is like:

$$a_{ik}^0 = a_{ik}^0 - \alpha\frac{\partial SEE(y, O_0^2)}{\partial a_{ik}^0}$$

PS: x_i^1 is the input of $O_i^1 = \sigma(x_i^1) = \frac{1}{1+e^{-(x_i^1)}}$
For i = 0:
For k=1:

$$a_{01}^0 = a_{01}^0 - \alpha\frac{\partial SEE(y, O_0^2)}{\partial a_{01}^0}$$

$$\frac{\partial SEE(y,O_0^2)}{\partial a_{01}^0} = \frac{\partial SEE(y,O_0^2)}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial O_1^1} \frac{\partial O_1^1}{\partial x_1^1} \frac{\partial x_1^1}{\partial a_{01}^0}$$

$$= -2(y-O_0^2) \cdot a_{10}^1 \cdot (\frac{1}{1+e^{-(a_{01}^0+O_1^0a_{11}^0+O_2^0a_{21}^0)}}) \cdot (1-\frac{1}{1+e^{-(a_{01}^0+O_1^0a_{11}^0+O_2^0a_{21}^0)}})$$

Therefore:

$$a_{01}^0 = a_{01}^0 + \alpha \cdot 2(y-O_0^2) \cdot a_{10}^1 \cdot (\frac{1}{1+e^{-(a_{01}^0+O_1^0a_{11}^0+O_2^0a_{21}^0)}}) \cdot (1-\frac{1}{1+e^{-(a_{01}^0+O_1^0a_{11}^0+O_2^0a_{21}^0)}})$$

For k=2:

$$a_{02}^0 = a_{02}^0 - \alpha \frac{\partial SEE(y,O_0^2)}{\partial a_{02}^0}$$

$$\frac{\partial SEE(y,O_0^2)}{\partial a_{02}^0} = \frac{\partial SEE(y,O_0^2)}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial O_2^1} \frac{\partial O_2^1}{\partial x_2^1} \frac{\partial x_2^1}{\partial a_{02}^0}$$

$$= -2(y-O_0^2) \cdot a_{20}^1 \cdot (\frac{1}{1+e^{-(a_{02}^0+O_1^0a_{12}^0+O_2^0a_{22}^0)}}) \cdot (1-\frac{1}{1+e^{-(a_{02}^0+O_1^0a_{12}^0+O_2^0a_{22}^0)}})$$

Therefore:

$$a_{02}^0 = a_{02}^0 + \alpha \cdot 2(y-O_0^2) \cdot a_{20}^1 \cdot (\frac{1}{1+e^{-(a_{02}^0+O_1^0a_{12}^0+O_2^0a_{22}^0)}}) \cdot (1-\frac{1}{1+e^{-(a_{02}^0+O_1^0a_{12}^0+O_2^0a_{22}^0)}})$$

For k=3:

$$a_{03}^0 = a_{03}^0 - \alpha \frac{\partial SEE(y,O_0^2)}{\partial a_{03}^0}$$

$$\frac{\partial SEE(y,O_0^2)}{\partial a_{03}^0} = \frac{\partial SEE(y,O_0^2)}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial O_3^1} \frac{\partial O_3^1}{\partial x_3^1} \frac{\partial x_3^1}{\partial a_{03}^0}$$

$$= -2(y-O_0^2) \cdot a_{30}^1 \cdot (\frac{1}{1+e^{-(a_{03}^0+O_1^0a_{13}^0+O_2^0a_{23}^0)}}) \cdot (1-\frac{1}{1+e^{-(a_{03}^0+O_1^0a_{13}^0+O_2^0a_{23}^0)}})$$

Therefore:

$$a_{03}^0 = a_{03}^0 + \alpha \cdot 2(y-O_0^2) \cdot a_{30}^1 \cdot (\frac{1}{1+e^{-(a_{03}^0+O_1^0a_{13}^0+O_2^0a_{23}^0)}}) \cdot (1-\frac{1}{1+e^{-(a_{03}^0+O_1^0a_{13}^0+O_2^0a_{23}^0)}})$$

For i=1:
For k=1:

$$a_{11}^0 = a_{11}^0 - \alpha \frac{\partial SEE(y,O_0^2)}{\partial a_{11}^0}$$

$$\frac{\partial SEE(y,O_0^2)}{\partial a_{11}^0} = \frac{\partial SEE(y,O_0^2)}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial O_1^1} \frac{\partial O_1^1}{\partial x_1^1} \frac{\partial x_1^1}{\partial a_{11}^0}$$

$$= -2(y-O_0^2) \cdot a_{10}^1 \cdot (\frac{1}{1+e^{-(a_{01}^0+O_1^0a_{11}^0+O_2^0a_{21}^0)}}) \cdot (1-\frac{1}{1+e^{-(a_{01}^0+O_1^0a_{11}^0+O_2^0a_{21}^0)}}) \cdot O_1^0$$

Therefore:

$$a_{11}^0 = a_{11}^0 + \alpha \cdot 2(y-O_0^2) \cdot a_{10}^1 \cdot (\frac{1}{1+e^{-(a_{01}^0+O_1^0a_{11}^0+O_2^0a_{21}^0)}}) \cdot (1-\frac{1}{1+e^{-(a_{01}^0+O_1^0a_{11}^0+O_2^0a_{21}^0)}}) \cdot O_1^0$$

For k=2:

$$a_{12}^0 = a_{12}^0 - \alpha \frac{\partial SEE(y, O_0^2)}{\partial a_{12}^0}$$

$$\frac{\partial SEE(y, O_0^2)}{\partial a_{12}^0} = \frac{\partial SEE(y, O_0^2)}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial O_2^1} \frac{\partial O_2^1}{\partial x_2^1} \frac{\partial x_2^1}{\partial a_{12}^0}$$

$$= -2(y - O_0^2) \cdot a_{20}^1 \cdot \left(\frac{1}{1 + e^{-(a_{02}^0 + O_1^0 a_{12}^0 + O_2^0 a_{22}^0)}}\right) \cdot \left(1 - \frac{1}{1 + e^{-(a_{02}^0 + O_1^0 a_{12}^0 + O_2^0 a_{22}^0)}}\right) \cdot O_1^0$$

Therefore:

$$a_{12}^0 = a_{12}^0 + \alpha \cdot 2(y - O_0^2) \cdot a_{20}^1 \cdot \left(\frac{1}{1 + e^{-(a_{02}^0 + O_1^0 a_{12}^0 + O_2^0 a_{22}^0)}}\right) \cdot \left(1 - \frac{1}{1 + e^{-(a_{02}^0 + O_1^0 a_{12}^0 + O_2^0 a_{22}^0)}}\right) \cdot O_1^0$$

For k=3:

$$a_{13}^0 = a_{13}^0 - \alpha \frac{\partial SEE(y, O_0^2)}{\partial a_{13}^0}$$

$$\frac{\partial SEE(y, O_0^2)}{\partial a_{13}^0} = \frac{\partial SEE(y, O_0^2)}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial O_3^1} \frac{\partial O_3^1}{\partial x_3^1} \frac{\partial x_3^1}{\partial a_{13}^0}$$

$$= -2(y - O_0^2) \cdot a_{30}^1 \cdot \left(\frac{1}{1 + e^{-(a_{03}^0 + O_1^0 a_{13}^0 + O_2^0 a_{23}^0)}}\right) \cdot \left(1 - \frac{1}{1 + e^{-(a_{03}^0 + O_1^0 a_{13}^0 + O_2^0 a_{23}^0)}}\right) \cdot O_1^0$$

Therefore:

$$a_{13}^0 = a_{13}^0 + \alpha \cdot 2(y - O_0^2) \cdot a_{30}^1 \cdot \left(\frac{1}{1 + e^{-(a_{03}^0 + O_1^0 a_{13}^0 + O_2^0 a_{23}^0)}}\right) \cdot \left(1 - \frac{1}{1 + e^{-(a_{03}^0 + O_1^0 a_{13}^0 + O_2^0 a_{23}^0)}}\right) \cdot O_1^0$$

For i=2:
For k=1:

$$a_{21}^0 = a_{21}^0 - \alpha \frac{\partial SEE(y, O_0^2)}{\partial a_{21}^0}$$

$$\frac{\partial SEE(y, O_0^2)}{\partial a_{21}^0} = \frac{\partial SEE(y, O_0^2)}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial O_1^1} \frac{\partial O_1^1}{\partial x_1^1} \frac{\partial x_1^1}{\partial a_{21}^0}$$

$$= -2(y - O_0^2) \cdot a_{10}^1 \cdot \left(\frac{1}{1 + e^{-(a_{01}^0 + O_1^0 a_{11}^0 + O_2^0 a_{21}^0)}}\right) \cdot \left(1 - \frac{1}{1 + e^{-(a_{01}^0 + O_1^0 a_{11}^0 + O_2^0 a_{21}^0)}}\right) \cdot O_2^0$$

Therefore:

$$a_{21}^0 = a_{21}^0 + \alpha \cdot 2(y - O_0^2) \cdot a_{10}^1 \cdot \left(\frac{1}{1 + e^{-(a_{01}^0 + O_1^0 a_{11}^0 + O_2^0 a_{21}^0)}}\right) \cdot \left(1 - \frac{1}{1 + e^{-(a_{01}^0 + O_1^0 a_{11}^0 + O_2^0 a_{21}^0)}}\right) \cdot O_2^0$$

For k=2:

$$a_{22}^0 = a_{22}^0 - \alpha \frac{\partial SEE(y, O_0^2)}{\partial a_{22}^0}$$

$$\frac{\partial SEE(y, O_0^2)}{\partial a_{22}^0} = \frac{\partial SEE(y, O_0^2)}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial O_2^1} \frac{\partial O_2^1}{\partial x_2^1} \frac{\partial x_2^1}{\partial a_{22}^0}$$

$$= -2(y - O_0^2) \cdot a_{20}^1 \cdot \left(\frac{1}{1 + e^{-(a_{02}^0 + O_1^0 a_{12}^0 + O_2^0 a_{22}^0)}}\right) \cdot \left(1 - \frac{1}{1 + e^{-(a_{02}^0 + O_1^0 a_{12}^0 + O_2^0 a_{22}^0)}}\right) \cdot O_2^0$$

Therefore:

$$a_{22}^0 = a_{22}^0 + \alpha \cdot 2(y - O_0^2) \cdot a_{20}^1 \cdot (\frac{1}{1 + e^{-(a_{02}^0 + O_1^0 a_{12}^0 + O_2^0 a_{22}^0)}}) \cdot (1 - \frac{1}{1 + e^{-(a_{02}^0 + O_1^0 a_{12}^0 + O_2^0 a_{22}^0)}}) \cdot O_2^0$$

For k=3:

$$a_{23}^0 = a_{23}^0 - \alpha \frac{\partial SEE(y, O_0^2)}{\partial a_{23}^0}$$

$$\frac{\partial SEE(y, O_0^2)}{\partial a_{23}^0} = \frac{\partial SEE(y, O_0^2)}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial O_3^1} \frac{\partial O_3^1}{\partial x_3^1} \frac{\partial x_3^1}{\partial a_{23}^0}$$

$$= 42(y - O_0^2) \cdot a_{30}^1 \cdot (\frac{1}{1 + e^{-(a_{03}^0 + O_1^0 a_{13}^0 + O_2^0 a_{23}^0)}}) \cdot (1 - \frac{1}{1 + e^{-(a_{03}^0 + O_1^0 a_{13}^0 + O_2^0 a_{23}^0)}}) \cdot O_2^0$$

Therefore:

$$a_{23}^0 = a_{23}^0 + \alpha \cdot 2(y - O_0^2) \cdot a_{30}^1 \cdot (\frac{1}{1 + e^{-(a_{03}^0 + O_1^0 a_{13}^0 + O_2^0 a_{23}^0)}}) \cdot (1 - \frac{1}{1 + e^{-(a_{03}^0 + O_1^0 a_{13}^0 + O_2^0 a_{23}^0)}}) \cdot O_2^0$$

1.2 Classification

We know that the generic weight update rule is as follows:

$$a^1_{kn} = a^1_{kn} - \alpha \frac{\partial CE(l,l')}{\partial a^1_{kn}}$$

$$CE(l,l') = -\sum_i l_i * log(l'_i)$$

PS: x^1_i is the input of $O^1_i = \sigma(x^1_i) = \frac{1}{1+e^{-(x^1_i)}}$ and x^2_i is the input of $O^2_i = softmax(x^2_i)$
For k=0:
For n=0:

$$a^1_{00} = a^1_{00} - \alpha \frac{\partial CE(l,l')}{\partial a^1_{00}}$$

$$\frac{\partial CE(l,l')}{\partial a^1_{00}} = \frac{\partial CE(l,l')}{\partial O^2_0} \cdot \frac{\partial O^2_0}{\partial x^2_0} \frac{\partial x^2_0}{\partial a^1_{00}}$$

$$= -\frac{l_0}{O^2_0} . softmax(x^2_0)(1 - softmax(x^2_0))$$

$$= -\frac{l_0}{O^2_0} . O^2_0(1 - O^2_0)$$

Therefore:

$$a^1_{00} = a^1_{00} - \alpha \cdot (-\frac{l_0}{O^2_0} . O^2_0(1 - O^2_0))$$

For n=1:

$$a^1_{01} = a^1_{01} - \alpha \frac{\partial CE(l,l')}{\partial a^1_{01}}$$

$$\frac{\partial CE(l,l')}{\partial a^1_{01}} = \frac{\partial CE(l,l')}{\partial O^2_0} \cdot \frac{\partial O^2_0}{\partial x^2_1} \frac{\partial x^2_1}{\partial a^1_{01}}$$

$$= -\frac{l_1}{O^2_1} . softmax(x^2_1)(1 - softmax(x^2_1))$$

$$= -\frac{l_1}{O^2_1} . O^2_1(1 - O^2_1)$$

Therefore:

$$a^1_{01} = a^1_{01} - \alpha \cdot (-\frac{l_1}{O^2_1} . O^2_1(1 - O^2_1))$$

For n=2:

$$a^1_{02} = a^1_{02} - \alpha \frac{\partial CE(l,l')}{\partial a^1_{02}}$$

$$\frac{\partial CE(l,l')}{\partial a^1_{02}} = \frac{\partial CE(l,l')}{\partial O^2_0} \cdot \frac{\partial O^2_0}{\partial x^2_2} \frac{\partial x^2_2}{\partial a^1_{02}}$$

$$= -\frac{l_2}{O_2^2}.softmax(x_2^2)(1 - softmax(x_2^2))$$

$$= -\frac{l_2}{O_2^2}.O_2^2(1 - O_2^2)$$

Therefore:

$$a_{02}^1 = a_{02}^1 - \alpha \cdot (-\frac{l_2}{O_2^2}.O_2^2(1 - O_2^2))$$

For k=1:
For n=0:

$$a_{10}^1 = a_{10}^1 - \alpha \frac{\partial CE(l,l')}{\partial a_{10}^1}$$

$$\frac{\partial CE(l,l')}{\partial a_{10}^1} = \frac{\partial CE(l,l')}{\partial O_0^2} . \frac{\partial O_0^2}{\partial x_0^2} \frac{\partial x_0^2}{\partial a_{10}^1}$$

$$= -\frac{l_0}{O_0^2}.softmax(x_0^2)(1 - softmax(x_0^2)).O_1^1$$

$$= -\frac{l_0}{O_0^2}.O_0^2(1 - O_0^2).O_1^1$$

Therefore:

$$a_{10}^1 = a_{10}^1 - \alpha \cdot (-\frac{l_0}{O_0^2}.O_0^2(1 - O_0^2).O_1^1)$$

For n=1:

$$a_{11}^1 = a_{11}^1 - \alpha \frac{\partial CE(l,l')}{\partial a_{11}^1}$$

$$\frac{\partial CE(l,l')}{\partial a_{11}^1} = \frac{\partial CE(l,l')}{\partial O_0^2} . \frac{\partial O_0^2}{\partial x_1^2} \frac{\partial x_1^2}{\partial a_{11}^1}$$

$$= -\frac{l_1}{O_1^2}.softmax(x_1^2)(1 - softmax(x_1^2)).O_1^1$$

$$= -\frac{l_1}{O_1^2}.O_1^2(1 - O_1^2).O_1^1$$

Therefore:

$$a_{11}^1 = a_{11}^1 - \alpha \cdot (-\frac{l_1}{O_1^2}.O_1^2(1 - O_1^2).O_1^1)$$

For n=1:

$$a_{12}^1 = a_{12}^1 - \alpha \frac{\partial CE(l,l')}{\partial a_{12}^1}$$

$$\frac{\partial CE(l,l')}{\partial a_{12}^1} = \frac{\partial CE(l,l')}{\partial O_0^2} . \frac{\partial O_0^2}{\partial x_2^2} \frac{\partial x_2^2}{\partial a_{12}^1}$$

$$= -\frac{l_2}{O_2^2}.softmax(x_2^2)(1 - softmax(x_2^2)).O_1^1$$

$$= -\frac{l_2}{O_2^2}.O_2^2(1 - O_2^2).O_1^1$$

Therefore:

$$a_{12}^1 = a_{12}^1 - \alpha \cdot (-\frac{l_2}{O_2^2}.O_2^2(1 - O_2^2).O_1^1)$$

For k=2:
For n=0:

$$a_{20}^1 = a_{20}^1 - \alpha \frac{\partial CE(l, l')}{\partial a_{20}^1}$$

$$\frac{\partial CE(l, l')}{\partial a_{20}^1} = \frac{\partial CE(l, l')}{\partial O_0^2} . \frac{\partial O_0^2}{\partial x_0^2} \frac{\partial x_0^2}{\partial a_{20}^1}$$

$$= -\frac{l_0}{O_0^2}.softmax(x_0^2)(1 - softmax(x_0^2)).O_2^1$$

$$= -\frac{l_0}{O_0^2}.O_0^2(1 - O_0^2).O_2^1$$

Therefore:

$$a_{20}^1 = a_{20}^1 - \alpha \cdot (-\frac{l_0}{O_0^2}.O_0^2(1 - O_0^2).O_2^1)$$

For n=1:

$$a_{21}^1 = a_{21}^1 - \alpha \frac{\partial CE(l, l')}{\partial a_{21}^1}$$

$$\frac{\partial CE(l, l')}{\partial a_{21}^1} = \frac{\partial CE(l, l')}{\partial O_0^2} . \frac{\partial O_0^2}{\partial x_1^2} \frac{\partial x_1^2}{\partial a_{21}^1}$$

$$= -\frac{l_1}{O_1^2}.softmax(x_1^2)(1 - softmax(x_1^2)).O_2^1$$

$$= -\frac{l_1}{O_1^2}.O_1^2(1 - O_1^2).O_2^1$$

Therefore:

$$a_{21}^1 = a_{21}^1 - \alpha \cdot (-\frac{l_1}{O_1^2}.O_1^2(1 - O_1^2).O_2^1)$$

For n=2:

$$a_{22}^1 = a_{22}^1 - \alpha \frac{\partial CE(l, l')}{\partial a_{22}^1}$$

$$\frac{\partial CE(l, l')}{\partial a_{22}^1} = \frac{\partial CE(l, l')}{\partial O_0^2} . \frac{\partial O_0^2}{\partial x_2^2} \frac{\partial x_2^2}{\partial a_{22}^1}$$

$$= -\frac{l_2}{O_2^2}.softmax(x_2^2)(1 - softmax(x_2^2)).O_2^1$$

$$= -\frac{l_2}{O_2^2}.O_2^2(1 - O_2^2).O_2^1$$

Therefore:

$$a_{22}^1 = a_{22}^1 - \alpha \cdot (-\frac{l_2}{O_2^2}.O_2^2(1 - O_2^2).O_2^1)$$

For k=3:
For n=0:

$$a_{30}^1 = a_{30}^1 - \alpha\frac{\partial CE(l,l')}{\partial a_{30}^1}$$

$$\frac{\partial CE(l,l')}{\partial a_{30}^1} = \frac{\partial CE(l,l')}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial x_0^2}\frac{\partial x_0^2}{\partial a_{30}^1}$$

$$= -\frac{l_0}{O_0^2}.softmax(x_0^2)(1 - softmax(x_0^2)).O_3^1$$

$$= -\frac{l_0}{O_0^2}.O_0^2(1 - O_0^2).O_3^1$$

Therefore:

$$a_{30}^1 = a_{30}^1 - \alpha \cdot (-\frac{l_0}{O_0^2}.O_0^2(1 - O_0^2).O_3^1)$$

For n=1:

$$a_{31}^1 = a_{31}^1 - \alpha\frac{\partial CE(l,l')}{\partial a_{31}^1}$$

$$\frac{\partial CE(l,l')}{\partial a_{31}^1} = \frac{\partial CE(l,l')}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial x_1^2}\frac{\partial x_1^2}{\partial a_{31}^1}$$

$$= -\frac{l_1}{O_1^2}.softmax(x_1^2)(1 - softmax(x_1^2)).O_3^1$$

$$= -\frac{l_1}{O_1^2}.O_1^2(1 - O_1^2).O_3^1$$

Therefore:

$$a_{31}^1 = a_{31}^1 - \alpha \cdot (-\frac{l_1}{O_1^2}.O_1^2(1 - O_1^2).O_3^1)$$

For n=2:

$$a_{32}^1 = a_{32}^1 - \alpha\frac{\partial CE(l,l')}{\partial a_{32}^1}$$

$$\frac{\partial CE(l,l')}{\partial a_{32}^1} = \frac{\partial CE(l,l')}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial x_2^2}\frac{\partial x_2^2}{\partial a_{32}^1}$$

$$= -\frac{l_2}{O_2^2}.softmax(x_2^2)(1 - softmax(x_2^2)).O_3^1$$

$$= -\frac{l_2}{O_2^2}.O_2^2(1 - O_2^2).O_3^1$$

Therefore:

$$a_{32}^1 = a_{32}^1 - \alpha \cdot (-\frac{l_2}{O_2^2}.O_2^2(1 - O_2^2).O_3^1)$$

$$a_{ik}^0 = a_{ik}^0 - \alpha \frac{\partial CE(l,l')}{\partial a_{ik}^0}$$

For i=0:
For k=1:

$$a_{01}^0 = a_{01}^0 - \alpha \frac{\partial CE(l,l')}{\partial a_{01}^0}$$

$$\frac{\partial CE(l,l')}{\partial a_{01}^0} = \frac{\partial CE(l,l')}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial x_0^2} \frac{\partial x_0^2}{\partial O_1^1} \frac{\partial O_1^1}{\partial x_1^1} \frac{\partial x_1^1}{\partial a_{01}^0}$$

$$+ \frac{\partial CE(l,l')}{\partial O_1^2} \cdot \frac{\partial O_1^2}{\partial x_1^2} \frac{\partial x_1^2}{\partial O_1^1} \frac{\partial O_1^1}{\partial x_1^1} \frac{\partial x_1^1}{\partial a_{01}^0}$$

$$+ \frac{\partial CE(l,l')}{\partial O_2^2} \cdot \frac{\partial O_2^2}{\partial x_2^2} \frac{\partial x_2^2}{\partial O_1^1} \frac{\partial O_1^1}{\partial x_1^1} \frac{\partial x_1^1}{\partial a_{01}^0}$$

$$= -\frac{l_0}{O_0^2} \cdot O_0^2 \cdot (1 - O_0^2) a_{10}^1 \sigma(x_1^1) \cdot (1 - \sigma(x_1^1)) \cdot O_0^0$$

$$-\frac{l_1}{O_1^2} \cdot O_1^2 \cdot (1 - O_1^2) a_{11}^1 \sigma(x_1^1) \cdot (1 - \sigma(x_1^1)) \cdot O_0^0$$

$$-\frac{l_2}{O_2^2} \cdot O_2^2 \cdot (1 - O_2^2) a_{12}^1 \sigma(x_1^1) \cdot (1 - \sigma(x_1^1)) \cdot O_0^0$$

Therefore:

$$a_{01}^0 = a_{01}^0 + \alpha(\frac{l_0}{O_0^2} \cdot O_0^2 \cdot (1 - O_0^2) a_{10}^1 \sigma(x_1^1) \cdot (1 - \sigma(x_1^1))$$

$$+\frac{l_1}{O_1^2} \cdot O_1^2 \cdot (1 - O_1^2) a_{11}^1 \sigma(x_1^1) \cdot (1 - \sigma(x_1^1))$$

$$+\frac{l_2}{O_2^2} \cdot O_2^2 \cdot (1 - O_2^2) a_{12}^1 \sigma(x_1^1) \cdot (1 - \sigma(x_1^1))$$

For k=2:

$$a_{02}^0 = a_{02}^0 - \alpha \frac{\partial CE(l,l')}{\partial a_{02}^0}$$

$$\frac{\partial CE(l,l')}{\partial a_{02}^0} = \frac{\partial CE(l,l')}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial x_0^2} \frac{\partial x_0^2}{\partial O_2^1} \frac{\partial O_2^1}{\partial x_2^1} \frac{\partial x_2^1}{\partial a_{02}^0}$$

$$+ \frac{\partial CE(l,l')}{\partial O_1^2} \cdot \frac{\partial O_1^2}{\partial x_1^2} \frac{\partial x_1^2}{\partial O_2^1} \frac{\partial O_2^1}{\partial x_2^1} \frac{\partial x_2^1}{\partial a_{02}^0}$$

$$+ \frac{\partial CE(l,l')}{\partial O_2^2} \cdot \frac{\partial O_2^2}{\partial x_2^2} \frac{\partial x_2^2}{\partial O_2^1} \frac{\partial O_2^1}{\partial x_2^1} \frac{\partial x_2^1}{\partial a_{02}^0}$$

$$= -\frac{l_0}{O_0^2} \cdot O_0^2 \cdot (1 - O_0^2) a_{20}^1 \sigma(x_2^1) \cdot (1 - \sigma(x_2^1)) \cdot O_0^0$$

$$-\frac{l_1}{O_1^2}\cdot O_1^2\cdot(1-O_1^2)a_{21}^1\sigma(x_2^1)\cdot(1-\sigma(x_2^1))\cdot O_0^0$$

$$-\frac{l_2}{O_2^2}\cdot O_2^2\cdot(1-O_2^2)a_{22}^1\sigma(x_2^1)\cdot(1-\sigma(x_2^1))\cdot O_0^0$$

Therefore:

$$a_{02}^0=a_{02}^0+\alpha(\frac{l_0}{O_0^2}\cdot O_0^2\cdot(1-O_0^2)a_{20}^1\sigma(x_2^1)\cdot(1-\sigma(x_2^1))$$

$$+\frac{l_1}{O_1^2}\cdot O_1^2\cdot(1-O_1^2)a_{21}^1\sigma(x_2^1)\cdot(1-\sigma(x_2^1))$$

$$+\frac{l_2}{O_2^2}\cdot O_2^2\cdot(1-O_2^2)a_{22}^1\sigma(x_2^1)\cdot(1-\sigma(x_2^1))$$

For k=3:

$$a_{03}^0=a_{03}^0-\alpha\frac{\partial CE(l,l')}{\partial a_{03}^0}$$

$$\frac{\partial CE(l,l')}{\partial a_{03}^0}=\frac{\partial CE(l,l')}{\partial O_0^2}\cdot\frac{\partial O_0^2}{\partial x_0^2}\frac{\partial x_0^2}{\partial O_3^1}\frac{\partial O_3^1}{\partial x_3^1}\frac{\partial x_3^1}{\partial a_{03}^0}$$

$$+\frac{\partial CE(l,l')}{\partial O_1^2}\cdot\frac{\partial O_1^2}{\partial x_1^2}\frac{\partial x_1^2}{\partial O_3^1}\frac{\partial O_3^1}{\partial x_3^1}\frac{\partial x_3^1}{\partial a_{03}^0}$$

$$+\frac{\partial CE(l,l')}{\partial O_2^2}\cdot\frac{\partial O_2^2}{\partial x_2^2}\frac{\partial x_2^2}{\partial O_3^1}\frac{\partial O_3^1}{\partial x_3^1}\frac{\partial x_3^1}{\partial a_{03}^0}$$

$$=-\frac{l_0}{O_0^2}\cdot O_0^2\cdot(1-O_0^2)a_{30}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))\cdot O_0^0$$

$$-\frac{l_1}{O_1^2}\cdot O_1^2\cdot(1-O_1^2)a_{31}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))\cdot O_0^0$$

$$-\frac{l_2}{O_2^2}\cdot O_2^2\cdot(1-O_2^2)a_{32}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))\cdot O_0^0$$

Therefore:

$$a_{03}^0=a_{03}^0+\alpha(\frac{l_0}{O_0^2}\cdot O_0^2\cdot(1-O_0^2)a_{30}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))$$

$$+\frac{l_1}{O_1^2}\cdot O_1^2\cdot(1-O_1^2)a_{31}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))$$

$$+\frac{l_2}{O_2^2}\cdot O_2^2\cdot(1-O_2^2)a_{32}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))$$

For i=1:
For k=1:

$$a_{11}^0=a_{11}^0-\alpha\frac{\partial CE(l,l')}{\partial a_{11}^0}$$

$$\frac{\partial CE(l,l')}{\partial a^0_{11}} = \frac{\partial CE(l,l')}{\partial O^2_0} \cdot \frac{\partial O^2_0}{\partial x^2_0} \frac{\partial x^2_0}{\partial O^1_1} \frac{\partial O^1_1}{\partial x^1_1} \frac{\partial x^1_1}{\partial a^0_{11}}$$

$$+ \frac{\partial CE(l,l')}{\partial O^2_1} \cdot \frac{\partial O^2_1}{\partial x^2_1} \frac{\partial x^2_1}{\partial O^1_1} \frac{\partial O^1_1}{\partial x^1_1} \frac{\partial x^1_1}{\partial a^0_{11}}$$

$$+ \frac{\partial CE(l,l')}{\partial O^2_2} \cdot \frac{\partial O^2_2}{\partial x^2_2} \frac{\partial x^2_2}{\partial O^1_1} \frac{\partial O^1_1}{\partial x^1_1} \frac{\partial x^1_1}{\partial a^0_{11}}$$

$$= -\frac{l_0}{O^2_0} \cdot O^2_0 \cdot (1 - O^2_0) a^1_{10} \sigma(x^1_1) \cdot (1 - \sigma(x^1_1)) \cdot O^0_1$$

$$-\frac{l_1}{O^2_1} \cdot O^2_1 \cdot (1 - O^2_1) a^1_{11} \sigma(x^1_1) \cdot (1 - \sigma(x^1_1)) \cdot O^0_1$$

$$-\frac{l_2}{O^2_2} \cdot O^2_2 \cdot (1 - O^2_2) a^1_{12} \sigma(x^1_1) \cdot (1 - \sigma(x^1_1)) \cdot O^0_1$$

Therefore:

$$a^0_{11} = a^0_{11} + \alpha(\frac{l_0}{O^2_0} \cdot O^2_0 \cdot (1 - O^2_0) a^1_{10} \sigma(x^1_1) \cdot (1 - \sigma(x^1_1)) \cdot O^0_1$$

$$+\frac{l_1}{O^2_1} \cdot O^2_1 \cdot (1 - O^2_1) a^1_{11} \sigma(x^1_1) \cdot (1 - \sigma(x^1_1)) \cdot O^0_1$$

$$+\frac{l_2}{O^2_2} \cdot O^2_2 \cdot (1 - O^2_2) a^1_{12} \sigma(x^1_1) \cdot (1 - \sigma(x^1_1)) \cdot O^0_1$$

For k=2:

$$a^0_{12} = a^0_{12} - \alpha \frac{\partial CE(l,l')}{\partial a^0_{12}}$$

$$\frac{\partial CE(l,l')}{\partial a^0_{12}} = \frac{\partial CE(l,l')}{\partial O^2_0} \cdot \frac{\partial O^2_0}{\partial x^2_0} \frac{\partial x^2_0}{\partial O^1_2} \frac{\partial O^1_2}{\partial x^1_2} \frac{\partial x^1_2}{\partial a^0_{12}}$$

$$+ \frac{\partial CE(l,l')}{\partial O^2_1} \cdot \frac{\partial O^2_1}{\partial x^2_1} \frac{\partial x^2_1}{\partial O^1_2} \frac{\partial O^1_2}{\partial x^1_2} \frac{\partial x^1_2}{\partial a^0_{12}}$$

$$+ \frac{\partial CE(l,l')}{\partial O^2_2} \cdot \frac{\partial O^2_2}{\partial x^2_2} \frac{\partial x^2_2}{\partial O^1_2} \frac{\partial O^1_2}{\partial x^1_2} \frac{\partial x^1_2}{\partial a^0_{12}}$$

$$= -\frac{l_0}{O^2_0} \cdot O^2_0 \cdot (1 - O^2_0) a^1_{20} \sigma(x^1_2) \cdot (1 - \sigma(x^1_2)) \cdot O^0_1$$

$$-\frac{l_1}{O^2_1} \cdot O^2_1 \cdot (1 - O^2_1) a^1_{21} \sigma(x^1_2) \cdot (1 - \sigma(x^1_2)) \cdot O^0_1$$

$$-\frac{l_2}{O^2_2} \cdot O^2_2 \cdot (1 - O^2_2) a^1_{22} \sigma(x^1_2) \cdot (1 - \sigma(x^1_2)) \cdot O^0_1$$

Therefore:

$$a^0_{12} = a^0_{12} + \alpha(\frac{l_0}{O^2_0} \cdot O^2_0 \cdot (1 - O^2_0) a^1_{20} \sigma(x^1_2) \cdot (1 - \sigma(x^1_2)) \cdot O^0_1$$

$$+\frac{l_1}{O_1^2}\cdot O_1^2\cdot(1-O_1^2)a_{21}^1\sigma(x_2^1)\cdot(1-\sigma(x_2^1))\cdot O_1^0$$

$$+\frac{l_2}{O_2^2}\cdot O_2^2\cdot(1-O_2^2)a_{22}^1\sigma(x_2^1)\cdot(1-\sigma(x_2^1))\cdot O_1^0$$

For k=3:

$$a_{13}^0=a_{13}^0-\alpha\frac{\partial CE(l,l')}{\partial a_{13}^0}$$

$$\frac{\partial CE(l,l')}{\partial a_{13}^0}=\frac{\partial CE(l,l')}{\partial O_0^2}\cdot\frac{\partial O_0^2}{\partial x_0^2}\frac{\partial x_0^2}{\partial O_3^1}\frac{\partial O_3^1}{\partial x_3^1}\frac{\partial x_3^1}{\partial a_{13}^0}$$

$$+\frac{\partial CE(l,l')}{\partial O_1^2}\cdot\frac{\partial O_1^2}{\partial x_1^2}\frac{\partial x_1^2}{\partial O_3^1}\frac{\partial O_3^1}{\partial x_3^1}\frac{\partial x_3^1}{\partial a_{13}^0}$$

$$+\frac{\partial CE(l,l')}{\partial O_2^2}\cdot\frac{\partial O_2^2}{\partial x_2^2}\frac{\partial x_2^2}{\partial O_3^1}\frac{\partial O_3^1}{\partial x_3^1}\frac{\partial x_3^1}{\partial a_{13}^0}$$

$$=-\frac{l_0}{O_0^2}\cdot O_0^2\cdot(1-O_0^2)a_{30}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))\cdot O_1^0$$

$$-\frac{l_1}{O_1^2}\cdot O_1^2\cdot(1-O_1^2)a_{31}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))\cdot O_1^0$$

$$-\frac{l_2}{O_2^2}\cdot O_2^2\cdot(1-O_2^2)a_{32}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))\cdot O_1^0$$

Therefore:

$$a_{13}^0=a_{13}^0+\alpha(\frac{l_0}{O_0^2}\cdot O_0^2\cdot(1-O_0^2)a_{30}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))\cdot O_1^0$$

$$+\frac{l_1}{O_1^2}\cdot O_1^2\cdot(1-O_1^2)a_{31}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))\cdot O_1^0$$

$$+\frac{l_2}{O_2^2}\cdot O_2^2\cdot(1-O_2^2)a_{32}^1\sigma(x_3^1)\cdot(1-\sigma(x_3^1))\cdot O_1^0$$

For i=2:

For k=1:

$$a_{21}^0=a_{21}^0-\alpha\frac{\partial CE(l,l')}{\partial a_{21}^0}$$

$$\frac{\partial CE(l,l')}{\partial a_{21}^0}=\frac{\partial CE(l,l')}{\partial O_0^2}\cdot\frac{\partial O_0^2}{\partial x_0^2}\frac{\partial x_0^2}{\partial O_1^1}\frac{\partial O_1^1}{\partial x_1^1}\frac{\partial x_1^1}{\partial a_{21}^0}$$

$$+\frac{\partial CE(l,l')}{\partial O_1^2}\cdot\frac{\partial O_1^2}{\partial x_1^2}\frac{\partial x_1^2}{\partial O_1^1}\frac{\partial O_1^1}{\partial x_1^1}\frac{\partial x_1^1}{\partial a_{21}^0}$$

$$+\frac{\partial CE(l,l')}{\partial O_2^2}\cdot\frac{\partial O_2^2}{\partial x_2^2}\frac{\partial x_2^2}{\partial O_1^1}\frac{\partial O_1^1}{\partial x_1^1}\frac{\partial x_1^1}{\partial a_{21}^0}$$

$$=-\frac{l_0}{O_0^2}\cdot O_0^2\cdot(1-O_0^2)a_{10}^1\sigma(x_1^1)\cdot(1-\sigma(x_1^1))\cdot O_2^0$$

$$-\frac{l_1}{O_1^2} \cdot O_1^2 \cdot (1 - O_1^2) a_{11}^1 \sigma(x_1^1) \cdot (1 - \sigma(x_1^1)) \cdot O_2^0$$

$$-\frac{l_2}{O_2^2} \cdot O_2^2 \cdot (1 - O_2^2) a_{12}^1 \sigma(x_1^1) \cdot (1 - \sigma(x_1^1)) \cdot O_2^0$$

Therefore:

$$a_{21}^0 = a_{21}^0 + \alpha(\frac{l_0}{O_0^2} \cdot O_0^2 \cdot (1 - O_0^2) a_{10}^1 \sigma(x_1^1) \cdot (1 - \sigma(x_1^1)) \cdot O_2^0$$

$$+\frac{l_1}{O_1^2} \cdot O_1^2 \cdot (1 - O_1^2) a_{11}^1 \sigma(x_1^1) \cdot (1 - \sigma(x_1^1)) \cdot O_2^0$$

$$+\frac{l_2}{O_2^2} \cdot O_2^2 \cdot (1 - O_2^2) a_{12}^1 \sigma(x_1^1) \cdot (1 - \sigma(x_1^1)) \cdot O_2^0$$

For k=2:

$$a_{22}^0 = a_{22}^0 - \alpha \frac{\partial CE(l, l')}{\partial a_{22}^0}$$

$$\frac{\partial CE(l, l')}{\partial a_{22}^0} = \frac{\partial CE(l, l')}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial x_0^2} \frac{\partial x_0^2}{\partial O_2^1} \frac{\partial O_2^1}{\partial x_2^1} \frac{\partial x_2^1}{\partial a_{22}^0}$$

$$+\frac{\partial CE(l, l')}{\partial O_1^2} \cdot \frac{\partial O_1^2}{\partial x_1^2} \frac{\partial x_1^2}{\partial O_2^1} \frac{\partial O_2^1}{\partial x_2^1} \frac{\partial x_2^1}{\partial a_{22}^0}$$

$$+\frac{\partial CE(l, l')}{\partial O_2^2} \cdot \frac{\partial O_2^2}{\partial x_2^2} \frac{\partial x_2^2}{\partial O_2^1} \frac{\partial O_2^1}{\partial x_2^1} \frac{\partial x_2^1}{\partial a_{22}^0}$$

$$= -\frac{l_0}{O_0^2} \cdot O_0^2 \cdot (1 - O_0^2) a_{20}^1 \sigma(x_2^1) \cdot (1 - \sigma(x_2^1)) \cdot O_2^0$$

$$-\frac{l_1}{O_1^2} \cdot O_1^2 \cdot (1 - O_1^2) a_{21}^1 \sigma(x_2^1) \cdot (1 - \sigma(x_2^1)) \cdot O_2^0$$

$$-\frac{l_2}{O_2^2} \cdot O_2^2 \cdot (1 - O_2^2) a_{22}^1 \sigma(x_2^1) \cdot (1 - \sigma(x_2^1)) \cdot O_2^0$$

Therefore:

$$a_{22}^0 = a_{22}^0 + \alpha(\frac{l_0}{O_0^2} \cdot O_0^2 \cdot (1 - O_0^2) a_{20}^1 \sigma(x_2^1) \cdot (1 - \sigma(x_2^1)) \cdot O_2^0$$

$$+\frac{l_1}{O_1^2} \cdot O_1^2 \cdot (1 - O_1^2) a_{21}^1 \sigma(x_2^1) \cdot (1 - \sigma(x_2^1)) \cdot O_2^0$$

$$+\frac{l_2}{O_2^2} \cdot O_2^2 \cdot (1 - O_2^2) a_{22}^1 \sigma(x_2^1) \cdot (1 - \sigma(x_2^1)) \cdot O_2^0$$

For k=3:

$$a_{23}^0 = a_{23}^0 - \alpha \frac{\partial CE(l, l')}{\partial a_{23}^0}$$

$$
\begin{aligned}
\frac{\partial CE(l,l')}{\partial a_{23}^0} &= \frac{\partial CE(l,l')}{\partial O_0^2} \cdot \frac{\partial O_0^2}{\partial x_0^2} \frac{\partial x_0^2}{\partial O_3^1} \frac{\partial O_3^1}{\partial x_3^1} \frac{\partial x_3^1}{\partial a_{23}^0} \\
&+ \frac{\partial CE(l,l')}{\partial O_1^2} \cdot \frac{\partial O_1^2}{\partial x_1^2} \frac{\partial x_1^2}{\partial O_3^1} \frac{\partial O_3^1}{\partial x_3^1} \frac{\partial x_3^1}{\partial a_{23}^0} \\
&+ \frac{\partial CE(l,l')}{\partial O_2^2} \cdot \frac{\partial O_2^2}{\partial x_2^2} \frac{\partial x_2^2}{\partial O_3^1} \frac{\partial O_3^1}{\partial x_3^1} \frac{\partial x_3^1}{\partial a_{23}^0} \\
&= -\frac{l_0}{O_0^2} \cdot O_0^2 \cdot (1 - O_0^2) a_{30}^1 \sigma(x_3^1) \cdot (1 - \sigma(x_3^1)) \cdot O_2^0 \\
&- \frac{l_1}{O_1^2} \cdot O_1^2 \cdot (1 - O_1^2) a_{31}^1 \sigma(x_3^1) \cdot (1 - \sigma(x_3^1)) \cdot O_2^0 \\
&- \frac{l_2}{O_2^2} \cdot O_2^2 \cdot (1 - O_2^2) a_{32}^1 \sigma(x_3^1) \cdot (1 - \sigma(x_3^1)) \cdot O_2^0
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
a_{23}^0 = a_{23}^0 + \alpha(&\frac{l_0}{O_0^2} \cdot O_0^2 \cdot (1 - O_0^2) a_{30}^1 \sigma(x_3^1) \cdot (1 - \sigma(x_3^1)) \cdot O_2^0 \\
&+ \frac{l_1}{O_1^2} \cdot O_1^2 \cdot (1 - O_1^2) a_{31}^1 \sigma(x_3^1) \cdot (1 - \sigma(x_3^1)) \cdot O_2^0 \\
&+ \frac{l_2}{O_2^2} \cdot O_2^2 \cdot (1 - O_2^2) a_{32}^1 \sigma(x_3^1) \cdot (1 - \sigma(x_3^1)) \cdot O_2^0
\end{aligned}
$$