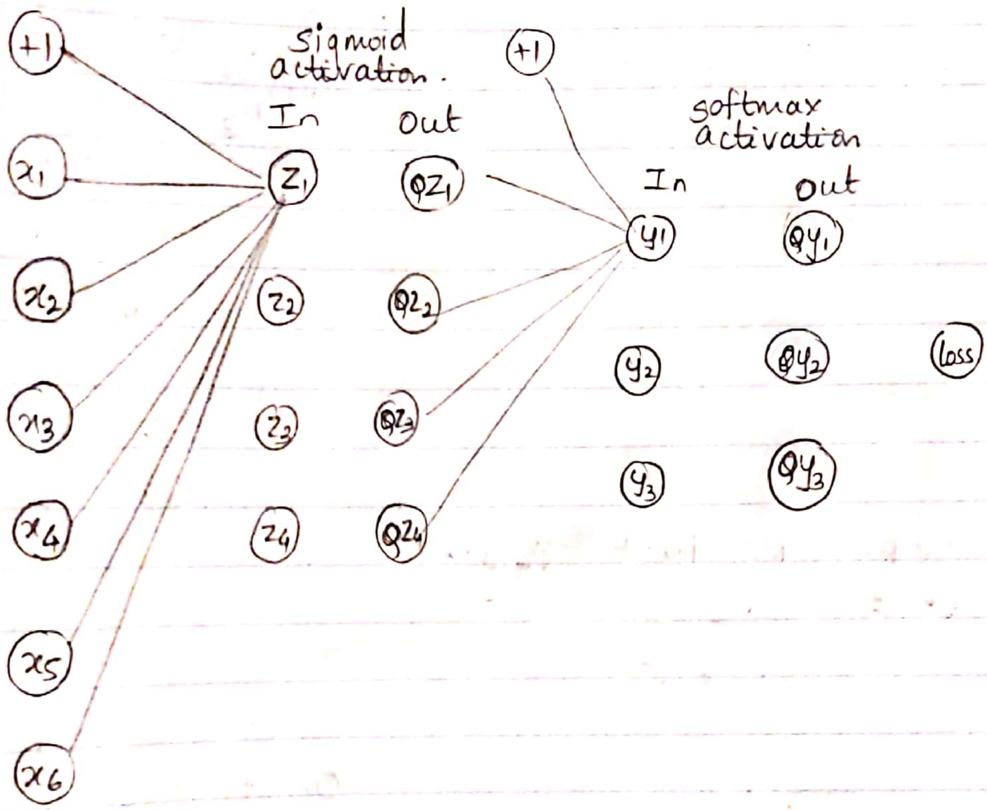
Problem 4 : Feed forward and Back Propagation.



a) Forward feed:

Linear combination at first layer

$$a_j = a_0 + \sum_{i=1} a_{ji} x_i$$

$$z_1 = b_0 + x_1 w_{11} + x_2 w_{12} + x_3 w_{13} + x_4 w_{14} + x_5 w_{15} + x_6 w_{16}$$
$$= 1 + (1\times1) + (1\times2) + (0\times-3) + (0\times0) + (1\times1) + (1\times-3)$$
$$= 2$$

Similarly z_2 to z_4 can be calulated using the weight matrix α given.

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & 2 & -3 & 0 & 1 & -3 \\ 3 & 1 & 2 & 1 & 0 & 2 \\ 2 & 2 & 2 & 2 & 2 & 1 \\ 1 & 0 & 2 & 1 & -2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 7 \\ 8 \\ 2 \end{bmatrix}$$

Activation of first layer using sigmoid

$$z_i = \frac{1}{1+\exp(-z_i)}$$

$$\phi z_1 = \frac{1}{1+\exp(2)} = 0.880797, \quad \phi z_2 = \frac{1}{1+\exp(7)} = 0.999088$$

$$\phi z_3 = \frac{1}{1+\exp(8)} = 0.999664, \quad \phi z_4 = \frac{1}{1+\exp(2)} = 0.880797$$

Linear combination at second layer

$$y_1 = b_0 + z_1 w_{21} + z_2 w_{22} + z_3 w_{23}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & -2 & 1 \\ 1 & -1 & 1 & 2 \\ 3 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0.880797 \\ 0.999088 \\ 0.999664 \\ 0.880797 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2.760442 \\ 3.642966 \\ 4.5226126 \end{bmatrix}$$

Activation of second layer using softmax

$$Qy_1 = \frac{e^y}{\sum e^y}$$

$$\begin{bmatrix} Qy_1 \\ Qy_2 \\ Qy_3 \end{bmatrix} = \begin{bmatrix} 0.108201 \\ 0.261521 \\ 0.630277 \end{bmatrix}$$

Loss $\therefore\ l = -\sum_{i=1}^{3} \hat{y}_i \times \log(Qy_i)$ log is to base e

$$= -0\times\log(0.108201) - 1\times\log(0.261521) - 0\times\log(0.630277)$$

$$= 1.3412$$

(b) Back propagation

(A) Back propagation for second layer

Updating weights,

$$\frac{\partial L}{\omega_{21}} = \frac{\partial L}{\partial Qy_1} \times \frac{\partial Qy_1}{\partial y_1} \times \frac{\partial y_1}{\partial \omega_{21}} + \frac{\partial L}{\partial Qy_2} \times \frac{\partial Qy_2}{\partial y_1} \frac{\partial y_1}{\partial \omega_{21}}$$

$$+ \frac{\partial L}{\partial Qy_3} \times \frac{\partial Qy_3}{\partial y_1} \times \frac{\partial y_1}{\partial \omega_{21}}$$

$$= \frac{-\hat{y}_1}{Qy_1} \times Qy_1 \times (1-Qy_1) \times QZ_1 + \frac{(-\hat{y}_2)}{Qy_2} \times (-Qy_2 \times Qy_1)\, QZ_1$$

$$\frac{(-\hat{y}_3)}{Qy_3}(-Qy_3 \times Qy_1)\, QZ_1$$

$$= -\hat{y}_1 (1-\phi y_1) \times \phi z_1 + \hat{y}_2 \times \phi y_1 \times \phi z_1 + \hat{y}_3 \times \phi y_1 \times \phi z_1$$

$$= -\hat{y}_1 \times \phi z_1 + \hat{y}_1 \phi z_1 \phi y_1 + \hat{y}_2 \times \phi z_1 \times \phi y_1 + \hat{y}_3 \phi y_1 \times \phi z_1$$

$$= -\hat{y}_1 \times \phi z_1 + \phi z_1 \phi y_1$$

$$w_{21}\text{new} = w_{21} - \frac{\alpha \partial L}{\partial w_{21}} = 1 - 1 \times (\phi y_1 - \hat{y}_1) \times \phi z_1$$

$$w_{21}\text{new} = 1 - 1(0.1082 - 0) \times 0.8808 = 0.9047$$

We can generalize $\frac{\partial L}{\partial w_{2i}} = (-\hat{y}_i \times \phi z_i + \phi z_i \times \phi y_i)$

There updated weights are as follows.

$$w_2\text{new} = B\text{new} = \begin{bmatrix} 0.9047 & 1.8919 & -2.1082 & 0.9047 \\ 1.6505 & -0.2622 & 1.7382 & 2.6505 \\ 2.4449 & 0.3703 & -1.6301 & 0.4449 \end{bmatrix}$$

(B) Back propagation for 1st layer

$$\frac{\partial L}{\partial w_{11}} = \left[\frac{\partial L}{\partial w_{21}} \frac{\partial w_{21}}{\partial \phi z_1} \frac{\partial \phi z_1}{\partial z_1} \frac{\partial z_1}{\partial w_{11}} + \frac{\partial L}{\partial w_{22}} \frac{\partial w_{22}}{\partial \phi z_1} \frac{\partial \phi z_1}{\partial z_1} \frac{\partial z_1}{\partial w_{11}} \right.$$

$$\left. + \frac{\partial L}{\partial w_{31}} \times \frac{\partial w_{31}}{\partial \phi z_1} \frac{\partial \phi z_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_{11}} \right]$$

$$= \left[(\phi y_1 - \hat{y}_1) w_{21} + (\phi y_2 - \hat{y}_2) w_{22} + (\phi y_3 - \hat{y}_3) w_{23} \right]$$

$$\times \phi z_1 (1 - \phi z_1) x_1$$

$$\therefore \frac{\partial L}{\partial \omega_{11}} = \left[(0.1082-0)\times 1 + (0.2615-1)\times 1 + (0.6303-0)\times 3 \right] \times (0.8808) \times (1-0.8808) \times 1$$

$$\frac{\partial L}{\partial \omega_{11}} = 0.1324$$

$$\therefore \omega_{11}\text{update} = \omega_{11} - \alpha \frac{\partial L}{\partial \omega_{11}}$$

$$= 1 - 1 \times 0.1324$$

$$= 0.8676$$

The matrix of updated weights is as follows.

$$W_{1new} = \alpha\text{-update} = \begin{bmatrix} 0.8676 & 1.8676 & -3 & 0 & 0.8676 & -3.1324 \\ 2.9986 & 0.9986 & 2 & 1 & -0.0014 & 1.9986 \\ 2.0005 & 2.0005 & 2 & 2 & 2.0005 & 1.0005 \\ 1.0775 & 0.0775 & 2 & 1 & -1.9224 & 2.0775 \end{bmatrix}$$

Updating 2nd Layer bias.

$$\frac{\partial L}{\partial B_{21}} = \frac{\partial L}{\partial Qy_1} \times \frac{\partial Qy_1}{\partial y_1} \times \frac{\partial y_1}{\partial B_{21}} + \frac{\partial L}{\partial Qy_2} \times \frac{\partial Qy_2}{\partial y_1} \times \frac{\partial y_1}{\partial B_{21}}$$

$$+ \frac{\partial L}{\partial Qy_3} \times \frac{\partial Qy_3}{\partial y_1} \times \frac{\partial y_1}{\partial B_{21}}$$

$$= (Qy_1 - \hat{y}_1)\, B_{21}$$

Scanned by CamScanner

$\therefore B_{21new} = B_{21} - \alpha \left[Qy_1 - \hat{y}_1 \right] B_{21}$

$= 1 - (0.1082 - 0) \times 1$

$= 0.8918$

B_{22}, B_{23} can be calculated as follows

$B_{22} = 1.7385$

$B_{23} = 0.3697$

Updating 1^{st} layer Bias.

$$\frac{\partial L}{\partial B_{11}} = \left(\sum_{i=1}^{3} \frac{\partial L}{\partial w_{1i}} \right) \times \frac{\partial Qz_1}{\partial z_1} \times \frac{\partial z_1}{\partial B_{11}}$$

$= 0.1324$

Therefore, B_{11}-new $= B_{11} - \alpha(0.1324) = 0.8676$

Similarly we can calculate updated values as below,

$B_{12} = 0.9986$

$B_{13} = 1.0005$

$B_{14} = 1.0775$

Scanned by CamScanner

Answers are as follows:

1(a) $a_1^{(1)} = 2 \Rightarrow z_1^{(1)} = 0.88079$

1(b) $a_3^{(1)} = 8 \Rightarrow z_3^{(1)} = 0.99966$

1(c) $b_2^{(1)} = 3.6430$

1(d) $\hat{y}_2(1) = 0.2615$

1(e) class 3

1(f) loss $= 1.3412$

2 (a) $B_{2,1} = 1.6505$

(b) Bias $= 0.8918$

(c) $\alpha_{3,4} = 2$

(d) Bias $= 0.9986$

(e) Class 2