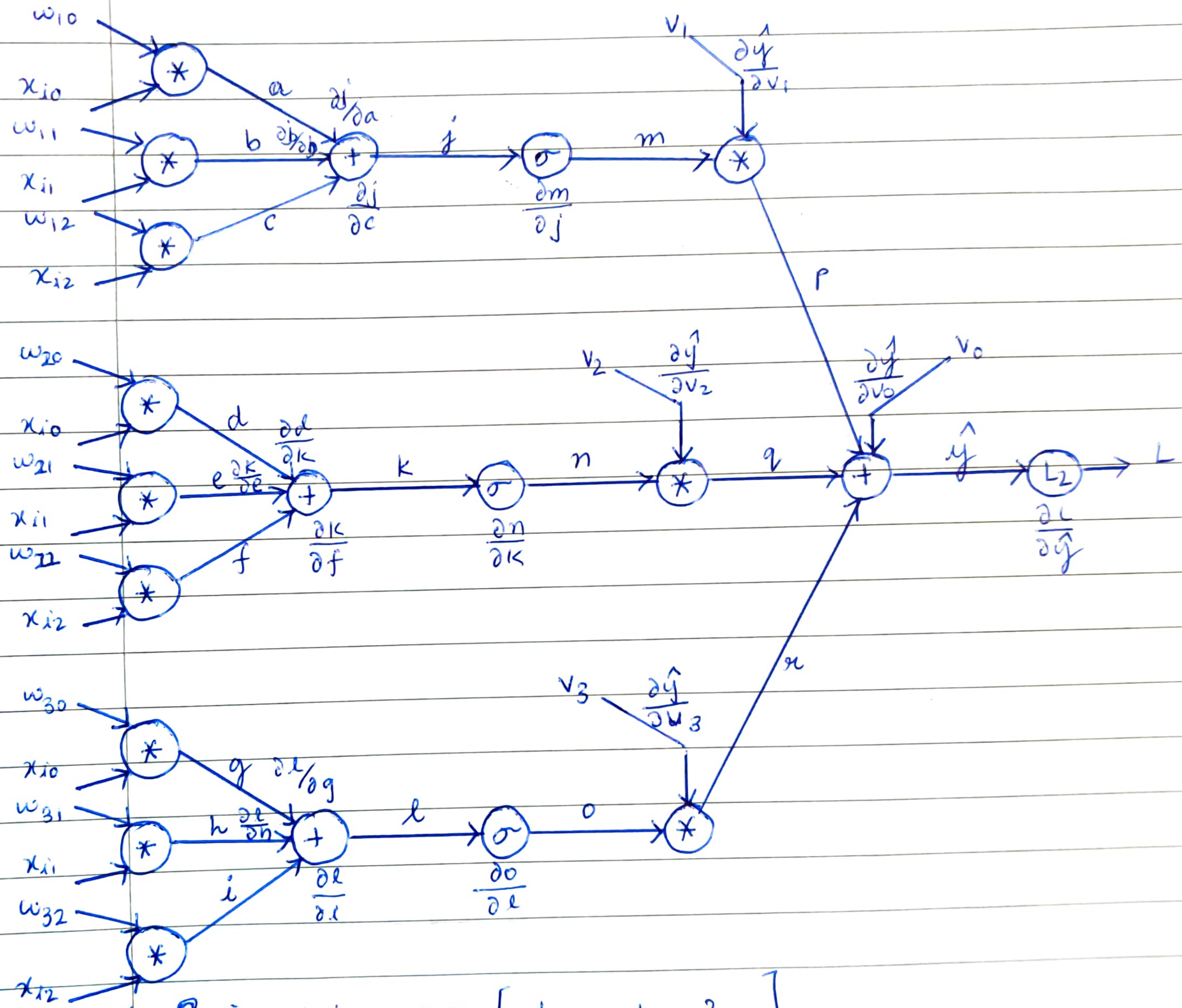
Deep Learning Assignment 2 - A20442409.

1.

(a).



(b) Point 1 : $x = [1 \; . \; 1 \; . \; 2]$

$a = 0.01 \quad b = 0.02 \quad c = 0.06$

$j = 0.01 + 0.02 + 0.06 = 0.09$

$m = \text{sigmoid}(j) = 0.52248$

$d = 0.03 \quad e = 0.01 \quad f = 0.04$

$k = 0.03 + 0.01 + 0.04 = 0.08$

$n = \text{sigmoid}(k) = 0.51999$

$g = 0.03 \quad h = 0.02 \quad i = 0.02$

$l = 0.03 + 0.02 + 0.02 = 0.07.$

$o = \text{sigmoid}(l) = 0.51749.$

$p = m v_1 = 0.01044$

$q = n v_2 = 0.015599$

$r = o v_3 = 0.02069$

$\hat{y}_1 = v_0 + p + q + r = \underline{0.05675}$

Point 2 $\quad x = [\,1 \quad 1 \quad 3\,]$

$a = 0.01 \qquad b = 0.02 \qquad c = 0.09$

$j = 0.01 + 0.02 + 0.09 = 0.12$

$m = \text{sigmoid}(j) = 0.52996$

$d = 0.03 \qquad e = 0.01 \qquad f = 0.06$

$k = 0.03 + 0.01 + 0.06 = 0.1$

$n = \text{sigmoid}(k) = 0.52498$

$g = 0.02 \qquad h = 0.03 \qquad i = 0.03$

$l = 0.02 + 0.03 + 0.03 = 0.08$

$o = \text{sigmoid}(l) = 0.51999$

$p = m v_1 = 0.01059$

$q = n v_2 = 0.01574$

$r = o v_3 = 0.02079$

$\hat{y} = v_0 + p + q + r = \underline{0.057148}$

Point 3 $\quad x = [\,1 \quad 2 \quad 2\,]$

$a = 0.01 \qquad b = 0.04 \qquad c = 0.06$

$j = 0.01 + 0.04 + 0.06 = 0.11$

$m = \text{sigmoid}(j) = 0.52747$

$d = 0.03 \qquad e = 0.02 \qquad f = 0.04$

$k = 0.03 + 0.02 + 0.04 = 0.09$

$n = \text{sigmoid}(k) = 0.52248$

$g = 0.02 \qquad h = 0.06 \qquad i = 0.02$

$l = 0.02 + 0.06 + 0.02 = 0.10$

$o = \text{sigmoid}(l) = 0.52498$

$p = m v_1 = 0.01054$

$q = n v_2 = 0.01567$

$r = o v_3 = 0.02079$

$\hat{y} = v_0 + p + q + r = \underline{0.05722}$

$L_2 \text{ loss} = (\hat{y}_i - y_i)^2$

$\frac{\partial \hat{y}}{\partial v_0} = \frac{\partial \hat{y}}{\partial p} = \frac{\partial \hat{y}}{\partial q} = \frac{\partial \hat{y}}{\partial r} = 1.$

$\frac{\partial p}{\partial v_1} = m \qquad \frac{\partial q}{\partial v_2} = n \qquad \frac{\partial r}{\partial v_3} = o$

$\frac{\partial m}{\partial j} = m(1-m) \qquad \frac{\partial n}{\partial k} = n(1-n) \qquad \frac{\partial o}{\partial l} = o(1-o)$

$\frac{\partial j}{\partial c} = \frac{\partial k}{\partial f} = \frac{\partial l}{\partial g} = \frac{\partial j}{\partial b} = \frac{\partial k}{\partial d} = \frac{\partial l}{\partial h} = \frac{\partial j}{\partial a} = \frac{\partial k}{\partial e} = \frac{\partial l}{\partial i}$

$\frac{\partial a}{\partial w_{10}} = \frac{\partial d}{\partial w_{20}} = \frac{\partial g}{\partial w_{30}} = 1.$

$\frac{\partial b}{\partial w_{11}} = x_{i1} \qquad \frac{\partial c}{\partial w_{12}} = x_{i2} \qquad \frac{\partial e}{\partial w_{21}} = x_{i1} \qquad \frac{\partial f}{\partial w_{22}} = x_{i2}$

$\frac{\partial h}{\partial w_{31}} = x_{i1} \qquad \frac{\partial i}{\partial w_{32}} = x_{i2}. \qquad \frac{\partial p}{\partial m} = v_1 \qquad \frac{\partial q}{\partial n} = v_2$

$\partial k / \partial o = v_3.$

Gradian decent of loss :

$\frac{\partial \hat{y}}{\partial w_{10}} = \frac{\partial \hat{y}}{\partial p} \cdot \frac{\partial p}{\partial m} \cdot \frac{\partial m}{\partial j} \cdot \frac{\partial j}{\partial a} \cdot \frac{\partial a}{\partial w_{10}}$

$= v_1 \cdot m(1-m) \cdot 1 \cdot 1$

$\frac{\partial \hat{y}}{\partial w_{10}} = v_1 \cdot n(1-n).$

Similarly

$$\frac{\partial \hat{y}}{\partial w_{11}} = v_1 \cdot m(1-m) \cdot x_{i1}$$

$$\frac{\partial \hat{y}}{\partial w_{12}} = v_1 \cdot m(1-m) \cdot x_{i2}$$

$$\frac{\partial \hat{y}}{\partial w_{20}} = v_2 \cdot n(1-n) \cdot 1$$

$$\frac{\partial \hat{y}}{\partial w_{21}} = v_2 \cdot n(1-n) \cdot x_{i1}$$

$$\frac{\partial \hat{y}}{\partial w_{22}} = v_2 \cdot n(1-n) \cdot x_{i2}$$

$$\frac{\partial \hat{y}}{\partial w_{30}} = v_3 \cdot o(1-o) \cdot 1$$

$$\frac{\partial \hat{y}}{\partial w_{31}} = v_3 \cdot o(1-o) \cdot x_{i1}$$

$$\frac{\partial \hat{y}}{\partial w_{32}} = v_3 \cdot o(1-o) \cdot x_{i2}$$

c). Gradients of the functions ~~wrt~~ with respect to all weights are: for $x = [1, 2]$

$$\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}(\hat{y}-y)^2 = 2(\hat{y}-y) \cdot 1$$

$$= 2(0.05675 - 8)$$

$$= -15.8866$$

$$\frac{\partial y}{\partial v} = \begin{bmatrix} \frac{\partial \hat{y}}{\partial v_0} & \frac{\partial \hat{y}}{\partial v_1} & \frac{\partial \hat{y}}{\partial v_2} & \frac{\partial \hat{y}}{\partial v_3} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.52248 & 0.51999 & 0.51749 \end{bmatrix}$$

$$\frac{\partial \hat{y}}{\partial w} = \begin{bmatrix} \partial\hat{y}/\partial w_{10} & \partial\hat{y}/\partial w_{11} & \partial\hat{y}/\partial w_{12} \\ \partial\hat{y}/\partial w_{20} & \partial\hat{y}/\partial w_{21} & \partial\hat{y}/\partial w_{22} \\ \partial\hat{y}/\partial w_{30} & \partial\hat{y}/\partial w_{22} & \partial\hat{y}/\partial w_{32} \end{bmatrix}.$$

$$= \begin{bmatrix} v_1 m(1-m) & v_1 m(1-m)x_{i1} & v_1 m(1-m)x_{i2} \\ v_2 n(1-n) & v_2 n(1-n)x_{i1} & v_2 n(1-n)x_{i2} \\ v_3 o(1-o) & v_3 o(1-o)x_{i1} & v_3 o(1-o)x_{i2} \end{bmatrix}$$

$$= \begin{bmatrix} 0.00499 & 0.00499 & 0.00996 \\ 0.00748 & 0.00748 & 0.0149 \\ 0.00998 & 0.00998 & 0.0199 \end{bmatrix}$$

$$\Delta v = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial v}$$

$$\Delta v = \begin{bmatrix} -15.8865 & -8.3 & -8.26 & -8.22 \end{bmatrix}$$

$$\Delta w = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w}$$

$$\Delta w = \begin{bmatrix} -0.07927 & -0.07927 & -0.15854 \\ -0.11896 & -0.11896 & -0.23793 \\ -0.15867 & -0.15867 & -0.31734 \end{bmatrix}.$$

for $x = [1 \;\; 1 \;\; 3]$ $\quad y = 11$

$$\hat{y} = 0.057148$$

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y) = -21.8857$$

$$\frac{\partial \hat{y}}{\partial v} = \begin{bmatrix} \frac{\partial \hat{y}}{\partial v_0} & \frac{\partial \hat{y}}{\partial v_1} & \frac{\partial \hat{y}}{\partial v_2} & \frac{\partial \hat{y}}{\partial v_3} \end{bmatrix} = [1 \;\; 0.52996 \;\; 0.5$$

0.5

$$\Delta v = \begin{bmatrix} -21.886 & -11.598 & -11.489 & -11.380 \end{bmatrix}.$$

6

$$\frac{\partial y}{\partial w} = \begin{bmatrix} 0.005 & 0.005 & 0.01 \\ 0.00712 & 0.00712 & 0.024 \\ 0.0099 & 0.0099 & 0.028 \end{bmatrix}$$

$$\Delta w \; \cancel{\frac{\partial y}{\partial w}} = \begin{bmatrix} -0.109 & -0.109 & -0.327 \\ -0.163 & -0.163 & -0.491 \\ -0.218 & -0.218 & -0.655 \end{bmatrix}$$

$\cancel{\Delta w} =$ for $x = [1, 2\ 2]$.

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y) = 2(0.05722 - 10)$$
$$= -19.885$$

$$\frac{\partial \hat{y}}{\partial w} = [1 \quad 0.527474 \quad 0.52248 \quad 0.51999]$$

$$\Delta v = [-19.885 \quad -10.489 \quad -10.389 \quad -10.439]$$

$$\frac{\partial y}{\partial w} = \begin{bmatrix} 0.00499 & 0.00998 & 0.00998 \\ 0.00748 & 0.0149 & 0.0149 \\ 0.00998 & 0.01996 & 0.01996 \end{bmatrix}$$

$$\Delta w = \begin{bmatrix} -0.099 & -0.198 & -0.198 \\ -0.148 & -0.297 & -0.297 \\ -0.198 & -0.396 & -0.396 \end{bmatrix}$$

(d) $$v_j \leftarrow v_j - \eta \sum_{i=1}^{m} (\hat{y}_j^{(i)} - y_{j}^{(i)}) z^{(i)}$$

$$\Delta v_j = \sum_{i=1}^{m} (\hat{y}_j^i - y_j^i) z^i$$

$$z^i = \text{sigmoid}(w_j^T x).$$

$$w \leftarrow w_j - \eta \sum_{i=1}^{m} \sum_{l=1}^{k} (\hat{y}_l^i - y_l^i) v_{kj} z_j (1 - z_j^i) x^i$$

$$z = [m \quad n \quad o]$$

for first datapoint $[(1,2), 8]$.

$$\Delta v_j = (\hat{y}_j^i - y_j^i)\,[1 \quad m \quad n \quad o].$$

$$\Delta v_j = [-15.8865 \quad -8.3 \quad -8.26 \quad -8.22]$$

$$\Delta w_1 = (-15.8865)(0.2)(0.52248)(1-0.52248)\,[1 \quad 1 \quad 2].$$

$$\Delta w_1 = [-0.07927 \quad -0.07927 \quad -0.15854]$$

$$\Delta w_2 = (-15.8864)(0.3)(0.51999)(1-0.51999)\,[1 \quad 1 \quad 2]$$

$$= [-0.11896 \quad -0.11896 \quad -0.23793]$$

$$\Delta w_3 = [-15.8864](0.4)(0.51749)(1-0.51749)\,[1 \quad 1 \quad 2]$$

$$= [-0.15867 \quad -0.15867 \quad -0.31734]$$

for the datapoint $[(1,3), 11]$.

$$\Delta v_j = [-21.886 \quad -11.598 \quad -11.489 \quad -11.380].$$

$$\Delta w_1\ (-21.886)(0.2)(0.522996)(1-0.522996)\,[1 \quad 1 \quad 3]$$

$$\Delta w_1 = [-0.109 \quad -0.109 \quad -0.327]$$

$$\Delta w_2 = (-21.886)(0.3)(0.52498)(1-0.52498)\,[1 \quad 1$$

$$\Delta w_2 = [-0.163 \quad -0.163 \quad -0.491]$$

$\Delta w_3 = (-21.886)(0.4)(0.51999)(1-0.51999)\,[1 \quad 1 \quad 3]$

$\Delta w_3 = [-0.218 \quad -0.218 \quad -0.655]$.

for the datapoint [(2,2), 10].

$\Delta v = [-19.885 \quad -10.489 \quad -10.389 \quad -10.439]$.

$\Delta w_1 = (-19.885)(0.2)(0.527474)(1-0.527474)\,[1 \quad 2 \quad 2]$

$\Delta w_1 \; [-0.099 \quad -0.099 \quad -0.198]$

$\Delta w_2 = (-19.885)(0.3)(0.52248)(1-0.52248)\,[1 \quad 1 \quad 3]$

$\Delta w_2 \; [-0.148 \quad -0.297 \quad -0.297]$

$\Delta w_3 = (-19.885)(0.4)(0.51999)(1-0.59999)\,[1 \quad 1 \quad 3]$

$\Delta w_3 = [-0.498 \quad -0.796 \quad -0.376]$

It is seen that the weights obtaind in part d is almost the same as obtained in part c, Hence the values are matching.

2.

a) $f(x, y) = (2x + 3y)^2$

$$\nabla f(x, y) = \begin{bmatrix} \partial f / \partial x \\ \partial f / \partial y \end{bmatrix}$$

$$\frac{\partial f}{\partial x} = 2(2x + 3y)2 = 8x + 12y$$

$$\frac{\partial f}{\partial y} = 2(2x + 3y)3 = 12x + 18y$$

$$\nabla f(x, y) = \begin{bmatrix} 8x + 12y \\ 12x + 18y \end{bmatrix}$$

b) $F(x, y) = \begin{bmatrix} x^2 + 2y \\ 3x + 4y^2 \end{bmatrix}.$

$$DF(x, y) = \begin{bmatrix} f_1(x, y) \\ f_2(x, y) \end{bmatrix} = \begin{bmatrix} f_1(x^2 + 2y) \\ f_2(3x + 4y^2) \end{bmatrix}$$

$$= \begin{bmatrix} \partial f_1 / \partial x & \partial f_1 / \partial y \\ \partial f_2 / \partial x & \partial f_2 / \partial y \end{bmatrix}$$

$$DF(x, y) = \begin{bmatrix} 2x & 2 \\ 3 & 8y \end{bmatrix}$$

$$DF(1, 2) = \begin{bmatrix} 2 & 2 \\ 3 & 16 \end{bmatrix}.$$

c) $G(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$ $\quad F(x, y) = \begin{bmatrix} x^2 + 2y \\ 3x + 4y^2 \end{bmatrix}$

$D(F \circ G)(x)$ using chain rule

$= DF(G(x)) D.G(x)$

$$D(F \circ G)(x) = \begin{bmatrix} 2x & 2 \\ 3 & 8x^2 \end{bmatrix} \begin{bmatrix} 1 \\ 2x \end{bmatrix} = \begin{bmatrix} 6x \\ 3 + 16x^3 \end{bmatrix}$$

$$D(F \circ G)(2) = \begin{bmatrix} 12 \\ 3 + (16 \times 8) \end{bmatrix} = \begin{bmatrix} 12 \\ 131 \end{bmatrix}$$

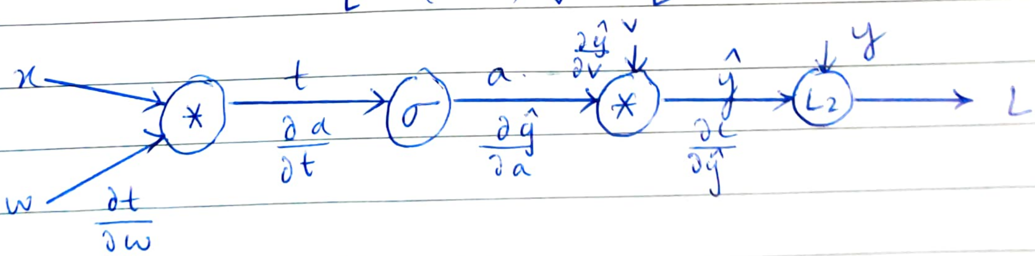
$D(F \circ G)(x)$ without using chain rule

$$F(G(x)) = \begin{bmatrix} 3x^2 \\ 3x + 4x^4 \end{bmatrix}$$

$$D[F(G(x))] = \begin{bmatrix} 6x \\ 3 + 16x^3 \end{bmatrix}$$

$$D[F(G(2))] = \begin{bmatrix} 12 \\ 3 + (16 \times 8) \end{bmatrix} = \begin{bmatrix} 12 \\ 131 \end{bmatrix}$$

d)



Point 1 $\quad x^T = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}$

$$w = \begin{bmatrix} 0.01 & 0.02 & 0.03 \\ 0.03 & 0.01 & 0.02 \\ 0.02 & 0.03 & 0.01 \end{bmatrix}$$

$v = \begin{bmatrix} 0.01 & 0.02 & 0.03 & 0.04 \end{bmatrix}$

$y = 8$

Forward Propogation for $x = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}$

$$a = \text{sigmoid}(wx) = \text{sigmoid}\begin{bmatrix} 0.09 \\ 0.08 \\ 0.07 \end{bmatrix} = \begin{bmatrix} 0.52248 \\ 0.51999 \\ 0.51749 \end{bmatrix}.$$

$$a' = \begin{bmatrix} 1 \\ 0.52248 \\ 0.51999 \\ 0.51749 \end{bmatrix}$$

$$\hat{y} = v a' = \begin{bmatrix} 0.01 & 0.02 & 0.03 & 0.04 \end{bmatrix}\begin{bmatrix} 1 \\ 0.52248 \\ 0.51999 \\ 0.51749 \end{bmatrix}$$

$$\hat{y} = 0.05675$$

$$L = 2(\hat{y} - y)^2 = (0.05675 - 8)^2$$
$$= 63.095.$$

Forward propogation for $x = \begin{bmatrix} 1 & 1 & 3 \end{bmatrix}$.

$y = 11$

$$a = \text{sigmoid}(wx) = \text{sigmoid}\begin{bmatrix} 0.12 \\ 0.1 \\ 0.08 \end{bmatrix} = \begin{bmatrix} 0.52996 \\ 0.52498 \\ 0.51999 \end{bmatrix}.$$

$$a' = \begin{bmatrix} 1 \\ 0.52996 \\ 0.52498 \\ 0.51999 \end{bmatrix}$$

$$\hat{y} = Va' = 0.057148$$

$$L = (\hat{y} - y)^2 = (0.057148 - 11)^2 = 119.746$$

Forward propogation for $x = [1 \ 2 \ 2]$

$y = 10$

$$H = \text{sigmoid}(w.x) = \text{sigmoid}\begin{bmatrix} 0.11 \\ 0.09 \\ 0.10 \end{bmatrix} = \begin{bmatrix} 0.52747 \\ 0.52248 \\ 0.51999 \end{bmatrix}$$

$$a' = \begin{bmatrix} 1 \\ 0.52747 \\ 0.52248 \\ 0.52498 \end{bmatrix}$$

$$\hat{y} = Va' = \begin{bmatrix} 0.01 & 0.02 & 0.03 & 0.04 \end{bmatrix} \begin{bmatrix} 1 \\ 0.52747 \\ 0.52248 \\ 0.52498 \end{bmatrix}$$

$$\hat{y} = 0.05722$$

$$L = (\hat{y} - y)^2$$

$$L = (0.05722 - 10)^2$$

$$L = 98.858$$

Backward propogation for $x = [1 \quad 1 \quad 2]$

$$\frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial v}$$

$$\frac{\partial L}{\partial v} = 2a'(\hat{y} - y) = \left[2\begin{bmatrix}1\\0.52248\\0.51999\\0.51749\end{bmatrix}(0.05675 - 8)\right]^T$$

$$= [-15.8865 \quad -8.3 \quad -8.26 \quad -8.22].$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial a}\frac{\partial a}{\partial w}$$

$$= x \times [a'(1-a')v^T \, 2(\hat{y} - y)]$$

$$\cancel{= \begin{bmatrix}1\\1\\2\end{bmatrix} \times \left[\begin{pmatrix}0.52996\\0.52498\\0.51999\end{pmatrix}\begin{pmatrix}1-0.52996\\1-0.52498\\1-0.51999\end{pmatrix} \cdot \begin{bmatrix}0.02\\0.03\\0.04\end{bmatrix} \cdot 2(0.057148 -\right.}$$

$$= \begin{bmatrix}1\\1\\2\end{bmatrix} \times \left[\begin{pmatrix}0.52248\\0.51999\\0.51749\end{pmatrix} \times \begin{pmatrix}1-0.52248\\1-0.51999\\1-0.51749\end{pmatrix} \cdot \begin{bmatrix}0.02\\0.03\\0.04\end{bmatrix}^T \cdot 2(0.05675 - 8)\right]^T$$

$$= \begin{bmatrix}1\\1\\2\end{bmatrix}[-0.07927 \quad -0.11896 \quad -0.15867].$$

$$= \begin{bmatrix}-0.07927 & -0.07927 & -0.15854\\ -0.11896 & -0.11896 & -0.23793\\ -0.15867 & -0.15867 & -0.31734\end{bmatrix}$$

//

Backward propogation of $x = [1, 1, 3]$.

$y = 11$

$$\frac{\partial L}{\partial V} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial v} = a' \cdot 2 \cdot (\hat{y} - y) = \begin{bmatrix} 1 \\ 0.52996 \\ 0.52498 \\ 0.51999 \end{bmatrix}^T \cdot 2 \cdot (0.0514 - 11)$$

$$= [-21.886 \quad -11.598 \quad -11.499 \quad -11.380]$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial a} \frac{\partial a}{\partial w} = x \times [a'(1-a') v^T 2(\hat{y} - y)]^T$$

$$= \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} \times \left[\begin{pmatrix} 0.52996 \\ 0.52498 \\ 0.51999 \end{pmatrix} \begin{pmatrix} 1 - 0.52996 \\ 1 - 0.52498 \\ 1 - 0.51199 \end{pmatrix} \cdot \begin{bmatrix} 0.02 \\ 0.03 \\ 0.04 \end{bmatrix}^T \cdot 2 \cdot (0.05714 - 11) \right]^T$$

$$= \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} [-0.109 \quad -0.163 \quad -0.218]$$

$$= \begin{bmatrix} -0.109 & -0.109 & -0.327 \\ -0.163 & -0.163 & -0.491 \\ -0.218 & -0.218 & -0.655 \end{bmatrix}$$

Backward propogation of $x = [1, 2, 2]$.

$y = 10$

$$\frac{\partial L}{\partial v} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial v} = a' \cdot 2 \cdot (\hat{y} - y) = \begin{bmatrix} 1 \\ 0.52747 \\ 0.52248 \\ 0.51999 \end{bmatrix} \cdot 2 \cdot (0.0572 - 10)$$

$$= [-19.885 \quad -10.489 \quad -10.389 \quad -10.439]$$

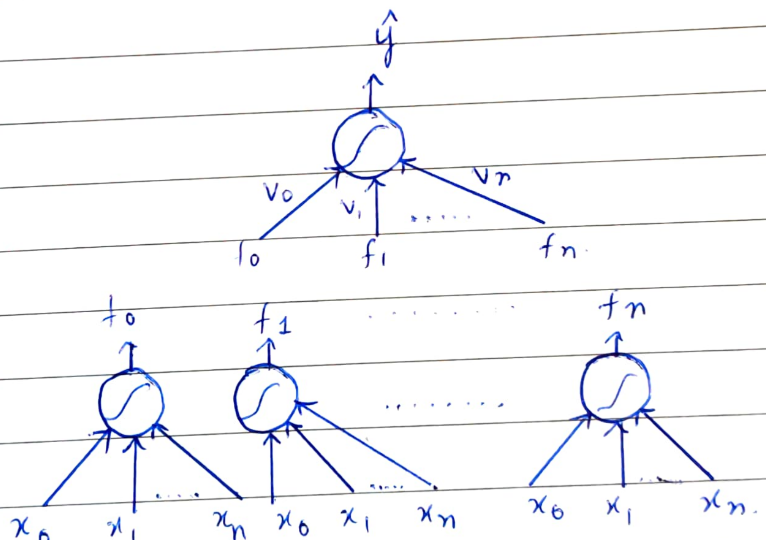
$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial a} \frac{\partial a}{\partial W} = x \times [a'(1-a') v^T 2(\hat{y} - y)]$$

$$= \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \left[\left(\begin{pmatrix} 0.52747 \\ 0.52248 \\ 0.51999 \end{pmatrix} \begin{pmatrix} 1-0.52747 \\ 1-0.52248 \\ 1-0.51999 \end{pmatrix} \right) \cdot \begin{bmatrix} 0.02 \\ 0.03 \\ 0.04 \end{bmatrix} \cdot 2.(0.05722 - 10) \right]^T$$

$$= \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \begin{bmatrix} -0.099 & -0.148 & -0.198 \end{bmatrix}$$

$$= \begin{bmatrix} -0.099 & -0.198 & -0.198 \\ -0.148 & -0.297 & -0.297 \\ -0.198 & -0.396 & -0.396 \end{bmatrix}$$

3.



given training set $= \{ x^{(i)}, y^{(i)} \}_{i=1}^{m}$

$$E = 1/2 \left(\sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)})^2 \right)$$

we have

$$\frac{\partial E}{\partial V} = \frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial V} = \left[2 \cdot \frac{1}{2} \cdot \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \right] f'^{(i)}$$

$$= \left[\sum_{i=1}^{m} \hat{y}^{(i)} - y^{(i)} \right] \left[\hat{y}^{(i)} (1 - \hat{y}^{(i)}) \right.$$

$$v \leftarrow v - \eta \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)})\, \hat{y}^{(i)} (1 - \hat{y}^{(i)})\, f^{(i)}$$

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial f_i} \frac{\partial f_i}{\partial w_j}$$

$$\frac{\partial E}{\partial w} = \left[\sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)})\right] v_j \left[f_j^{(i)} (1 - f_j^{(i)}) \cdot x^{(i)}\right].$$

$$w \leftarrow w - \eta \left[\sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)})\right] v_j \left[f_j^{(i)} \cdot (1 - f_j^{(i)}) \cdot x^{(i)}\right].$$