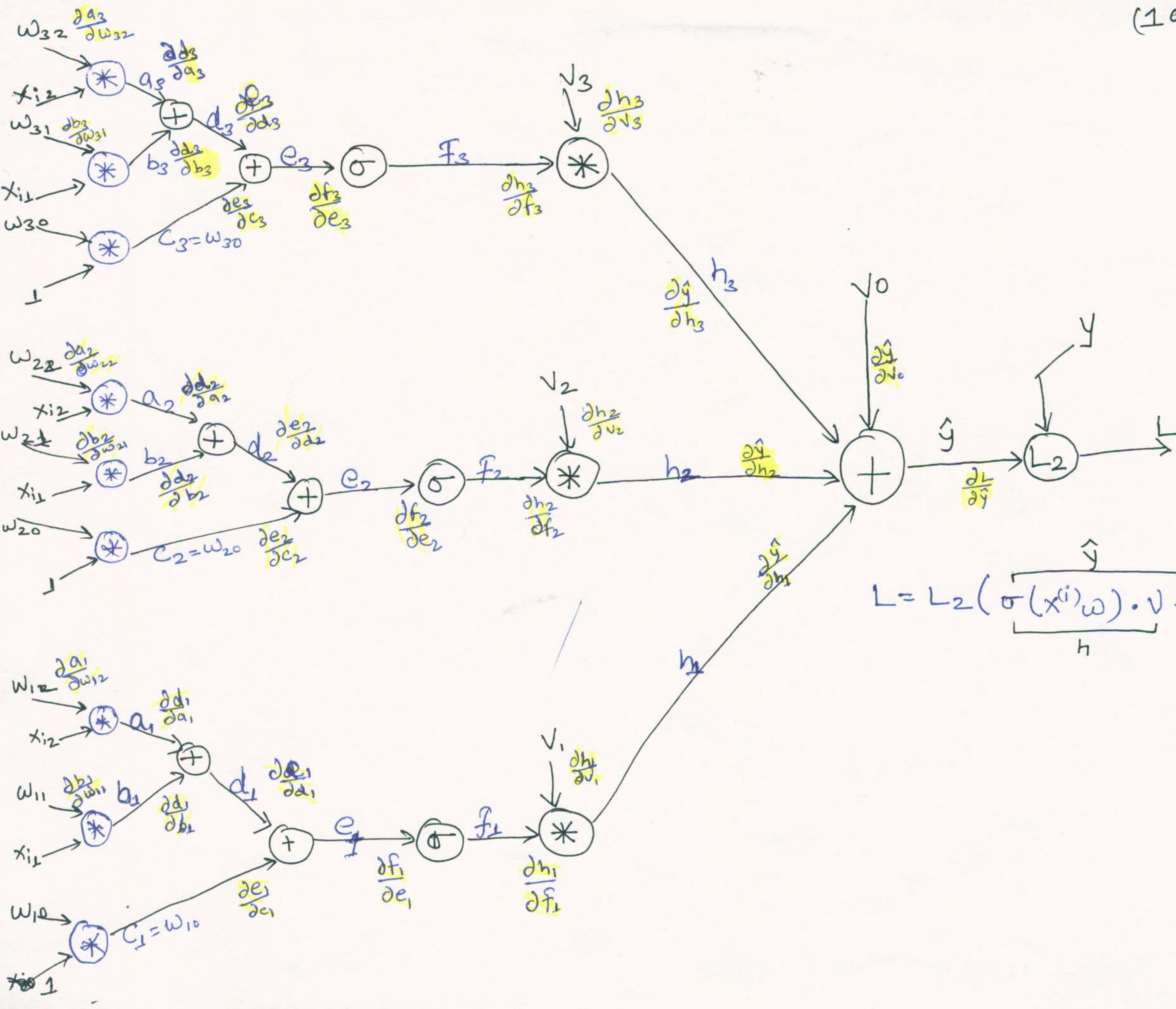
(1a.) Arinjay Jain
A20447307



$$L = L_2\Big(\overbrace{\sigma\underbrace{(x^{(i)}\omega)\cdot v}_{h} + v_0}^{\hat{y}},\ y\Big)$$

$$w_1 = (\overset{w_{10}}{0.01}, \overset{w_{11}}{0.02}, \overset{w_{12}}{0.03}) \qquad V = (\overset{v_0}{0.01}, \overset{v_1}{0.02}, \overset{v_2}{0.03}, \overset{v_3}{0.04})$$

$$w_2 = (\overset{w_{20}}{0.03}, \overset{w_{21}}{0.01}, \overset{w_{22}}{0.02}) \qquad x_1, y_1 = [(\overset{x_{11}}{1}, \overset{x_{12}}{2}), 8]$$

$$x_2, y_2 = [(\overset{x_{21}}{1}, \overset{x_{22}}{3}), 11]$$

$$w_3 = (\overset{w_{30}}{0.02}, \overset{w_{31}}{0.03}, \overset{w_{32}}{0.01}) \qquad x_3, y_3 = [(\overset{x_{31}}{2}, \overset{x_{33}}{2}), 10]$$

for x_1 $(x_{11} = 1, x_{12} = 2)$.

$a_1 = x_{12} \cdot w_{12} = 2 \cdot (0.03) = 0.06$; $b_1 = x_{11} \cdot w_{11} = 1 \cdot (0.02) = 0.02$; $c_1 = w_{10} = 0.01$

$d_1 = a_1 + b_1 = 0.06 + 0.02 = 0.08$; $e_1 = d_1 + c_1 = 0.09$, $\cancel{f_1 =}$

$$f_1 = \text{sigmoid}(e_1) = \frac{\exp(0.09)}{1+\exp(0.09)} = \cancel{0.5}\ \boxed{0.52 = f_1}$$

$h_1 = f_1 \cdot v_1 = 0.52 \times 0.02 = 0.0104$

$a_2 = x_{12} \cdot w_{22} = 2 \times 0.02 = 0.04$; $b_2 = x_{11} \cdot w_{21} = 1 \times 0.01 = 0.01$, $c_2 = w_{20} = 0.03$

$d_2 = a_2 + b_2 = 0.05$; $e_2 = d_2 + c_2 = 0.08$; $f_2 = \text{sigmoid}(e_2) = \dfrac{\exp(0.08)}{1+\exp(0.08)}$

$f_2 = 0.52$, $h_2 = f_2 \cdot v_2 = 0.52 \times 0.03 = 0.015$

$a_3 = x_{12} \cdot w_{32} = 2 \times (0.01) = 0.02$; $b_3 = x_{11} \cdot w_{31} = 1 \times 0.03 = 0.03$; $c_3 = 0.02$

$e_3 = d_3 + c_3 = a_3 + b_3 + c_3 = 0.05 + 0.02 = 0.07$; $f_3 = \text{sigmoid}(e_3) = \dfrac{\exp(0.07)}{1+\exp(0.07)}$

$f_3 = 0.517 \simeq 0.52$; $h_3 = f_3 \cdot v_3 = 0.52 \times 0.04 = 0.0207$

$$\hat{y}_1 = h_1 + h_2 + h_3 + v_0 = 0.0104 + 0.015 + 0.0207 + 0.01$$

$$\boxed{\hat{y}_1 = 0.0561}$$

for X_2 $(X_{21}=1, X_{22}=3)$.

$a_1 = X_{22} W_{12} = 3\times(0.03) = 0.09$; $b_1 = X_{21} \cdot W_{11} = 1\times(0.02) = 0.02$; $c_1 = W_{10} = 0.01$

$d_1 = a_1 + b_1 = 0.09 + 0.02 = 0.11$; $e_1 = d_1 + c_1 = 0.12$, $f_1 = \frac{\exp(0.12)}{1+\exp(0.12)} = 0.53$

$h_1 = f_1 \cdot V_1 = 0.53 \times 0.02 = 0.0106$

$a_2 = X_{22} \cdot W_{22} = 3\times 0.02 = 0.06$; $b_2 = X_{21} \cdot W_{21} = 1 \cdot 0.01 = 0.01$; $c_2 = W_{20} = 0.03$

$d_2 = a_2 + b_2 = 0.06 + 0.01 = 0.07$; $e_2 = d_2 + c_2 = 0.10$; $f_2 = \frac{\exp(0.10)}{1+\exp(0.10)} = 0.52$

$h_2 = f_2 \cdot V_2 = 0.52 \times 0.03 = 0.0156$

$a_3 = X_{22} \cdot W_{32} = 3 \cdot 0.01 = 0.03$; $b_3 = X_{21} \cdot W_{31} = 1\times 0.03 = 0.03$; $C_3 = 0.02$

$e_3 = 0.08$, $f_3 = 0.52$, $h_3 = f_3 \times V_3 = 0.52 \times 0.04 = 0.0207$.

$\hat{Y}_2 = h_1 + h_2 + h_3 + V_0 = 0.0106 + 0.0156 + 0.0207 + 0.01$

$\boxed{\hat{Y}_2 = 0.057}$

for X_3 $(X_{31}=2, X_{32}=2)$

$a_1 = X_{32} \cdot W_{12} = 2\times 0.03 = 0.06$; $b_1 = 0.04$; $c_1 = 0.01$; $e_1 = 0.11$; $f_1 = 0.53$; $h_1 = 0.0106$

$a_2 = 0.04$, $b_2 = 0.02$, $c_2 = 0.03$, $e_2 = 0.09$, $f_2 = 0.52$, $h_2 = 0.015$

$a_3 = 0.02$, $b_2 = 0.06$, $C_3 = 0.02$, $e_3 = 0.1$, $f_3 = 0.52$, $h_3 = 0.0207$

$\hat{Y}_3 = h_1 + h_2 + h_3 + V_0$

$\boxed{\hat{Y}_3 = 0.057}$

1c. Arinjay Jain (A20447307)

$\therefore L = (\hat{y} - y)^2 \Rightarrow \boxed{\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y)}$

Now backpropagation comes in a picture.

$\hat{y} = h_1 + h_2 + h_3 + v_0 \rightarrow \frac{\partial \hat{y}}{\partial h_1} = 1 = \frac{\partial \hat{y}}{\partial h_2} = \frac{\partial \hat{y}}{\partial h_3} = \frac{\partial \hat{y}}{\partial v_0} = 1$

$\frac{\partial h_1}{\partial v_1} = f_1$, $\frac{\partial h_2}{\partial v_2} = f_2$, $\frac{\partial h_3}{\partial v_3} = f_3$, $\frac{\partial h_1}{\partial f_1} = v_1$, $\frac{\partial h_2}{\partial f_2} = v_2$, $\frac{\partial h_3}{\partial f_3} = v_3$

$\frac{\partial f_1}{\partial e_1} = f_1(1-f_1)$, $\frac{\partial f_2}{\partial e_2} = f_2(1-f_2)$; $\frac{\partial f_3}{\partial e_3} = f_3(1-f_3)$; $\frac{\partial e_1}{\partial d_1} = 1$ $\frac{\partial e_1}{\partial c_1} = 1$

↳ derivation of sigmoid function.

$\frac{\partial d_1}{\partial a_1} = 1$ $\frac{\partial d_1}{\partial b_1} = 1$ $\frac{\partial a_1}{\partial w_{12}} = x_{i2}$ $\frac{\partial b_1}{\partial w_{11}} = x_{i1}$

$\frac{\partial c_1}{\partial w_{10}} = 1$, $\frac{\partial c_2}{\partial w_{20}} = 1$, $\frac{\partial c_3}{\partial w_{30}} = 1$ $\frac{\partial a_2}{\partial w_{22}} = x_{i2}$ $\frac{\partial b_2}{\partial w_{21}} = x_{i1}$

Gradients of Loss function with respect to w_1, w_2, w_3, v

$\frac{\partial a_3}{\partial w_{32}} = x_{i2}$ $\frac{\partial b_3}{\partial w_{31}} = x_{i1}$

\# $\frac{\partial \hat{y}}{\partial w_{10}} = \frac{\partial \hat{y}}{\partial h_1} \cdot \frac{\partial h_1}{\partial f_1} \cdot \frac{\partial f_1}{\partial e_1} \cdot \frac{\partial e_1}{\partial c_1} \cdot \frac{\partial c_1}{\partial w_{10}} = 1 \cdot v_1 \cdot f_1(1-f_1) \cdot 1 \cdot 1 = \underline{v_1 f_1(1-f_1)}$

\# $\frac{\partial \hat{y}}{\partial w_{11}} = \frac{\partial \hat{y}}{\partial h_1} \cdot \frac{\partial h_1}{\partial f_1} \cdot \frac{\partial f_1}{\partial e_1} \cdot \frac{\partial e_1}{\partial d_1} \cdot \frac{\partial d_1}{\partial b_1} \cdot \frac{\partial b_1}{\partial w_{11}} = 1 \cdot v_1 \cdot f_1(1-f_1) \cdot 1 \cdot 1 \cdot x_{i1} = v_1 f_1(1-f_1) x_{i1}$

$= \frac{\partial \hat{y}}{\partial w_{10}} \cdot x_{i1}$

\# $\frac{\partial \hat{y}}{\partial w_{12}} = \underline{v_1 \cdot f_1(1-f_1) x_{i2}}$; # $\frac{\partial \hat{y}}{\partial w_{20}} = \underline{v_2 (f_2(1-f_2))}$

\# $\frac{\partial \hat{y}}{\partial w_{21}} = v_2 f_2(1-f_2) x_{i1}$; # $\frac{\partial \hat{y}}{\partial w_{22}} = v_2 f_2(1-f_2) x_{i2}$

\# $\frac{\partial \hat{y}}{\partial w_{30}} = v_3 f_3 (1-f_3)$; # $\frac{\partial \hat{y}}{\partial w_{31}} = v_3 f_3(1-f_3) x_{i1}$, # $\frac{\partial \hat{y}}{\partial w_{32}} = v_3 f_3(1-f_3) x_{i2}$

\# $\frac{\partial \hat{y}}{\partial v_1} = \frac{\partial \hat{y}}{\partial h_1} \cdot \frac{\partial h_1}{\partial v_1} = 1 \cdot f_1 = f_1$, # $\frac{\partial \hat{y}}{\partial v_2} = f_2$, # $\frac{\partial \hat{y}}{\partial v_3} = f_3$

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y}-y) \Rightarrow$$ # for x_1 ($x_{11}=1$, $x_{12}=2$).

$$\hat{y}_{x_1} = 0.0561 \quad y = 8$$

$$\frac{\partial L}{\partial \hat{y}_{x_1}} = 2(0.0561-8) = -15.88$$

$$\frac{\partial \hat{y}}{\partial v} = \left[\frac{\partial \hat{y}}{\partial v_0}, \frac{\partial \hat{y}}{\partial v_1}, \frac{\partial \hat{y}}{\partial v_2}, \frac{\partial \hat{y}}{\partial v_3}\right] = [1, f_1, f_2, f_3] = [1, 0.52, 0.52, 0.52]$$

$$\frac{\partial \hat{y}}{\partial w} = \begin{bmatrix} \frac{\partial \hat{y}}{\partial w_{10}} & \frac{\partial \hat{y}}{\partial w_{11}} & \frac{\partial \hat{y}}{\partial w_{12}} \\ \frac{\partial \hat{y}}{\partial w_{20}} & \frac{\partial \hat{y}}{\partial w_{21}} & \frac{\partial \hat{y}}{\partial w_{22}} \\ \frac{\partial \hat{y}}{\partial w_{30}} & \frac{\partial \hat{y}}{\partial w_{31}} & \frac{\partial \hat{y}}{\partial w_{32}} \end{bmatrix} = \begin{bmatrix} v_1 f_1(1-f_1) & v_1 f_1(1-f_1)\cdot x_{11} & v_1 f_1(1-f_2) x_{12} \\ v_2 f_2(1-f_2) & v_2 f_2(1-f_2) x_{11} & v_2 f_2(1-f_2) x_{12} \\ v_3 f_3(1-f_3) & v_3 f_3(1-f_3) x_{11} & v_3 f_3(1-f_3) x_{12} \end{bmatrix}$$

$$\frac{\partial \hat{y}}{\partial w_{(x_1)}} = \begin{bmatrix} \cancel{0.00499} & \cancel{0.00499} & \cancel{0.00998} \\ \cancel{0.0075} & \cancel{0.0075} & \cancel{0.015} \\ & & \end{bmatrix} \begin{bmatrix} 0.00499 & 0.00499 & 0.00998 \\ 0.00748 & 0.00748 & 0.0149 \\ 0.00998 & 0.00998 & 0.0199 \end{bmatrix}$$

$$\Delta v = \frac{\partial l}{\partial v} = \left[\frac{\partial l}{\partial v_0}, \frac{\partial l}{\partial v_1}, \frac{\partial l}{\partial v_2}, \frac{\partial l}{\partial v_3}\right] = \left[\frac{\partial L}{\partial \hat{y}}\cdot\frac{\partial \hat{y}}{\partial v_0}, \frac{\partial L}{\partial \hat{y}}\cdot\frac{\partial \hat{y}}{\partial v_1}, \frac{\partial L}{\partial \hat{y}}\cdot\frac{\partial \hat{y}}{\partial v_2}, \frac{\partial L}{\partial \hat{y}}\cdot\frac{\partial \hat{y}}{\partial v_3}\right]$$

$$= \cancel{\frac{\partial L}{\partial \hat{y}}[1, 0.52,}\quad \frac{\partial L}{\partial \hat{y}}[1, f_1, f_2, f_3] = -15.88[1, 0.52, 0.52, 0.52]$$

$$\Delta v = \frac{\partial l}{\partial v} = [-15.88, -8.25, -8.25, -8.25]$$

$$\Delta w = \frac{\partial l}{\partial w} = \frac{\partial l}{\partial \hat{y}} \left[\frac{\partial \hat{y}}{\partial w}\right] \Rightarrow \begin{bmatrix} -0.079 & -0.079 & -0.1584 \\ -0.1187 & -0.1187 & -0.2375 \\ -0.1584 & -0.1584 & -0.3168 \end{bmatrix}$$

~~*~~ for x_2 2^{nd} training example.

$(x_{21} = 1, x_{22} = 3)$ $\hat{y}_{(x_2)} = 0.057$ $y = 11$ $\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y)$

$= 2(0.057 - 11)$

$$\frac{\partial \hat{y}}{\partial v} = [1, f_1, f_2, f_3] = [1, 0.53, 0.52, 0.52] \quad \boxed{\frac{\partial l}{\partial \hat{y}} = -21.886,}$$

$$\frac{\partial \hat{y}}{\partial w} = \begin{bmatrix} v_1 f_1 (1-f_1) & v_1 f_1 (1-f_1) x_{21} & v_1 f_1 (1-f_2) x_{22} \\ v_2 f_2 (1-f_2) & v_2 f_2 (1-f_2) x_{21} & v_2 f_2 (1-f_2) x_{22} \\ v_3 f_3 (1-f_3) & v_3 f_3 (1-f_3) x_{21} & v_3 f_3 (1-f_3) x_{22} \end{bmatrix}$$

$$= \begin{bmatrix} 0.00498 & 0.00498 & 0.0149 \\ 0.00748 & 0.00748 & 0.02244 \\ 0.00998 & 0.00998 & 0.02994 \end{bmatrix}$$

$$\nabla v = \frac{\partial l}{\partial v} = -21.886 [1, 0.53, 0.52, 0.52] = [-21.886, -11.599, -11.38, -11.38]$$

∇w $\xrightarrow{}$ Gradient of loss funtion w.r.t (ω)

$$\frac{\partial l}{\partial \omega} = \frac{\partial l}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \omega} = \cancel{-21.886} \begin{bmatrix} -0.1089, & -0.1089, & -0.326 \\ -0.1637, & -0.1637 & -0.4911 \\ -0.2184 & -0.2184 & -0.6552 \end{bmatrix}$$

\# for x_3 $(x_{31} = 2, x_{32} = 2)$ 3^{rd} training example.

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y) = 2(0.057 - 10) = -19.886,$$

$$\frac{\partial \hat{y}}{\partial v} = [1, f_1, f_2, f_3] = [1, 0.53, 0.52, 0.52]$$

for x_3 $\quad x_{31} = 2 \quad x_{32} = 2$

$$\frac{\partial \hat{y}}{\partial w} = \begin{bmatrix} v_1 f_1(1-f_1) & v_1 f_1(1-f_1) x_{31} & v_1 f_1(1-f_1) x_{32} \\ v_2 f_2(1-f_2) & v_2 f_2(1-f_2) x_{31} & v_2 f_2(1-f_2) x_{32} \\ v_3 f_3(1-f_3) & v_3 f_3(1-f_3) x_{31} & v_3 f_3(1-f_3) x_{32} \end{bmatrix}$$

$$\frac{\partial \hat{y}}{\partial w} = \begin{bmatrix} 0.00498 & 0.00996 & 0.00996 \\ 0.00748 & 0.01496 & 0.01496 \\ 0.00998 & 0.01996 & 0.01996 \end{bmatrix}$$

$$\Delta v = \frac{\partial l}{\partial v} = -19.886\,[1, 0.53, 0.52, 0.52] = [-19.886, -10.539, -10.34, -10.34]$$

$$\Delta w = \frac{\partial l}{\partial w} = \frac{\partial l}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w} = -19.886 \begin{bmatrix} 0.00498 & 0.00996 & \cdots \\ 0.00748 & 0.01496 & \cdots \\ 0.00998 & 0.01996 & \cdots \end{bmatrix}$$

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

1 d)

$$\Delta v_j = \sum_{i=1}^{m} 2(\hat{y}_i^{(i)} - y_j^{(i)}) \cdot z^{(i)} \qquad \Delta w_j = \sum_{1}^{m} 2(\hat{y}^{(i)} - y^{(i)}) \cdot v_j \, z_j^{(i)}(1 - z_j^{(i)}) \times 1$$

$z^{(i)} \rightarrow$ vector of sigmoid function $[f_1, f_2, f_3]$

for $x_1 = [1, 2]$ $\quad y = [8]$

$$\Delta v_j = 2(\hat{y}_j^{(i)} - y_i^{(i)}) \, [\underbrace{1}, f_1, f_2, f_3] \rightarrow \text{adding a bias term}$$

$$= 2(0.0561 - 8] \, [1, 0.52, 0.52, 0.52]$$

$$= [-15.888, -8.262, -8.262, -8.262].$$

$$\Delta w_j = 2(\hat{y}^{(i)} - y^{(i)}) \cdot v_j \cdot z_j^{(i)}(1 - z_j^{(i)}) \, x^{(i)}$$

$$\Delta w_1 = (-15.888) \times 0.02 \times 0.52(1 - 0.52) \cdot [\underline{1}\ 1\ 2] \rightarrow \text{adding a bias in } x^{(i)}$$

$$\Delta w_1 = [-0.792 \ \ -0.792 \ \ -0.1584]$$

In the same way, we get Δw_2 & Δw_3

$$\Delta w_2 = 2(\hat{y}^{(i)} - y^{(i)}) \cdot v_2 \cdot z_2^{(i)}(1 - z_2^{(i)} \, x^{(i)})]$$

$$\Delta w_2 = [-0.1187, -0.1187, -0.2375]$$

$$\rightarrow \Delta w_3 = 2\,[\hat{y}^{(i)} - y^{(i)}] \cdot v_3 \cdot z_3^{(i)}(1 - z_3^{(i)}) \, x^{(i)}$$

$$= [-0.1584, -0.1584, -0.3168].$$

These are the same results we got in part 1(c)

for $x_2 = [1, 3]$.

$$\Delta v_2 = 2[\hat{y}_2^{(i)} - y_2^{(i)}] \, [1, f_1, f_2, f_3]$$

$$\Delta v_2 = [-21.886, -11.599, -11.38, -11.38],$$

$$\Delta w_j = 2(\hat{y}^{(i)} - y^{(i)}) \, v_j \, z_j^{(i)}(1 - z_j^{(i)}) \, x^{(i)}$$

$$\Delta w_1 = [-0.1089, -0.1089, -0.326]$$

$$\Delta w_2 = [-0.1637, -0.1637, -0.4911]$$

$$\Delta w_3 = [-0.2184, -0.2184, -0.6552].$$

Similarly for $x_3 = [2, 2]$

$$\Delta v_3 = [-19.886, -10.539, -10.34, -10.34]$$

$$\Delta w_1 = [-0.099 \ \ -0.198 \ \ -0.198]$$

$$\Delta w_2 = [-0.148 \ \ -0.297 \ \ -0.297]$$

$$\Delta w_3 = [-0.198 \ \ -0.396 \ \ -0.396]$$

Arinjay Jain (A20447307)

2. Vector Computation graph:→

$\nabla f(x, y) = (\partial f/\partial x, \partial f/\partial y)$

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

$\partial f/\partial x = 4(2x+3y)$ $\partial f/\partial y = 6(2x+3y)$.

$$\nabla f(x,y) = \begin{bmatrix} 4(2x+3y) \\ 6(2x+3y) \end{bmatrix}$$

b) Let $F(x,y) = \begin{bmatrix} x^2+2y \\ 3x+4y^2 \end{bmatrix}$. Compute the Jacobian matix $DF(1,2)$.

Jacobian matix →

$$\begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x & 2 \\ 3 & 8y \end{bmatrix} \xRightarrow{x=1,\, y=2} \begin{bmatrix} 2 & 2 \\ 3 & 16 \end{bmatrix}$$

c) $G(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$ $D(f\circ G)(2) = f'(G(2))\, G'(2)$

$G'(x) = \begin{bmatrix} 1 \\ 2x \end{bmatrix}$ $G'(2) = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$ $f'(G(2)) = \begin{bmatrix} 2\cdot 1 & 2 \\ 3 & 8\cdot 4 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 3 & 32 \end{bmatrix}$

$$D(f\circ G)(2) = \begin{bmatrix} 2 & 2 \\ 3 & 32 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 2+2\times 4 \\ 3+128 \end{bmatrix} = \begin{bmatrix} 12 \\ 131 \end{bmatrix}$$

Method 2

$f\circ G(x) = f(G(x))$ $G(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$ x → go at x in f, x^2 → go at y in f.

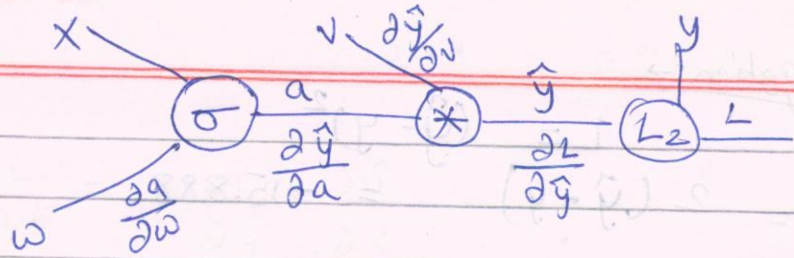
$$f\circ G(x) = f(G(x)) = \begin{bmatrix} x^2+2x^2 \\ 3x+4x^4 \end{bmatrix} = \begin{bmatrix} 3x^2 \\ 3x+4x^4 \end{bmatrix}_{2\times 1}$$

$$D(f\circ G) = f'\circ G(x) = \begin{bmatrix} 6x \\ 3+16x^3 \end{bmatrix} \underset{(x=2)}{=} \begin{bmatrix} 6\cdot 2 \\ 3+16(2)^3 \end{bmatrix}$$

$$= \begin{bmatrix} 12 \\ 131 \end{bmatrix}$$

2d ~~2~~ d task 1 2-d.

task1



$$L = L_2(\sigma(\omega x)v, y)$$

$$x^T = [1 \;\; 1 \;\; 2]$$

$$\omega = \begin{bmatrix} \omega_1^T \\ \omega_2^T \\ \omega_3^T \end{bmatrix} = \begin{bmatrix} 0.01 & 0.02 & 0.03 \\ 0.03 & 0.01 & 0.02 \\ 0.02 & 0.03 & 0.01 \end{bmatrix}$$

$$v = \begin{array}{cccc} [\,0.01 & 0.02 & 0.03 & 0.04\,] \\ v_0 & v_1 & v_2 & v_3 \end{array}$$

Forward Pass:→

$a = \sigma(\omega x)$ $\quad$ $\cancel{\hat{y} = a.v}$ $\quad$ $\hat{y} = v.a$

$$\omega x = \underset{3\times3}{\begin{bmatrix} 0.01 & 0.02 & 0.03 \\ 0.03 & 0.01 & 0.02 \\ 0.02 & 0.03 & 0.01 \end{bmatrix}} \underset{3\times1}{\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}} = \begin{bmatrix} 0.09 \\ 0.08 \\ 0.07 \end{bmatrix}$$

$$a = \sigma(x\omega) = \begin{bmatrix} \text{sigmoid}(0.09) \\ \text{sigmoid}(0.08) \\ \text{sigmoid}(0.07) \end{bmatrix} = \begin{bmatrix} 0.52248 \\ 0.51998 \\ 0.5174 \end{bmatrix}$$

$$\simeq \begin{bmatrix} 0.52 \\ 0.52 \\ 0.52 \end{bmatrix}$$

$\hat{y} = v.a$

$$= [0.01 \;\; 0.02 \;\; 0.03 \;\; 0.04] \begin{bmatrix} 1 \\ 0.52 \\ 0.52 \\ 0.52 \end{bmatrix}$$

bias term.

$$\boxed{\hat{y} = 0.056}$$

task 1 2 d)

Back propagation:→

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

$$\frac{\partial L}{\partial \hat{y}} = 2(\hat{y} - y) = -15.888.$$

$$\frac{\partial \hat{y}}{\partial v} = \begin{bmatrix} \partial\hat{y}/\partial v_0 \\ \partial\hat{y}/\partial v_1 \\ \partial\hat{y}/\partial v_2 \\ \partial\hat{y}/\partial v_3 \end{bmatrix} = \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.52 \\ 0.52 \\ 0.52 \end{bmatrix}$$

$$\frac{\partial \hat{y}}{\partial a} = \begin{bmatrix} \partial\hat{y}/\partial a_1 \\ \partial\hat{y}/\partial a_2 \\ \partial\hat{y}/\partial a_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0.02 \\ 0.03 \\ 0.04 \end{bmatrix}$$

$$\frac{\partial a}{\partial w} = \begin{bmatrix} \partial a_1/\partial w_1 \\ \partial a_2/\partial w_2 \\ \partial a_3/\partial w_3 \end{bmatrix} = \cancel{\begin{bmatrix} a_1(1-a_1) \\ a_2(1-a_2) \\ a_3(1-a_3) \end{bmatrix} \times \begin{bmatrix} x^T \\ x^T \\ x^T \end{bmatrix}}$$

derivation of sigmoid

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

$$= \cancel{\begin{bmatrix} 0.52(1-0.52) \\ 0.52(1-0.52) \\ 0.52(1-0.52) \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 2 \\ 1 & 1 & 2 \end{bmatrix}}$$

$$= \cancel{\begin{matrix} 0.249 \\ 0.249 \\ 0.249 \end{matrix}} \quad \{\because a = \sigma(wx)\} \quad \boxed{\frac{\partial a}{\partial w} = a(1-a)x}$$

$$\frac{\partial a}{\partial w} = \begin{bmatrix} \partial a_1/\partial w_1 \\ \partial a_2/\partial w_2 \\ \partial a_3/\partial w_3 \end{bmatrix} = \begin{bmatrix} a_1(1-a_1)\,x_1^T \\ a_2(1-a_2)\,x_2^T \\ a_3(1-a_3)\,x_3^T \end{bmatrix} = \begin{bmatrix} 0.249[1\ 1\ 2] \\ 0.249[1\ 1\ 2] \\ 0.249[1\ 1\ 2] \end{bmatrix}$$

$$\frac{\partial a}{\partial w} = \begin{bmatrix} 0.249 & 0.249 & 0.499 \\ 0.249 & 0.249 & 0.499 \\ 0.249 & 0.249 & 0.499 \end{bmatrix}$$

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

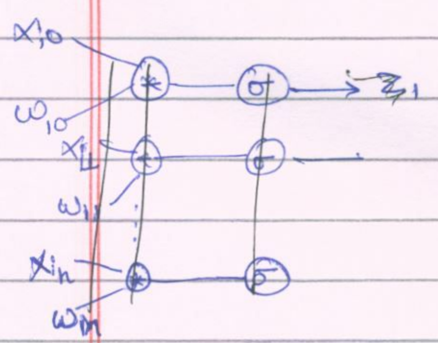
$$\underset{-15.88}{\downarrow} \quad \begin{bmatrix} 1 \\ 0.52 \\ 0.52 \\ 0.52 \end{bmatrix} = \begin{bmatrix} -15.88 \\ -8.262 \\ -8.262 \\ -8.262 \end{bmatrix}$$

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

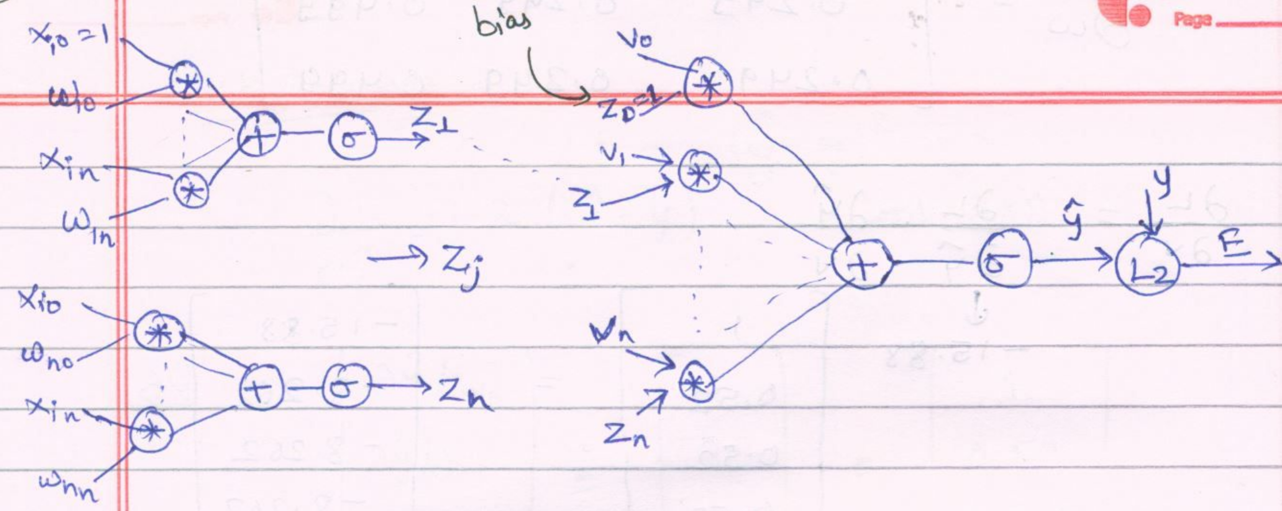
$$= -15.888 \begin{bmatrix} (0.02)[0.249 \quad 0.249 \quad 0.499] \\ (0.03)[0.249 \quad 0.249 \quad 0.499] \\ (0.04)[0.249 \quad 0.249 \quad 0.499] \end{bmatrix}$$

$$\boxed{\frac{\partial L}{\partial w} = \begin{bmatrix} -0.0792 & -0.0792 & -0.1584 \\ -0.1187 & -0.1187 & -0.2375 \\ -0.1584 & -0.1584 & -0.3168 \end{bmatrix}}$$ #

③



3



$$E = \frac{1}{2} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2$$

$$\hat{y} = \text{Sigmoid}(z^T \cdot v) = \text{Sigmoid}(v_0 + z_1 v_1 + \dots + z_n v_n)$$

$$z_j = \text{Sigmoid}(w_j^T x).$$

$$\left\{ \frac{\partial E}{\partial v} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial v} \right\} \# \qquad \frac{\partial E}{\partial \hat{y}} = \sum_{i=1}^{m} (\hat{y}_i - y_i)$$

$$\frac{\partial \hat{y}}{\partial v} = \frac{\partial}{\partial v} \text{Sigmoid}(z^T v) = \hat{y}^{(i)}(1-\hat{y}^{(i)}) \cdot z^{(i)}$$

$$\boxed{\frac{\partial E}{\partial v} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot \hat{y}^{(i)}(1-\hat{y}^{(i)}) \cdot z^{(i)}} \#$$