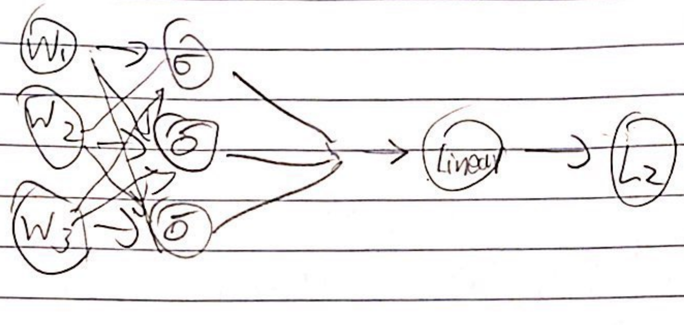
Subowen Yan A20430537

HW 2

1

a)




b)

$$X \cdot W^T = \begin{bmatrix} 1, & 1, 2 \\ 1, & 1, 3 \\ 1, & 2, 2 \end{bmatrix} \cdot \begin{bmatrix} 0.01 & 0.03 & 0.02 \\ 0.02 & 0.01 & 0.03 \\ 0.03 & 0.02 & 0.01 \end{bmatrix} = \begin{bmatrix} 0.09 & 0.08 & 0.07 \\ 0.12 & 0.1 & 0.08 \\ 0.11 & 0.09 & 0.1 \end{bmatrix}$$

$$z = \text{sigmoid}(W.X) = \begin{array}{c} \begin{bmatrix} 0.5224848 & 0.51998934 & 0.51749286 \\ 0.5299641 & 0.52497919 & 0.51998934 \\ 0.527472 & 0.5224848 & 0.52497919 \end{bmatrix} \end{array} \begin{array}{l} x_0 \\ x_1 \\ x_2 \end{array}$$

$\frac{1}{1+e^{-x}}$

$$L = z \cdot V^T = \begin{bmatrix} 1 & & \\ 1 & \text{sigmoid}(W.X) & \\ 1 & & \end{bmatrix} \cdot \begin{bmatrix} 0.01 \\ 0.02 \\ 0.03 \\ 0.04 \end{bmatrix} = \begin{bmatrix} 0.05672 \\ 0.05714823 \\ 0.05722316 \end{bmatrix}$$

$$L_2 = (8 - 0.05672)^2 + (11 - 0.057148)^2 + (10 - 0.05722)^2 = 281.70052$$

c) $\frac{\partial L}{\partial V_0} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V_0} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot L_0^{(i)} = (0.05672 - 8) * 1 + (0.05714823 - 11) \times 1$
$+ (0.05722316 - 10) \times 1 = -28.83$

$\frac{\partial L}{\partial V_1} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial V_1} = \sum_{i=1}^{M} (\hat{y}^{(i)} - y^{(i)}) \cdot L_1^{(i)} = (0.05672 - 8) * 0.5224848 + (0.05714823$
$- 11) * 0.5299641 + (0.05722316 - 11) * 0.527472 3 = -15.19$

$\frac{\partial L}{\partial V_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V_2} = \sum_{i=1}^{M} (\hat{y}^{(i)} - y^{(i)}) \cdot L_2^{(i)} = (0.05672 - 8) * 0.51998934 + (0.05714823$
$- 11) * 0.52497919 + (0.05722316 - 11) * 0.5224848 = -15.07$

$\frac{\partial L}{\partial V_3} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V_3} = \sum_{i=1}^{M} (\hat{y}^{(i)} - y^{(i)}) \cdot L_3^{(i)} = (0.05672 - 8) * 0.51749286) + (0.05714823$
$- 11) * 0.51998934 + (0.05722316 - 11) * 0.52492919 = -15.02$

$\frac{\partial L}{\partial w_{10}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_{10}} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot V_1 \cdot z_1^{(i)}(1-z_1^{(i)}) \cdot 1 = (0.05672-8) * 0.02 * 0.52248482 * (1-0.52248482) + (0.057148-11) * 0.02 * 0.52996405(1-0.52996405) + (0.057723 - 10) * (1-0.5274723) = -0.1437$

$\frac{\partial L}{\partial w_{11}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_{11}} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot V_1 \cdot z_1^{(i)}(1-z_1^{(i)}) \cdot x_1^{(i)} = (0.05672-8) \times 0.02 \times 0.52248482 * (1-0.5248482) * 1 + (0.057148-11) * 0.02 \times 0.52996405(1-0.52996405) * 1 + (0.057723 - 10) * (1-0.5274723) \times 2 = -0.1933$

$\frac{\partial L}{\partial w_{12}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_{12}} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot V_1 \cdot z_1^{(i)}(1-z_1^{(i)}) \cdot x_2^{(i)} = (0.05672-8) \times 0.02 * 0.52248482 * (1-0.5248482) * 2 + (0.057148-11) * 0.02 * 0.52996405(1-0.52996405) * 3 + (0.057723 - 10) * (1-0.527472\,3) \times 2 = -0.34$

$\frac{\partial L}{\partial w_{20}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_{20}} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot V_2 \cdot z_2^{(i)}(1-z_2^{(i)}) \cdot 1 = (0.05672-8) \times 0.03 * (0.51998934) * (1-0.51998934) + (0.057148-11) * 0.52497919 * (1-0.52497919) * 0.03 + (0.057223-10) * 0.52248482 * (1-0.52248482) \times 0.03 = -0.21577$

$\frac{\partial L}{\partial w_{21}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_{21}} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot V_2 \, z_2^{(i)} (1-z_2^{(i)}) x_1^{(i)} = -0.29$

$\frac{\partial L}{\partial w_{22}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_{22}} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot V_2 \, z_2^{(i)} (1-z_2^{(i)}) x_2^{(i)} = -0.513$

$\frac{\partial L}{\partial w_{30}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_{30}} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot V_3 \cdot z_3^{(i)}(1-z_3^{(i)}) \cdot 1 = (0.05672-8) \times 0.04 \times 0.51749286 * (1-0.51749286) + (0.057148-11) * 0.51998934 \times 0.04(1-0.51998934) + (0.057223-10) * 0.52497919 * 0.04 * (1-0.52497919) = -0.28777$

$\frac{\partial L}{\partial w_{31}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_{31}} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot V_3 \cdot z_3^{(i)}(1-z_3^{(i)}) \cdot x_1^{(i)} = -0.387$

$\frac{\partial L}{\partial w_{32}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_{32}} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot V_3 \cdot z_3^{(i)}(1-z_3^{(i)}) x_2^{(i)} = -0.68479$

d) $\frac{\partial L}{\partial V} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot z^{(i)} = (-28.83, -15.19, -15.07, -15.02)$

$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1} = \sum_{i=1}^{m} (\hat{y} - y^{(i)}) \cdot V_1 \cdot z_1^{(i)}(1-z_1^{(i)}) x^{(i)} = (-0.1437, -0.1933, -0.34)$

$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_2} = \sum_{i=1}^{m} (\hat{y} - y^{(i)}) \cdot V_2 \cdot z_2^{(i)}(1-z_2^{(i)}) \cdot x^{(i)} = (-0.21577, -0.29, -0.513)$

$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_3} = \sum_{i=1}^{m} (\hat{y} - y^{(i)}) \cdot V_3 \, z_3^{(i)}(1-z_3^{(i)}) x^{(i)} = (-0.2877, -0.387, -0.68479)$

It's same with part (c)

2

a) $\nabla f(x,y) = \left(\frac{df}{dx}, \frac{df}{dy}\right) = [(2x+3y)\cdot 2\cdot 2,\ 2(2x+3y)\cdot 3] = (8x+12y, 12x+18y)$

b) $DF = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x & 2 \\ 3 & 8y \end{bmatrix} \Big|_{x=1, y=2} = \begin{bmatrix} 2 & 2 \\ 3 & 16 \end{bmatrix}$

c) $d(F \circ G)(2) = d\begin{bmatrix} x^2+2x^2 \\ 3x+4x^4 \end{bmatrix}(2) = \begin{bmatrix} 6x \\ 3+16x^3 \end{bmatrix} \Big|_{x=2} = \begin{bmatrix} 12 \\ 131 \end{bmatrix}$

$d(F) = \begin{bmatrix} 2x & 2 \\ 3 & 8y \end{bmatrix}$ $\quad d(G) = \begin{bmatrix} 1 \\ 2x \end{bmatrix}$ $\quad d(F)\cdot d(G) = \begin{bmatrix} 2x+4x \\ 3+16x\cdot x^2 \end{bmatrix} = \begin{bmatrix} 6x \\ 3+16x^3 \end{bmatrix} \Big|_{x=2} \begin{bmatrix} 12 \\ 131 \end{bmatrix}$

d) $\frac{\partial L}{\partial V_1} = \frac{\partial L}{\partial \hat{y}^{(1)}} \cdot \frac{\partial \hat{y}_1}{\partial V_1} = [\hat{y}^{(1)} - y][z_0^{(1)}\ z_1^{(1)}\ z_2^{(1)}\ z_3^{(1)}]^T$

$= [-7.943\ \ -4.146\ \ -4.13\ \ -4.11]^T$

$\frac{\partial L}{\partial W_1^{(1)}} = \frac{\partial L}{\partial \hat{y}^{(1)}} \cdot \frac{\partial \hat{y}^{(1)}}{\partial z^{(1)}} \cdot \frac{\partial z^{(1)}}{\partial W^{(1)}} = [\hat{y}^{(1)} - y^{(1)}] \begin{bmatrix} V_1 z_1^{(1)}(1-z_1^{(1)}) x_0 & V_1 z_1^{(1)}(1-z_1^{(1)}) x_1 & V_1 z_1^{(1)}(1-z_1^{(1)}) x_1 \\ V_2 z_2^{(1)}(1-z_2^{(1)}) x_1 & V_2 z_2^{(1)}(1-z_2^{(1)}) x_2 & V_2 z_2^{(1)}(1-z_2^{(1)}) x_2 \\ V_3 z_3^{(1)}(1-z_3^{(1)}) x_2 & V_3 z_3^{(1)}(1-z_3^{(1)}) x_3 & V_3 z_3^{(1)}(1-z_3^{(1)}) x_3 \end{bmatrix}$

$= \begin{bmatrix} -0.03964 & -0.03964 & -0.07928 \\ -0.0545 & -0.0545 & -0.1636 \\ -0.04956 & -0.09913 & -0.991 \end{bmatrix}$

$\frac{\partial L}{\partial V_2^{(1)}} = \frac{\partial L}{\partial \hat{y}^{(2)}} \cdot \frac{\partial \hat{y}_2}{\partial V_i} = [\hat{y}^{(2)} - y][z_0^{(2)}\ z_1^{(2)}\ z_2^{(2)}\ z_3^{(2)}] = [-10.94\ \ -5.79\ \ -5.74\ \ -5.69]$

$\frac{\partial L}{\partial V_3} = \frac{\partial L}{\partial \hat{y}_3} \cdot \frac{\partial \hat{y}_3}{\partial V_3} = [\hat{y}^{(3)} - y][z_0^{(3)}\ z_1^{(3)}\ z_2^{(3)}\ z_3^{(3)}] = [-9.943\ \ -5.245\ \ -5.195\ \ -5.22]$

$\frac{\partial L}{\partial V} = \sum \frac{\partial L}{\partial V_i} = [-28.83,\ -15.19,\ -15.07,\ -15.02]$

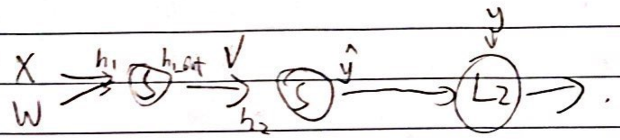
$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial y_2} \cdot \frac{\partial y_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial W_2} = [\hat{y}^{(i)} - y^{(i)}] \begin{bmatrix} V_1 z_1(1-z_1) x_0 & V_1 (1-z_1) z_1 x_1 & V_1(1-z_1) z_1 x_1 \\ V_2 z_2(1-z_2) x_1 & V_2 (1-z_2) z_2 x_2 & V_2(1-z_2) z_2 x_2 \\ V_3 (1-z_3) x_2 z_3 & V_3 (1-z_3) z_3 x_3 & V_3(1-z_3) z_3 x_3 \end{bmatrix}$

$$\begin{bmatrix} -0.059479 & -0.059479 & -0.118959 \\ -0.081867 & -0.081867 & -0.2455997 \\ -0.07442 & -0.14884 & -0.14884 \end{bmatrix}$$

$$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial y_3} \cdot \frac{\partial y_3}{\partial z_3} \cdot \frac{\partial z}{\partial w_3} = \begin{bmatrix} -0.07934 & -0.07934 & -0.15867 \\ -0.10925 & -0.10925 & -0.32776 \\ -0.09918 & -0.19836 & -0.19836 \end{bmatrix}$$

$$\sum \frac{\partial L}{\partial w_i} = \begin{bmatrix} -0.1437 & -0.1933 & -0.34 \\ -0.21577 & -0.29 & -0.513 \\ -0.2877 & -0.387 & -0.68479 \end{bmatrix}$$

3




$$\frac{\partial L}{\partial V} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_2} \frac{\partial h_2}{\partial V} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot \hat{y}^{(i)} (1 - \hat{y}^{(i)}) h_{1,out}^{(i)}$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_2} \cdot \frac{\partial h_2}{\partial h_{1,out}} \frac{\partial h_{1,out}}{\partial h_1} \frac{\partial h_1}{\partial w} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot \hat{y}^{(i)} (1 - \hat{y}^{(i)}) \cdot V \cdot h_{1,out}^{(i)}$$

$$\cdot (1 - h_{1,out}^{(i)}) x^{(i)}$$