Student Name: Mohammad Ahmad Khattab Mousa

Student Number: 2002639

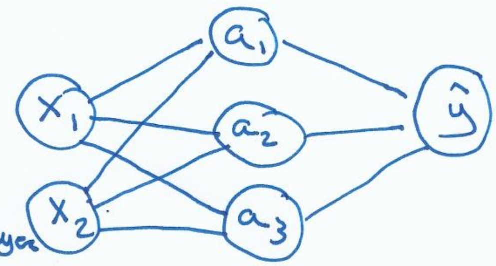
Solution to Assignment 1

$\boxed{1}$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \quad y = 32$$

$$W^{[1]} = \begin{pmatrix} 2 & 1 & 3 \\ 2 & -1 & 1 \end{pmatrix}, \quad b^{[1]} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

$$W^{[2]} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \quad b^{[2]} = (1)$$



1a)

assume input to a neuron layer is raw_i & output of a layer is act_i

$$\therefore raw^{[1]} = W^{[1]^T} X + b^{[1]}$$

$$= \begin{pmatrix} 2 & 2 \\ 1 & -1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix}$$

if unity activation function $\Rightarrow act^{[1]} = a = raw^{[1]}$

$$= \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix}$$

$$\therefore raw^{[2]} = W^{[2]^T} act^{[1]} + b^{[2]}$$

$$= (3 \;\; 1 \;\; 2) \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix} + 1 = 36$$

$$\boxed{\therefore \hat{y} = 36}$$

1b) ReLU activation function

$\therefore raw^{[1]} = \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix}$

$\therefore act^{[1]} = \begin{pmatrix} 9 \\ \boxed{0} \\ 5 \end{pmatrix}$

$\therefore raw^{[2]} = \begin{pmatrix} 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} 9 \\ 0 \\ 5 \end{pmatrix} + 1 = 38$

$\therefore \hat{y} = ReLU(38) = 38 \quad \therefore \boxed{\hat{y} = 38}$

1c) network with identity activation function

$J = (\hat{y} - y)^2 \Longrightarrow \dfrac{\partial J}{\partial \hat{y}} = 2(\hat{y} - y)$ $\quad \hat{y} = 36$

$= 8 \quad y = 32$

$\dfrac{\partial J}{\partial \hat{y}} = 8$

$\hat{y} = \cancel{re}\ raw^{[2]} = \sum_{i=1}^{3} w_{i1}^{[2]} act_i^{[1]} + b_1^{[2]}$

$\therefore \dfrac{\partial \hat{y}}{\partial raw^{[2]}} = 1$

$\dfrac{\partial raw^{[2]}}{\partial b_1^{[2]}} = 1$

$\therefore \dfrac{\partial J}{\partial b_1^{[2]}} = \dfrac{\partial J}{\partial \hat{y}} * \dfrac{\partial \hat{y}}{\partial raw^{[2]}} * \dfrac{\partial raw^{[2]}}{\partial b_1^{[2]}}$

$= 8 * 1 * 1$

$\therefore \boxed{\dfrac{\partial J}{\partial b_1^{[2]}} = 8}$

2). $raw_i^{[1]} = \sum_{j=1}^{2} w_{ji} x_j + b_i^{[1]}$

$\cancel{\partial raw_1^{[}}$

$raw_1^{[2]} = \sum_{i=1}^{3} w_{i1}^{[2]} act_i^{[1]} + b_1^{[2]}$

$\therefore \frac{\partial J}{\partial w_{21}^{[2]}} = \frac{\partial J}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial raw_1^{[2]}} * \frac{\partial raw_1^{[2]}}{\partial w_{21}^{[2]}}$

$= 8 * 1 * a_2 = 8 * -2$

$\boxed{\therefore \frac{\partial J}{\partial w_{21}^{[2]}} = -16}$

3) $\frac{\partial J}{\partial b_2^{[1]}} = \frac{\partial J}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial raw_1^{[2]}} * \frac{\partial raw_1^{[2]}}{\partial a_2} * \frac{\partial a_2}{\partial b_2^{[1]}}$

$= 8 * 1 * \underset{\hookrightarrow w_{21}^{[2]}}{1} * 1$

$\therefore \boxed{\frac{\partial J}{\partial b_2^{[1]}} = 8}$

4) $\frac{\partial J}{\partial w_{13}^{[1]}} = \frac{\partial J}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial raw_1^{[2]}} * \frac{\partial raw_1^{[2]}}{\partial a_3} * \frac{\partial a_3}{\partial raw_3^{[1]}} * \frac{\partial raw_3^{[1]}}{\partial w_{13}^{[1]}}$

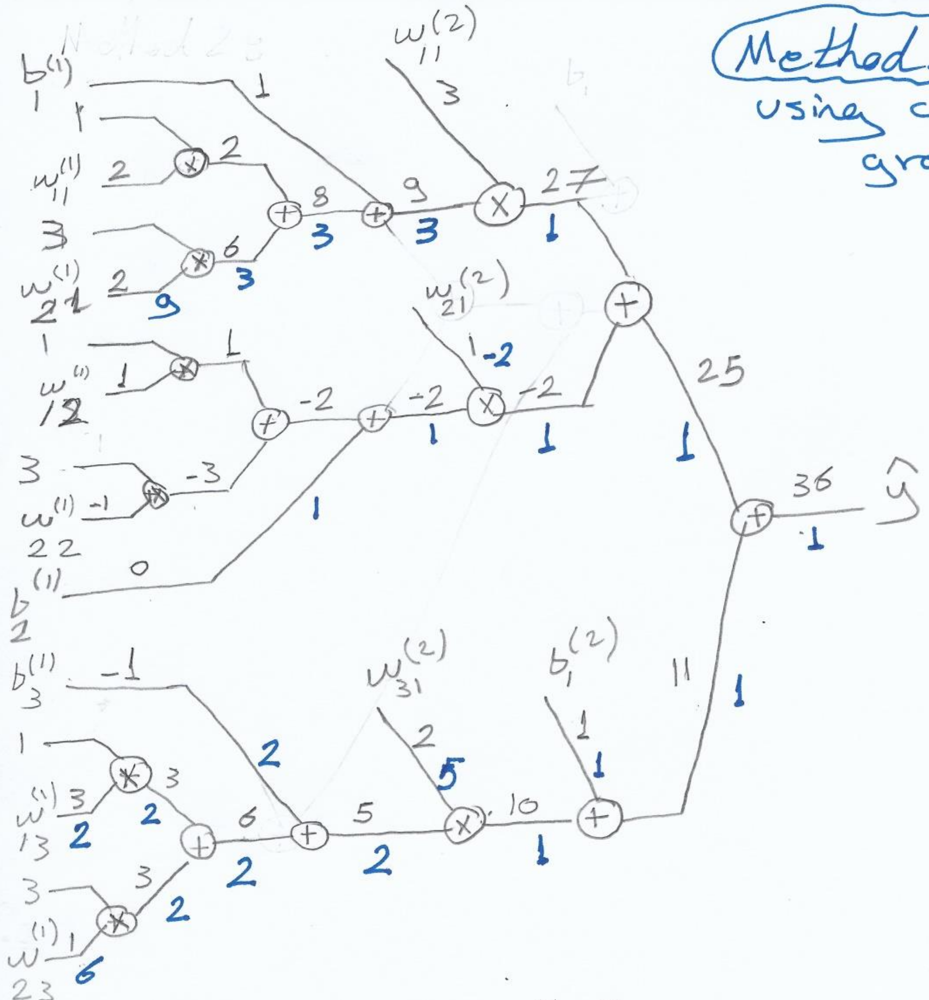
$= 8 * 1 * w_{31}^{[2]} * 1 * x_1$

$= 8 * 2 * 1 = 16$

$\therefore \boxed{\frac{\partial J}{\partial w_{13}^{[1]}} = 16}$

3

Method 2
using computational graph



$$\frac{\partial J}{\partial \hat{y}} = 2(\hat{y} - y) = 2 * 4 = 8$$

$$\frac{\partial J}{\partial b^{(2)}_1} = \frac{\partial J}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial b^{[2]}_1} = 8 * 1 = 8$$

$$\frac{\partial J}{\partial w^{[2]}_{21}} = \frac{\partial J}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial w^{[2]}_{21}} = 8 * (-2) = -16$$

$$\frac{\partial J}{\partial b^{[1]}_2} = \frac{\partial J}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial b^{[1]}_2} = 8 * 1 = 8$$

$$\frac{\partial J}{\partial w^{[1]}_{13}} = \frac{\partial J}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial w^{[1]}_{13}} = 8 * 2 = 16$$

1d)

$$\therefore\ b_2^{[1]} = b_2^{[1]} - \eta \frac{\partial J}{\partial b_2^{[1]}}$$

$$= 0 - 2 * 8 \Rightarrow \boxed{b_2^{[1]} = -16}$$

$$w_{13}^{[1]} = w_{13}^{[1]} - \eta \frac{\partial J}{\partial w_{13}^{[1]}}$$

$$= 3 - 2 * 16 = -29 \Rightarrow \boxed{w_{13}^{[1]} = -29}$$

1e) Test dataset is used to assess the performance (generalization) of a fully specified model.

Yes, test dataset could be used to assess the out_of_sample error as these are unseen samples & if the model fits these samples, it's assumed that minimal overfitting has taken place during training (or training noise has minimal impact over the model)

$\boxed{2}$ $f = \sin g_1 + g_2^2$

$\therefore \frac{\partial f}{\partial g_1} = \cos g_1 = \cos(x_1 e^{x_2})$

$\frac{\partial f}{\partial g_2} = 2 g_2 = 2(x_1 + x_2^2)$

$\frac{\partial g_1}{\partial x_1} = e^{x_2}$; $\frac{\partial g_1}{\partial x_2} = x_1 e^{x_2}$

$\frac{\partial g_2}{\partial x_1} = 1$, $\frac{\partial g_2}{\partial x_2} = 2x_2$

$\therefore \frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial g_1} * \frac{\partial g_1}{\partial x_1} + \frac{\partial f}{\partial g_2} * \frac{\partial g_2}{\partial x_1}$

$= \cos(x_1 e^{x_2}) * e^{x_2} + 2(x_1 + x_2^2)$

$\frac{\partial f}{\partial x_2} = \frac{\partial f}{\partial g_1} * \frac{\partial g_1}{\partial x_2} + \frac{\partial f}{\partial g_2} * \frac{\partial g_2}{\partial x_2}$

$= \cos(x_1 e^{x_2}) * x_1 e^{x_2} + 2(x_1 + x_2^2) * (2x_2)$

$= x_1 e^{x_2} \cos(x_1 e^{x_2}) + 4x_2^3 + 4x_1 x_2$

6M

$\boxed{3}$

1) $f(z) = \frac{1}{1+e^{-z}}$

$\therefore \frac{\partial f}{\partial z} = \frac{-1}{(1+e^{-z})^2} * -1 * e^{-z} = \frac{e^{-z}}{(1+e^{-z})^2}$

2) let $\underset{1\times1}{g} = \underset{1\times D}{w^T} \underset{D\times1}{x}$

$\frac{\partial g}{\partial w_i} = x_i \Rightarrow \frac{\partial g}{\partial w} = x$

$\frac{\partial f}{\partial w_i} = \frac{\partial f}{\partial g} * \frac{\partial g}{\partial w_i} = \frac{e^{-w^T x}}{(1+e^{-w^T x})^2} * x_i$

3) $\frac{\partial J}{\partial w} = \begin{bmatrix} \frac{\partial J}{\partial w_1} \\ \vdots \\ \frac{\partial J}{\partial w_D} \end{bmatrix}$

$\frac{d}{dx}(|x|) = \frac{x}{|x|}$

$J(w) = \frac{1}{2} \sum_{i=1}^{m} \left[\left| \sum_{j=1}^{D} w_j x_j^{(i)} - y^{(i)} \right| \right]$

$\therefore \frac{\partial J(w)}{\partial w_k} = \frac{1}{2} \sum_{i=1}^{m} \frac{w_k x_k^{(i)} - y^{(i)}}{|w_k x_k^{(i)} - y^{(i)}|}$

$\therefore \frac{\partial J(w)}{\partial w} = \frac{1}{2} \begin{bmatrix} \sum_{i=1}^{m} \frac{w_1 x_1^{(i)} - y^{(i)}}{|w_1 x_1^{(i)} - y^{(i)}|} \\ \vdots \\ \sum_{i=1}^{m} \frac{w_D x_D^{(i)} - y^{(i)}}{|w_D x_D^{(i)} - y^{(i)}|} \end{bmatrix}$

7

4) $J(w) = \frac{1}{2}\left[\sum_{i=1}^{m} (w^T x^{(i)} - y^{(i)})^2\right] + \lambda \sum_{j=1}^{D} w_j^2$

$\therefore \frac{\partial J}{\partial w_k} = \frac{1}{2}\left[\sum_{i=1}^{m} 2(w_k x_k^{(i)} - y^{(i)})(x_k^{(i)})\right] + 2\lambda w_k$

$\therefore \frac{\partial J}{\partial w_k} = \sum_{i=1}^{m} w_k x_k^{(i)^2} - \sum_{i=1}^{m} y^{(i)} x_k^{(i)} + 2\lambda w_k$

$$\therefore \frac{\partial J}{\partial w} = \begin{bmatrix} \frac{\partial J}{\partial w_1} \\ \vdots \\ \frac{\partial J}{\partial w_D} \end{bmatrix}$$

5) let $u = w^T x^{(i)}$

$\therefore J(u) = \sum_{i=1}^{m} \left[y^{(i)} \log \frac{1}{1+e^{-u}} + (1-y^{(i)}) \log\left(1 - \frac{1}{1+e^{-u}}\right)\right]$

$\therefore$ let $v = \frac{1}{1+e^{-u}}$

$\therefore J(v) = \sum_{i=1}^{m} \left[y^{(i)} \log v + (1-y^{(i)}) \log(1-v)\right]$

$\therefore \frac{\partial J}{\partial v} = \sum_{i=1}^{m} \left[\frac{y^{(i)}}{v} + (1-y^{(i)}) * \frac{-1}{v}\right]$

$= \sum_{i=1}^{m} \left[\frac{2y^{(i)}+1}{v} \right]$

$\frac{\partial v}{\partial u} = \frac{-1}{(1+e^{-u})^2} * -e^{-u} = \frac{e^{-u}}{(1+e^{-u})^2}$

$\frac{\partial u}{\partial w_k} = x_k^{(i)}$

$\therefore \frac{\partial J}{\partial w_k} = \frac{\partial J}{\partial v} * \frac{\partial v}{\partial u} * \frac{\partial u}{\partial w_k}$

$= \sum_{i=1}^{m} (2y^{(i)}+1)(1+e^{-w^T x^{(i)}}) \frac{e^{-w^T x^{(i)}}}{(1+e^{-w^T x^{(i)}})^2} * x_k^{(i)}$

$= \sum_{i=1}^{m} x_k^{(i)} (2y^{(i)}+1) \frac{e^{-w^T x^{(i)}}}{1+e^{-w^T x^{(i)}}}$

8

$$\therefore \frac{\partial J}{\partial w} = \begin{bmatrix} \frac{\partial J}{\partial w_1} \\ \vdots \\ \frac{\partial J}{\partial w_D} \end{bmatrix}$$

b) $f = \tanh(w^T x)$

$$\therefore \frac{\partial f}{\partial w_k} = \text{sech}^2(w^T x)\, x_k$$

$$\therefore \nabla_w f = \text{sech}^2(w^T x) \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix}$$

$$\nabla_w f = \text{sech}^2(w^T x)\, x$$

6