Part I: Derivative of Tanh

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\sinh(x) = \frac{1}{2}(e^x - e^{-x})$$

$$\cosh(x) = \frac{1}{2}(e^x + e^{-x})$$

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$f'(x) = \frac{d}{dx}(\tanh x) = \operatorname{sech}^2 x$$

$$f'(x) = 1 - f(x)^2$$

$$\operatorname{sech}^2 x = 1 - \tanh^2 x$$

$$\Downarrow \quad \operatorname{sech}^2 x + \tanh^2 x = 1$$

Hence proved $\boxed{f'(x) = 1 - f(x)^2}$

→ Part II Forward - backward Pass

1. $h1 = 0.7 \times -0.3 + 0.5 \times 0.15 + 1 \times 0.9 = 0.765$

$\ldots \times 0.8 + 1 \times -0.14$

$h_2 = 0.5 \times 0.2 + 0.7 \times \ldots$

$= 0.52$

$\Rightarrow \hat{y} = 0.7 \times 0.765 + 0.52 \times 0.25 - 0.1$

$= 0.5355 + 0.13 - 0.1$

$= 0.5655$

$\hat{y}$ = predicted value $= 0.5655$

2) $MSE = \frac{1}{2}(y - \hat{y})^2$

$= \frac{1}{2}(0.5 - 0.5655)^2$

$$\boxed{MSE = 0.002145}$$

3) Find the gradient using back-propogation

$$\nabla J_i = \begin{bmatrix} \frac{\partial J_i}{\partial w_1} \\ \frac{\partial J_i}{\partial w_2} \\ \vdots \\ \frac{\partial J_i}{\partial w_9} \end{bmatrix}$$

$$\rightarrow \quad h_1 = i_1 w_1 + i_2 w_3 + w_5$$

$$h_2 = i_1 w_2 + i_2 w_4 + w_6$$

$$\hat{y} = (i_1 w_1 + i_2 w_3 + w_5) \times w_7 + (i_1 w_2 + i_2 w_4 + w_6) \times w_8 + w_9$$

$$\frac{\partial Error}{\partial w_9} = \frac{\partial Error}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w_9}$$

$$= (y - \hat{y})(-1) \times 1$$

$$\boxed{\frac{\partial Error}{\partial w_9} = -(y - \hat{y})}$$

$$\boxed{\frac{\partial Error}{\partial w_7} = (y - \hat{y})(-1) \times h_1}$$

$$\frac{\partial Error}{\partial w_8} = \frac{\partial Error}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w_8}$$

$$\boxed{\frac{\partial Error}{\partial w_8} = (y - \hat{y})(-1) \times h_2}$$

$$\partial Error = \partial Error \times \partial \hat{y} \times \frac{\partial h_2}{\partial w_1}$$

$$\frac{\partial w_6}{} \quad \frac{}{\partial \hat{y}} \quad \partial h_2 \quad \partial w_6$$

$$= (y-\hat{y})(-1) \quad = (y-\hat{y})(-1) \times w_8$$

$$\times w_8 \times 1$$

$$\boxed{\frac{\partial Error}{\partial w_6} = (y-\hat{y}) \times (-1) \times w_8}$$

$$\frac{\partial Error}{\partial w_5} = \frac{\partial Error}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial h_1} \times \frac{\partial h_1}{\partial w_5}$$

$$= (y-\hat{y}) \times (-1) \times w_7$$

$$\boxed{\frac{\partial Error}{\partial w_5} = (y-\hat{y})(-1) \times w_7}$$

$$\frac{\partial Error}{\partial w_4} = \frac{\partial Error}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial h_2} \times \frac{\partial h_2}{\partial w_4}$$

$$\frac{\partial Error}{\partial w_4} = (y-\hat{y}) \times (-1) \times w_8\, i_2$$

$$\boxed{\frac{\partial Error}{\partial w_4} = (y-\hat{y}) \times (-1) \times w_8 \times i_2}$$

$$\frac{\partial Error}{} = \frac{\partial Error}{\partial \hat{y}} \times \frac{\partial \hat{y}}{} \times \frac{\partial h_1}{\partial w_3}$$

$$\partial w_3 \qquad \hat{y} \qquad \partial h_1$$

$$\boxed{\frac{\partial Error}{\partial w_3} = (y-\hat{y}) \times (-1) \times w_7 \times i_2}$$

$$\frac{\partial Error}{\partial w_2} = \frac{\partial Error}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial h_2} \times \frac{\partial h_2}{\partial w_2}$$

$$\boxed{\frac{\partial Error}{\partial w_2} = (y-\hat{y}) \times -1 \times w_8 \times i_1}$$

$$\boxed{\frac{\partial Error}{\partial w_1} = (y-\hat{y}) \times -1 \times w_7 \times i_1}$$

Gradients computed using back prop

$$\Rightarrow \nabla J_i = \begin{bmatrix} (y-\hat{y}) \times (-w_7) \times i_1 \\ (y-\hat{y}) \times (-w_8) \times i_1 \\ (y-\hat{y}) \times (-w_7) \times i_2 \\ (y-\hat{y}) \times (-w_8) \times i_2 \\ (y-\hat{y}) \times (-w_7) \\ (y-\hat{y}) \times (-w_8) \\ (y-\hat{y}) \times (-h_1) \\ (y-\hat{y}) \times (-h_2) \end{bmatrix}$$

$$\begin{bmatrix} (y-\hat{y} \\ -(y-\hat{y}) \end{bmatrix}$$

$y-\hat{y} = 0.5 - 0.5655 = -0.0655$

$h_1, h_2 = 0.765, 0.52$

$i_1, i_2 = 0.7, 0.5$

$w_7, w_8 = 0.7, 0.25$

$$\nabla J_i = 0.0655 \begin{bmatrix} 0.7 \times 0.7 \\ 0.25 \times 0.7 \\ 0.7 \times 0.5 \\ 0.25 \times 0.5 \\ 0.7 \\ 0.25 \\ 0.765 \\ 0.52 \\ 1 \end{bmatrix}$$

$$\Rightarrow 0.0655 \begin{bmatrix} 0.49 \\ 0.175 \\ 0.35 \\ 0.125 \\ 0.7 \\ 0.25 \\ 0.765 \\ 0.52 \\ 1 \end{bmatrix} = \begin{array}{c} \begin{bmatrix} 0.032095 \\ 0.0114 \\ 0.022 \\ 0.008 \\ 0.045 \\ 0.016 \\ 0.0501 \\ 0.0340 \\ 0.0655 \end{bmatrix} \end{array} \begin{array}{c} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ w_8 \\ w_9 \end{array}$$

4. Update the weights and the bias

$\ldots = -0.1 - 0.03 \times 0.0655$

$w_9 = w_9 - \alpha \times 0.0655 \rightarrow$

$w_8 = w_8 - \alpha \times 0.0340 \rightarrow 0.25 - 0.03 \times 0.0340$

$w_7 = w_7 - \alpha \times 0.0501 \rightarrow 0.7 - 0.03 \times 0.0501$

$w_6 = w_6 - \alpha \times 0.016 \rightarrow -0.14 - 0.03 \times 0.016$

$w_5 = w_5 - \alpha \times 0.045 \rightarrow 0.9 - 0.03 \times 0.045$

$w_4 = w_4 - \alpha \times 0.008 \rightarrow 0.2 - 0.03 \times 0.008$

$w_3 = w_3 - \alpha \times 0.022 \rightarrow 0.15 - 0.03 \times 0.022$

$w_2 = w_2 - \alpha \times 0.0114 \rightarrow 0.8 - 0.03 \times 0.0114$

$w_1 = w_1 - \alpha \times 0.032095$

$\Downarrow$

$-0.3 - 0.03 \times 0.032095$

$$\boxed{\begin{array}{ll} w_9 = -0.101 & w_5 = 0.89865 \\ w_8 = 0.24898 & w_4 = 0.19976 \\ w_7 = 0.698497 & w_3 = 0.14934 \\ w_6 = -0.14048 & w_2 = 0.799658 \\ w_1 = -0.30096285 & \end{array}}$$

6. $h1 = w1 \times 0.7 + w3 \times 0.5 + w5 \times 1$

$h2 = w2 \times 0.7 + w4 \times 0.5 + w6 \times 1$

$h1 = 0.7626$

$h2 = 0.519$

$output = h1 \times w7 + h2 \times w8 + w9$

$= 0.56089 \Rightarrow \hat{y} = 0.5608$

$MSE = \frac{1}{2}(0.56089 - 0.5)^2$

$= 0.001854$

$\boxed{\text{Task-2 Q.5 is written after part 3}}$

Part III

i) $W1 = 32$ $K = 10$ $P = 0$

$H1 = 28$ $F = 5$

$D1 = 3$ $S = 1$

$W2 = (32 - 5 + 2 \times 0)/1 + 1 = 28$

$H2 = (28 - 5 + 2 \times 0)/1 + 1 = 24$

$D2 = 10$

$28 \times 24 \times 10$

output size after the first layer =

2) No. of parameters in the layer

$\Downarrow$

$5 \times 5 \times 3 + 1 = 76$ params (+1 for bias

$\Downarrow$

$76 \times 10 = 760$ parameters

No. of parameters = 760

3) output size $\Rightarrow$ pad = 1

$$W2 = \left[(32 - 5 + 2 \times 1) / 1 + 1 \right]$$

$$H2 = \left[(28 - 5 + 2 \times 1) / 1 + 1 \right]$$

$W2, H2 = 30, 26$

$$\boxed{W2, H2, D2 = 30 \times 26 \times 10}$$

4) grayscale = 1

$\Downarrow$

$5 \times 5 \times 1 + 1 = 26$ params

$\boxed{= 26 \times 10 = 260 \text{ parameters}}$

5) Since it is multiclass classification problem, the most suitable activation function for the output layer would be $\boxed{\text{softmax function}}$

6) Softmax is invariant to constant shifts

$\text{total} = e^{x_1} + e^{x_2} + e^{x_3}$

$\Downarrow$

$x_1\ x_2\ x_3$ are the inputs

probabilities are calculated using $\Rightarrow$ $\dfrac{e^{x_1}}{\text{total}}, \dfrac{e^{x_2}}{\text{total}}, \dfrac{e^{x_3}}{\text{total}}$.

On adding a constant, we get

$\text{new} = e^{(x_1+c)} + e^{(x_2+c)} + e^{(x_3+c)}$

$\Downarrow$

$\Rightarrow \dfrac{e^{(x_1+c_1)}}{\quad}, \dfrac{e^{(x_2+c_2)}}{\text{new}}, \dfrac{e^{(x_3+c_3)}}{\text{new}}$

new

$$\Rightarrow \frac{e^{x_1} \times e^{c_1}}{e^{c_1}(e^{x_1}+e^{x_2}+e^{x_3})}, \frac{e^{x_2} \times e^{c_2}}{e^{c_2}(e^{x_1}+e^{x_2}+e^{x_3})}, \frac{e^{x_3} \times e^{c_3}}{e^{c_3}(e^{x_1}+e^{x_2}+e^{x_3})}$$

$$\Rightarrow \boxed{\frac{e^{x_1}}{total}, \frac{e^{x_2}}{total}, \frac{e^{x_3}}{total}}$$

$\Downarrow$

From the above, we can see that the softmax activation function is invariant to constant shifts in the input values

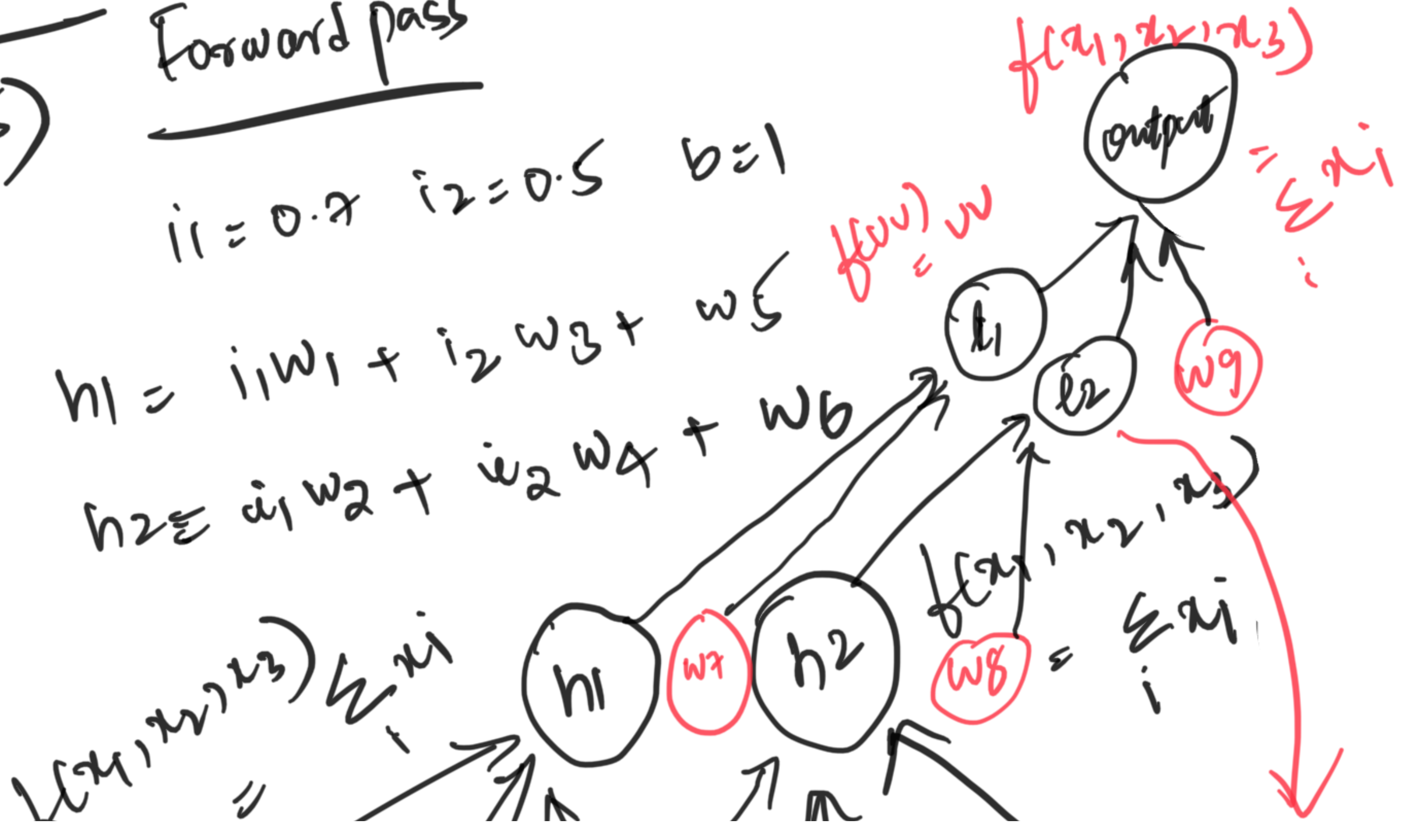
Task 2

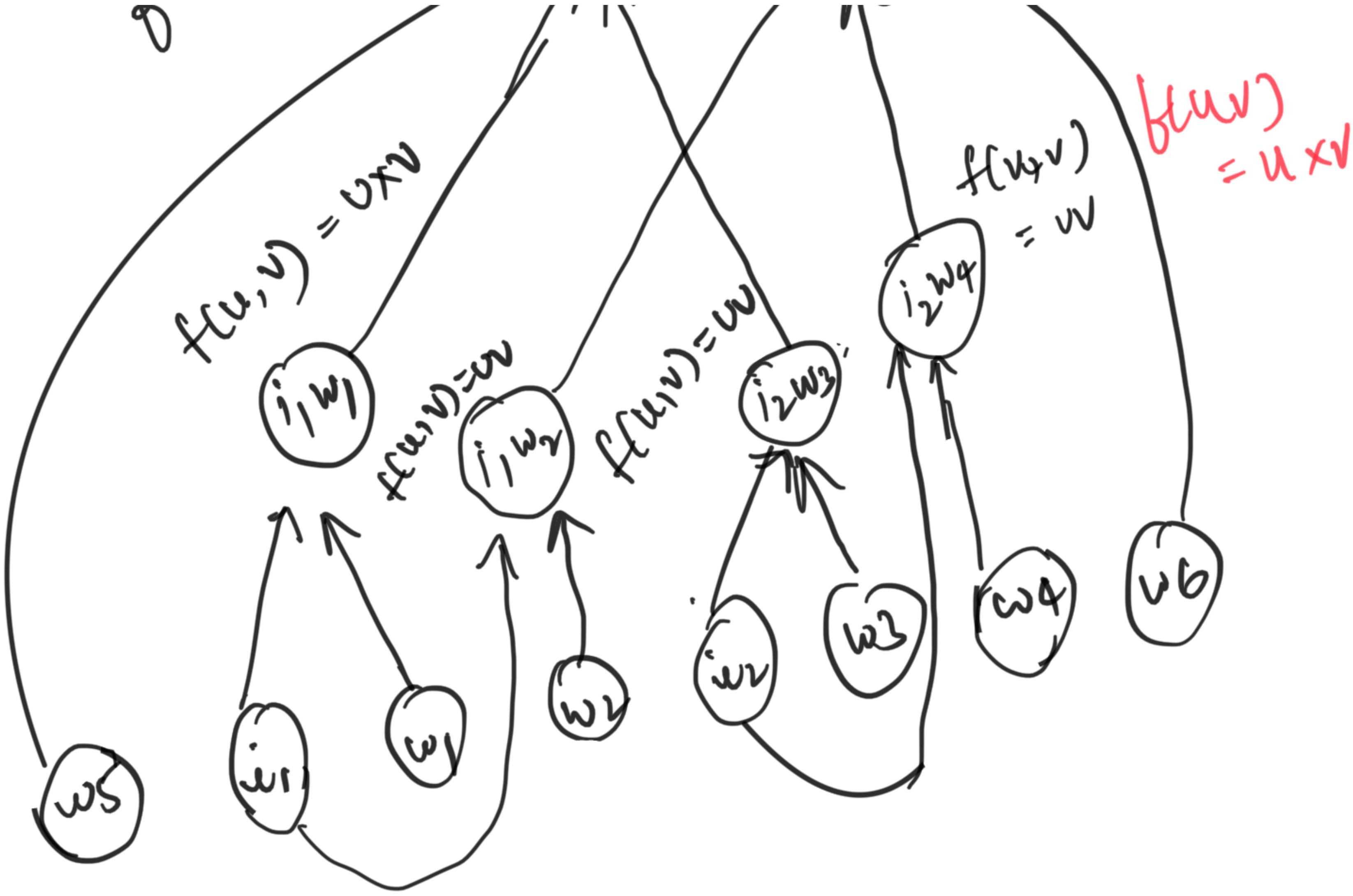
Q5) Forward pass

$i_1 = 0.7 \quad i_2 = 0.5 \quad b = 1$

$h1 = i_1 w_1 + i_2 w_3 + w_5$

$h2 = i_1 w_2 + i_2 w_4 + w_6$





⇒ <u>Backward passes:-</u>

$$\text{error} = \frac{1}{2}(y - \hat{y})^2$$