Q1 a) Input $x = (0.5 \ -1 \ 2)^T$, $Y = (1 \ 1 \ 0)^T$

$$W_1 = \begin{pmatrix} 0.2 & -0.3 & 0.4 \\ 0.1 & 0.2 & -0.5 \\ -0.3 & 0.1 & 0.2 \\ 0.4 & -0.1 & 0.3 \end{pmatrix} \quad W_2 = \begin{pmatrix} 0.1 & 0.2 & -0.1 & 0.1 \\ -0.2 & 0.3 & 0.1 & -0.3 \\ 0.2 & -0.1 & 0.3 & 0.2 \end{pmatrix}$$

$$Z_1 = \begin{pmatrix} 0.2 & -0.3 & 0.4 \\ 0.1 & 0.2 & -0.5 \\ -0.3 & 0.1 & 0.2 \\ 0.4 & -0.1 & 0.3 \end{pmatrix} \begin{pmatrix} 0.5 \\ -1 \\ 2 \end{pmatrix} = (1.2 \ -1.15 \ 0.15 \ 0.9)^T$$

$h = \text{ReLU}(z_1) = \max(0, z_1)$

$= (1.2 \ 0 \ 0.15 \ 0.9)^T$

$Z_2 = W_2 h$

$$= \begin{pmatrix} 0.1 & 0.2 & -0.1 & 0.1 \\ 0.2 & 0.3 & 0.1 & -0.3 \\ 0.2 & -0.1 & 0.3 & 0.2 \end{pmatrix} \begin{pmatrix} 1.2 \\ 0 \\ 0.15 \\ 0.9 \end{pmatrix} = \begin{pmatrix} 0.195 \\ -0.495 \\ 0.465 \end{pmatrix}$$

$\sum e^{z_2} = 1.215 + 0.61 + 1.592 = 3.417$

$$\hat{Y} = \text{softmax}(z_2) = \begin{pmatrix} e^{0.195} \\ e^{-0.495} \\ e^{0.465} \end{pmatrix} = \begin{pmatrix} 1.215 \\ 0.610 \\ 1.592 \end{pmatrix}$$

$$\hat{Y} = \begin{pmatrix} 1.215/3.417 \\ 0.61/3.417 \\ 1.592/3.417 \end{pmatrix} = \begin{pmatrix} 0.356 \\ 0.179 \\ 0.466 \end{pmatrix}$$

$\therefore$ $h = (1.2 \ 0 \ 0.15 \ 0.9)^T$

$\hat{Y} = (0.356 \ 0.179 \ 0.466)^T$

Q1b) $\mathcal{L}_{CE} = (y, \hat{y}) = -\sum_{i=1}^{N} y_i \ln(\hat{y}_i)$

$= -(\ln(0.356) + \ln(0.179) + 0(\ln(0.466))$

$= -(-1.032 + (1.721)) = 2.753$

$$\frac{\partial \mathcal{L}_{CE}}{\partial W_2} = \frac{\partial L_{CE}}{\partial z_2}(h)$$

$$= \begin{pmatrix} -0.644 \\ -0.821 \\ 0.466 \end{pmatrix} (1.2 \;\; 0 \;\; 0.15 \;\; 0.9)$$

$$= \begin{pmatrix} -0.773 & 0 & -0.0966 & -0.58 \\ -0.985 & 0 & -0.123 & -0.739 \\ 0.559 & 0 & 0.0699 & 0.419 \end{pmatrix}$$

$$\frac{\partial L_{CE}}{\partial h} = W_2^T \left(\frac{\partial L_{CE}}{\partial z_2}\right)$$

$$= \begin{pmatrix} 0.1 & -0.2 & 0.2 \\ 0.2 & 0.3 & -0.1 \\ -0.1 & 0.1 & 0.3 \\ 0.1 & -0.3 & 0.2 \end{pmatrix} \begin{pmatrix} -0.644 \\ -0.821 \\ 0.466 \end{pmatrix} = \begin{pmatrix} 0.193 \\ -0.422 \\ 0.122 \\ 0.275 \end{pmatrix}$$

$$\frac{\partial L_{CE}}{\partial z_1} = \frac{\partial L_{CE}}{\partial h} \; ReLU(z_1)$$

$$= \begin{pmatrix} 0.193 \\ 0 \\ 0.122 \\ 0.275 \end{pmatrix}$$

$$\frac{\partial L_{CE}}{\partial W_1} = \frac{\partial L_{CE}}{\partial z_1}(X^T)$$

$$= \begin{pmatrix} 0.193 \\ 0 \\ 0.122 \\ 0.275 \end{pmatrix} (0.5 \;\; -1 \;\; 2)$$

$$\frac{\partial L_{CE}}{\partial W_1} = \begin{pmatrix} 0.0965 & -0.193 & 0.386 \\ 0 & 0 & 0 \\ 0.061 & -0.122 & 0.244 \\ 0.138 & -0.275 & 0.55 \end{pmatrix} //$$

2