Given inputs

$$X = \begin{bmatrix} 0.1 & -0.5 \\ -0.7 & 1.3 \\ 0.5 & -1.5 \\ 1.3 & 0.8 \end{bmatrix}$$

(a)

weight matrix for hidden layer

$$W = \begin{bmatrix} 0.2 & 0.5 & 0.5 \\ -0.2 & 1.5 & 0.9 \end{bmatrix}$$

Biases vector for hidden layer

$$B = \begin{bmatrix} 0.5 \end{bmatrix}$$

$$\begin{bmatrix} 0.3 \\ 0.1 \end{bmatrix}$$

weight matrix for output layer

$$W = \begin{bmatrix} -0.9 & -0.6 \\ 0.6 & 0.8 \\ 0.7 & 0.3 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

(b)

Affine Transformation:—

$$y = Wx + b$$

$$= \begin{bmatrix} 0.1 & -0.5 \\ -0.7 & 1.3 \\ 0.5 & -1.5 \\ 1.3 & 0.8 \end{bmatrix} \begin{bmatrix} 0.2 & 0.5 & 0.5 \\ -0.2 & 1.5 & 0.9 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.3 \\ 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.12 & -0.7 & -0.4 \\ -0.4 & 1.6 & 0.82 \\ 0.4 & -2 & -1.1 \\ 0.1 & 1.85 & 1.37 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.3 \\ 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.62 & -0.4 & 0.1 \\ 0.1 & 1.9 & 0.92 \\ 0.9 & -1.7 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 0.6 & 2.15 & 1.47 \end{bmatrix}$$

(c)

Considering ReLU as activation function, output of hidden layer would be,

$$= \begin{bmatrix} 0.62 & 0 & 0.1 \\ 0.1 & 1.9 & 0.92 \\ 0.9 & 0 & 0 \\ 0.6 & 2.15 & 1.47 \end{bmatrix}$$

(d)

Affine Transformation;

$$y = wx + b$$

$$= \begin{bmatrix} 0.62 & 0 & 0.1 \\ 0.1 & 1.9 & 0.92 \\ 0.9 & 0 & 0 \\ 0.6 & 2.15 & 1.47 \end{bmatrix} \begin{bmatrix} -0.9 & -0.6 \\ 0.6 & 0.8 \\ 0.7 & 0.3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -0.488 & -0.342 \\ 1.694 & 1.736 \\ -0.81 & -0.54 \\ 1.779 & 1.801 \end{bmatrix}$$

(c)

$$\text{softmax}(z_i) = \frac{e^{z_i}}{}$$

$$\frac{e^{z_j}}{\sum_j e^{z_j}}$$

$$= \frac{e^{(-0.488)}}{e^{(-0.488)} + e^{(-0.342)}}$$

$$= 0.463$$

for $z_i = -0.342$

$$= \frac{e^{(-0.342)}}{e^{(-0.488)} + e^{(-0.342)}}$$

$$= 0.536$$

Similarly for every row, we get outputs as,

$$\begin{bmatrix} 0.463 & 0.537 \\ 0.489 & 0.511 \end{bmatrix}$$

| | |
|---|---|
| 0.432 | 0.568 |
| 0.495 | 0.505 |

(f)

using argmax to predict labels,

| 0 | 1 |
|---|---|
| 0.463 | 0.537 |
| 0.489 | 0.511 |
| 0.432 | 0.568 |
| 0.495 | 0.505 |

argmax →

| |
|---|
| 1 |
| 1 |
| 1 |
| 1 |

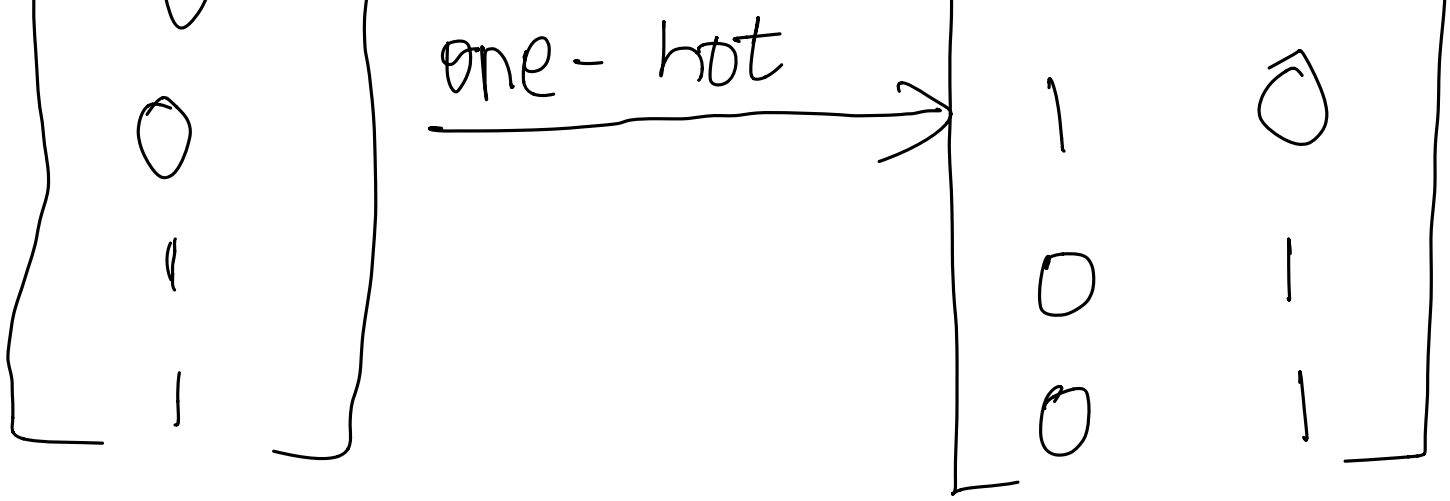
(g)

using one-hot encoding to convert labels,

| | |
|---|---|
| 0 | 1 |
| 0 | 1 |
| 0 | 1 |
| 0 | 1 |

→

| | |
|---|---|
| 1 | 0 |
| 1 | 0 |
| 1 | 0 |
| 1 | 0 |




(h) accuracy :- $\frac{\text{Actual} - \text{Predicted}}{\text{Actual}}$

$$= \frac{\text{No. of correct Predictions}}{\text{No. of Actual Predictions}}$$

$$= \frac{2}{4}$$

$$= 50\%$$

(i)

$$CE\ Loss = -\, y \cdot \log(\hat{y})$$

$$= -\left[0 \log(0.463) + 1 \log(0.537)\right]$$

$$\Rightarrow + 0.27$$

Similarly,

$$= -\left[0 \log(0.489) + 1 \log(0.511)\right]$$

$$= + 0.29$$

$$= -\left[0 \log(0.432) + 1 \log(0.568)\right]$$

$$= + 0.24$$

$$= -\left[0 \log(0.495) + 1 \log(0.505)\right]$$

$= +0.29$

Average CE Loss

$$= \frac{0.27 + 0.29 + 0.24 + 0.29}{4}$$

$$= 0.2725$$