SECTION A

Part A.

Ans 1

Q a) learning rate = 0.01

input ○—w_1—○ (b_1)—w_2—○ (b_2) output   activation func^n ReLU

given input $x = [1, 2, 3]$

$z_1 = w_1 x + b_1$ & $h = \max(0, z_1)$

MSE Loss $\Rightarrow \frac{1}{n} \sum_{i=1}^{n} (y_{target,i} - y_{predicted,i})^2$

let $w_1 = 0.5$, $w_2 = 0.5$, $b_1 = 0.1$ & $b_2 = 0.1$

* For $x_1 = 1$

$z_1 \Rightarrow w_1 x_1 + b_1 \Rightarrow 0.5(1) + 0.1 \Rightarrow 0.6$

$\therefore h = \max(0, 0.6) \Rightarrow 0.6$

$\therefore$ predicted $\Rightarrow w_2 h + b_2 \Rightarrow 0.5(0.6) + 0.1 \Rightarrow 0.4$

* For $x_2 = 2$

$z_2 \Rightarrow w_1 x_2 + b_1 \Rightarrow 0.5(2) + 0.1 \Rightarrow 1.1$

$h \Rightarrow \max(0, 1.1) \Rightarrow 1.1$

Predicted $\Rightarrow w_2 h + b_2 \Rightarrow 0.5(1.1) + 0.5 \Rightarrow 0.65$

* For $x_3 = 3$

$z_3 \Rightarrow w_1 x_3 + b_1 \Rightarrow 0.5(3) + 0.1 \Rightarrow 1.6$

$h \Rightarrow \max(0, 1.6) = 1.6$

Predicted $\Rightarrow (0.5)(1.6) + 0.1 \Rightarrow 0.8 + 0.1 = 0.9$

$\therefore$ MSE $\Rightarrow \frac{1}{3}\left[(0.4-0.3)^2+(0.65-4)^2+(0.9-5)^2\right]$

$\Rightarrow \frac{1}{3}(6.76+11.22+16.81)$

$\Rightarrow 11.59$

* Gradients

For $x=1$

$\frac{\partial L}{\partial pre} = 2(\text{predicted} - \text{target}) = 2(0.4-3) = -5.2$

$\frac{\partial L}{\partial w_2} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial w_2} \Rightarrow -5.2 \times h \Rightarrow -3.12$

$\frac{\partial L}{\partial b_2} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial b_2} \Rightarrow -5.2(0.5)(0.6) \Rightarrow -1.56$

$\therefore \frac{\partial L}{\partial w_1} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial h} \cdot \frac{\partial h}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}$

$\Rightarrow -5.2 \times w_2 \times (1)(x) \Rightarrow -2.6$

$\therefore \frac{\partial L}{\partial b_1} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial h} \cdot \frac{\partial h}{\partial z_1} \cdot \frac{\partial z_1}{\partial b_1} \Rightarrow -2.6$

$w_1 = w_1 - \alpha \frac{\partial L}{\partial w_1} \Rightarrow 0.5 - 0.01(-2.6)$

$w_2 = w_2 - \alpha \frac{\partial L}{\partial w_2} \Rightarrow 0.5 - 0.01(-3.12)$

$b_1 \Rightarrow b_1 - \alpha \frac{\partial L}{\partial b_1} \Rightarrow 0.1 - 0.01(-2.6)$

$b_2 \Rightarrow b_2 - \alpha \frac{\partial L}{\partial b_2} \Rightarrow 0.1 - 0.01(-1.56)$

Gradients for $x = 2$

$$\frac{\partial L}{\partial pre} = 2(0.65-4) = -6.70$$

$$\frac{\partial L}{\partial w_2} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial w_2} \Rightarrow (-6.7)(h) \Rightarrow -6.7 \times 1.1 \Rightarrow -7.37$$

$$\frac{\partial L}{\partial b_2} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial b_2} \Rightarrow (-6.7)(0.55) \Rightarrow -3.685$$

$$\frac{\partial L}{\partial w_1} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial h} \cdot \frac{\partial h}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1} \Rightarrow (-6.7)(0.5)(1)(2)$$
$$\Rightarrow -6.7$$

$$\frac{\partial L}{\partial b_1} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial h} \cdot \frac{\partial h}{\partial z_1} \cdot \frac{\partial z_1}{\partial b_1} \Rightarrow (-6.7)(0.5)(1)(1)$$
$$\Rightarrow -3.35$$

$$w_1 \Rightarrow w_1 - \alpha \frac{\partial L}{\partial w_1} \Rightarrow 0.5 - 0.01(-6.7) \Rightarrow 0.567$$

$$w_2 \Rightarrow w_2 - \alpha \frac{\partial L}{\partial w_2} \Rightarrow 0.5 - 0.01(-7.37) \Rightarrow 0.1237$$

$$b_1 \Rightarrow b_1 - \alpha \frac{\partial L}{\partial b_1} \Rightarrow 0.1 - 0.01(-3.35) \Rightarrow 0.1335$$

$$b_2 \Rightarrow b_2 - \alpha \frac{\partial L}{\partial b_2} \Rightarrow 0.1 - 0.01(-3.685) = 0.1365$$

Gradients for $x = 3$

$$\frac{\partial L}{\partial pre} \Rightarrow 2(0.9-5) \Rightarrow -8.2$$

$$\frac{\partial L}{\partial w_2} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial w_2} \Rightarrow -8.2 \cdot 4 \Rightarrow -13.12$$

$$\frac{\partial L}{\partial b_2} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial b_2} \Rightarrow -8.2 \times 1 \Rightarrow -8.2$$

$$\frac{\partial L}{\partial w_1} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial h} \cdot \frac{\partial h}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}$$

$$\Rightarrow (-8.2)(w_2)(1)(x) = (-8.2)(0.5)(5) = -12.30$$

$$\frac{\partial L}{\partial b_1} \Rightarrow \frac{\partial L}{\partial pre} \cdot \frac{\partial pre}{\partial h} \cdot \frac{\partial h}{\partial z_1} \cdot \frac{\partial z_1}{\partial b_1}$$

$$\Rightarrow (-8.2)(w_2)(1)(1) \Rightarrow -8.2(0.5) \Rightarrow -4.10$$

$$w_2 = w_2 - \alpha \frac{\partial L}{\partial w_2} \Rightarrow 0.5 - 0.01(-13.12) = 0.6312$$

$$w_1 \Rightarrow w_1 - \alpha \frac{\partial L}{\partial w_1} \Rightarrow 0.5 - 0.01(-12.3) = 0.623$$

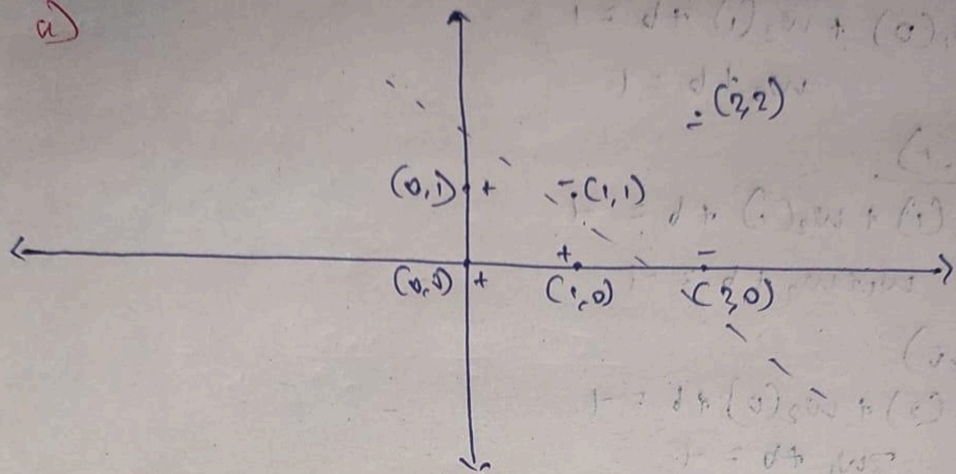
$$b_1 \Rightarrow b_1 - \alpha \frac{\partial L}{\partial b_1} \Rightarrow 0.1 - 0.01(-8.2) = 0.182$$

$$b_2 = b_2 - \alpha \frac{\partial L}{\partial b_2} \Rightarrow 0.1 - 0.01(-4.1) \Rightarrow 0.141$$

PART B.

Part B)

a)



Yes, the given points are linearly separable as a decision boundary line can be drawn b/w them.

b) To find the maximum margin hyperplane, we need to satisfy : $y_i (w x_i + b) \geq 1$

$w = [w_1 \; w_2]$ & b are weights & biases

for the points closest to boundary we have

$y_i (w x_i + b) = 1$

for (+1), (0,1) & (1,0) are closest

for (-1), (1,1) & (2,0) are closest

for (1,0)

$w_1 (1) + w_2 (0) + b = 1$

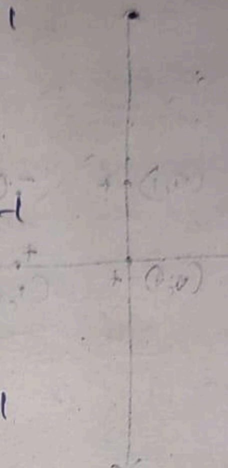
$w_1 + b = 1$

For (0,1)

$$w_1(0) + w_2(1) + b = 1$$
$$w_2 + b = 1$$

For (1,1)

$$w_1(1) + w_2(1) + b = -1$$
$$\therefore w_1 + w_2 + b = -1$$



For (2,0)

$$w_1(2) + w_2(0) + b = -1$$
$$2w_1 + b = -1$$

$w_1 + b = 1$ — (1) , $b \Rightarrow 1 - w_1$

$w_2 + b = 1$ — (2)

$w_1 + w_2 + b = -1$ — (3)

$2w_1 + b = -1$ — (4)

$\therefore$ in eq. (2) $w_2 + 1 - w_1 = 1 \Rightarrow w_2 = w_1$

$\therefore$ in eq (3) $w_1 + w_2 + b = -1 \Rightarrow w_1 + \cancel{w_1} + 1 - \cancel{w_1} =$

$w_1 \Rightarrow -2$

$b = 1 - (-2) \Rightarrow 3$ & $w_2 = -2$

$\therefore$ The eq.^{n} of max. margin hyperplane is

$$-2x_1 - 2x_2 + 3 = 0$$
$$\boxed{x_1 + x_2 = 1.5}$$

Support vectors are

(1,0) , (0,1) for +1

(1,1) , (2,0) for −1

$w = [-1, -1]$

$b = 1.5$

PART C.

a) Calculate the margin of the classifier

The margin m is given by :

$$m = \frac{2}{\|w\|} \quad , \text{ where } \|w\| = \sqrt{w_1^2 + w_2^2}$$

$\therefore \|w\| = \sqrt{(-2)^2 + 0^2} \Rightarrow 2$

$\therefore m = \frac{2}{2} \Rightarrow 1$

b) Identify the support vectors

To find the support vectors, we check if each samples satisfies $y_i(w \cdot x_i + b) = 1$

* Sample 1 : $x_1 = 1$, $x_2 = 2$ & $y_i = +1$

$w \cdot x + b \Rightarrow w_1 x_1 + w_2 x_2 + b$ {given $b = 5$}

$\Rightarrow (-2 \cdot 1) + (0 \cdot 2) + 5$

$\Rightarrow -2 + 5 \Rightarrow 3$

$y \times (w \cdot x + b) = 1 \times 3 \Rightarrow 3 \neq 1$

$\therefore$ This is not a support vector

* Sample 2 : $x_1 = 2$, $x_2 = 3$, $y_i = +1$

$w_1 x_1 + w_2 x_2 + b \Rightarrow (-2 \cdot 2) + (0 \cdot 3) + 5$

$\Rightarrow -4 + 5 = 1$

$y \cdot (w \cdot x + b) \Rightarrow 1 \times 1 = 1$

$\therefore$ This is a support vector

* Sample 3 :- $x_1 = 3$, $x_2 = 3$, $y = -1$

$w \cdot x + b \Rightarrow w_1 x_1 + w_2 x_2 + b$

$\Rightarrow (-2 \cdot 3) + (0 \cdot 3) + 5$

$\Rightarrow -6 + 5 = -1$

$y \times (w \cdot x + b) \Rightarrow -1 \times -1 \Rightarrow 1$

∴ This is a support vector

* Sample 4 :- $x_1 = 4$, $x_2 = 1$, $y = -1$

$w \cdot x + b \Rightarrow w_1 x_1 + w_2 x_2 + b$

$\Rightarrow (-2 \cdot 4) + (0 \cdot 1) + 5$

$\Rightarrow -8 + 5 \Rightarrow -3$

$y \times (w \cdot x + b) = -1 \times -3 \Rightarrow 3 \neq 1$

∴ This is not a support vector

∴ Only sample 2 & 3 are support vectors.

c) Predict the class of a new point $x_1 = 1$, $x_2 = 3$

Substitute x_1 & x_2 in decision function

$w \cdot x + b \Rightarrow (-2 \cdot 1) + (0 \cdot 3) + 5$

$\Rightarrow -2 + 5 = 3$

Since $w \cdot x + b$ is positive, ∴ belongs to (+1) class.

<u>SECTION C</u>

<u>PART 2.</u>
Performance Metrics:

Logistic Activation:

Test Set Accuracy: 0.8518

Final Validation Accuracy: 0.8488

Final Training Accuracy: 0.8632

--

Tanh Activation:

Test Set Accuracy: 0.8813

Final Validation Accuracy: 0.8815

Final Training Accuracy: 0.9396

--

Relu Activation:

Test Set Accuracy: 0.8847

Final Validation Accuracy: 0.8833

Final Training Accuracy: 0.9349

--

Identity Activation:

Test Set Accuracy: 0.8535

Final Validation Accuracy: 0.8519

Final Training Accuracy: 0.8745

Best performing activation function: relu

Test accuracy: 0.8847

PART 3.

Best Configuration Summary:

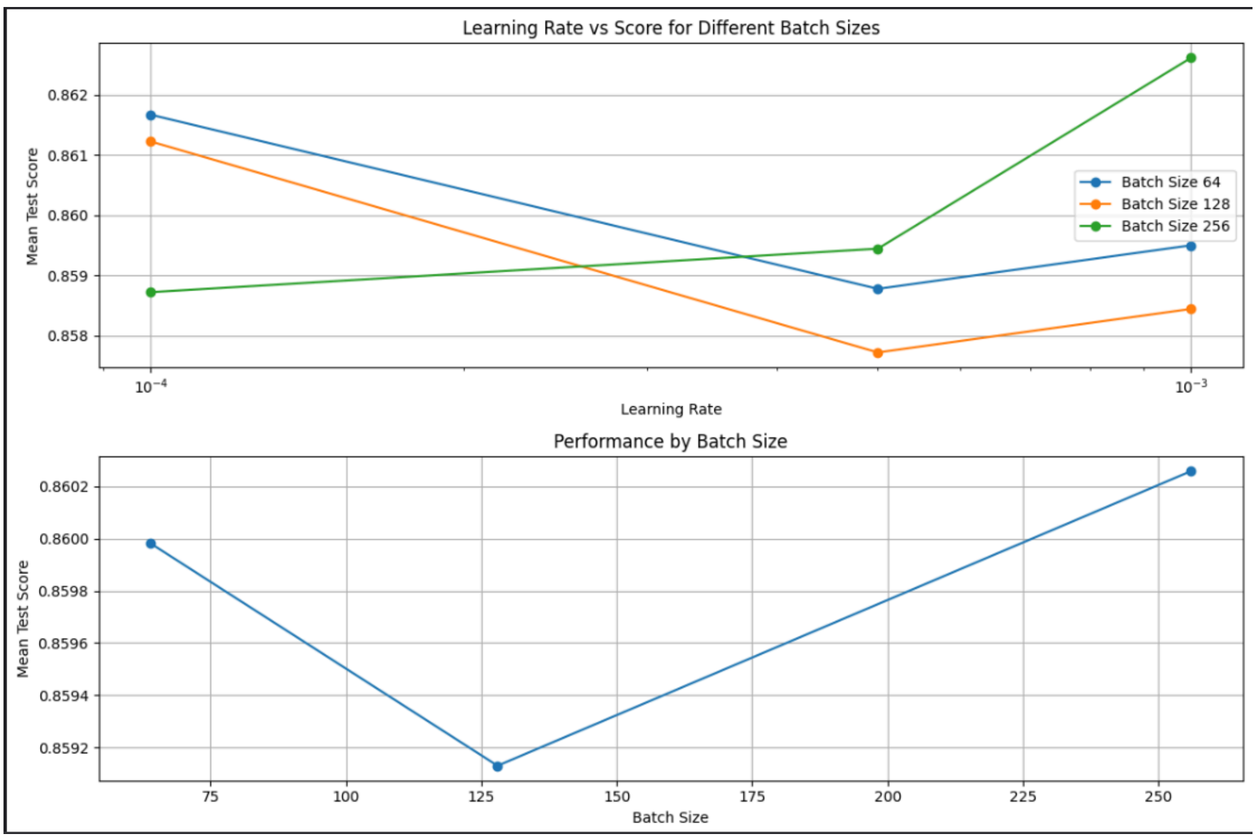
architecture: (128, 64, 32)

solver: adam

batch_size: 256

learning_rate: 0.001

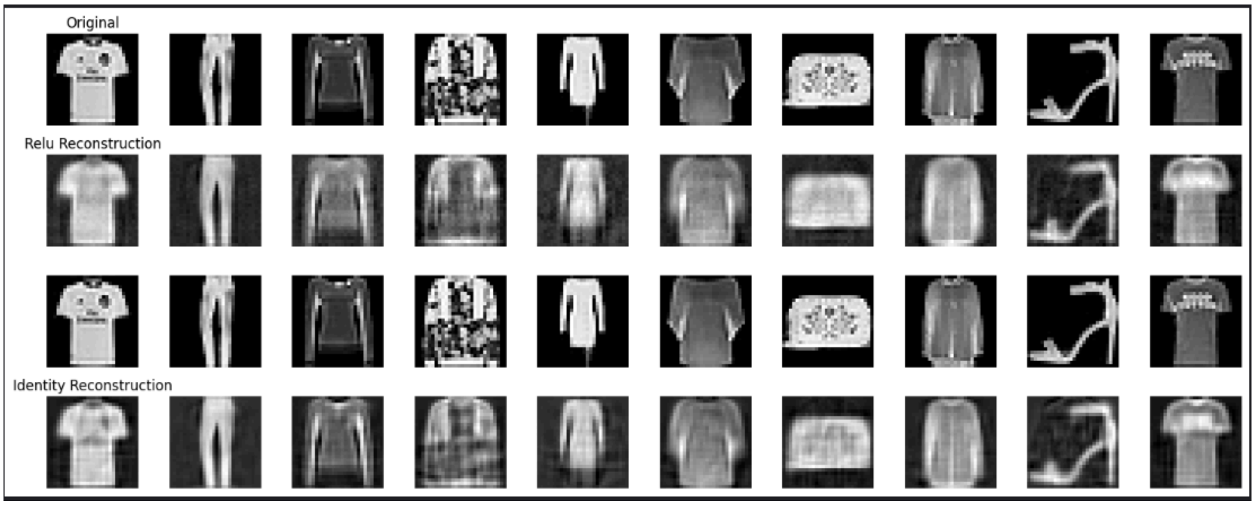
test_accuracy: 0.8885



PART 4.

1. **Visual Quality Comparison:**

- The original images (top row) show various clothing items like t-shirts, pants, dresses, and shoes
- ReLU reconstruction (second row) appears to produce slightly blurrier/fuzzier reconstructions compared to Identity reconstruction
- Identity reconstruction (bottom row) seems to preserve more details and produces sharper reconstructions that look closer to the original images



Relu Activation:
Training Loss: 0.010516
Validation Loss: 0.010506

Identity Activation:
Training Loss: 0.007924
Validation Loss: 0.007916

Identity activation performs better than ReLU for this reconstruction task, as evidenced by:

- ~25% lower loss values
- Better visual quality in reconstructions
- Better preservation of fine details and edges

Both models show very small differences between training and validation losses (around 0.00001), indicating good generalization without overfitting

PART 5.

Despite using a simpler architecture in Part 5, the performance is still decent for several reasons:

a) Quality of Extracted Features:

- The autoencoder has learned meaningful, compressed representations of the images
- Identity activation performed better (87.67%) than ReLU (85.20%), consistent with the lower reconstruction loss we observed
- These features capture essential patterns needed for classification

b) Dimensionality Reduction:

- The autoencoder has effectively reduced the dimensionality while preserving important information
- This makes the classification task easier for the smaller MLP
- Reduced noise and irrelevant features in the data

c) Transfer Learning Effect:

- The feature extraction process is essentially a form of transfer learning
- The autoencoder has already learned useful image representations
- The classifier only needs to learn the mapping from these high-level features to classes