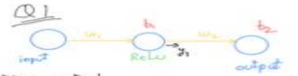
Section-A

Q1



taking initial weights :→ $w_1 = 0.4 \quad b_1 = 0$
$w_2 = 0.6 \quad b_2 = 0$

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad y = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

For z_1 :→

$z_1^{(1)} = 0.4 \times 1 + 0 = 0.4$

$z_1^{(2)} = 0.4 \times 2 + 0 = 0.8$

$z_1^{(3)} = 0.4 \times 3 + 0 = 1.2$

For y_1 :→

$y_1^{(1)} = \mathrm{ReLu}(z_1^{(1)}) = 0.4$

$y_1^{(2)} = 0.8$

$y_1^{(3)} = 1.2$

For z_2 :→

$z_2^{(1)} = 0.6 \times 0.4 + 0 = 0.24$

$z_2^{(2)} = 0.6 \times 0.8 + 0 = 0.48$

$z_2^{(3)} = 0.6 \times 1.2 + 0 = 0.72$

For y_2 :⇒

$y_2^{(1)} = \mathrm{ReLu}(z_2^{(1)}) = \max(z_2^{(1)}, 0)$
$= 0.24$

$y_2^{(2)} = 0.48$

$y_2^{(3)} = 0.72$

MSE :→

$$\left[[y^{(1)} - y_2^{(1)}]^2 + [y^{(2)} - y_2^{(2)}]^2 + [y^{(3)} - y_2^{(3)}]^2 \right] \times \frac{1}{3}$$

$$\Rightarrow \left[[3 - 0.24]^2 + [4 - 0.48]^2 + [5 - 0.72]^2 \right] \times \frac{1}{3}$$

≈ 12.7754

---

update w_2 :⇒

$w_2 \leftarrow w_2 - \eta \frac{\partial L}{\partial w_2}$ → $\frac{1}{3} \sum y_1^{(i)} [y - y_2^{(i)}]$

$w_2 \leftarrow w_2 - 0.01 \times \frac{1}{3} [(3-0.24)0.4 + (4-0.48)0.8 + (5-0.72)1.2]$

$w_2 \leftarrow w_2 - \frac{0.01}{3} \times 9.056$

$w_2 \leftarrow w_2 - 0.03018$

$w_2 \leftarrow 0.6 - 0.03018$

$\boxed{w_2 = 0.56982}$

$b_2 \leftarrow 0 - 0.01 \times \frac{1}{3} [(3-0.24) + (4-0.48) + (5-0.72)]$

$b_2 \leftarrow 0 - 0.01 \times \frac{1}{3} [10.56]$

$b_2 \leftarrow 0 - 0.0352$

$\boxed{b_2 = 0.0352}$

update w_1

$w_1 = w_1 - \eta \frac{\partial L}{\partial w_1}$ → $\frac{1}{3} \sum_i [y^{(i)} - y_2^{(i)}] x^{(i)} w_2$

$w_1 = w_1 - \eta \left[\frac{1}{3} [(3-0.24) \times 1 + (4-0.48) 2 + (5-0.72) 3] \times 0.6 \right]$

$w_1 = 0.4 - \frac{0.01}{3} \times 0.6 [22.64]$

$\boxed{w_1 = 0.35472}$

$b_1 = b_1 - \eta \frac{\partial L}{\partial b_1}$ → $\frac{1}{3} \sum (y^{(i)} - y_2^{(i)}) w_2$

$b_1 = 0 - 0.01 \times \frac{1}{3} [(3-0.24) + (4-0.48) + (5-0.72)] \, 0.6$

$\boxed{b_1 = -0.02112}$

Q3//

sample number	x_1	x_2	y
1	1	2	+1
2	2	3	+1
3	3	3	-1
4	4	1	-1

$w_1 = -2 \quad w_2 = 0 \quad b = 5$

$w^T x + b = 0$

$\Rightarrow \begin{bmatrix} -2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 5 = 0$

$\Rightarrow -2x_1 + 5 = 0 \Rightarrow x_1 = \frac{5}{2}$

a] Margin: $\Rightarrow \frac{2}{\|w\|} \Rightarrow \frac{2}{\sqrt{(-2)^2 + 0}}$

$\Rightarrow \frac{2}{\sqrt{4}} \Rightarrow 1$

b] Finding support vectors: $\rightarrow$

They satisfy $\rightarrow \underline{y(w^T x + b) = 1}$

$(1,2) \longrightarrow +1([-2 \ 0]\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 5)$

$\Rightarrow +1(-2+5) \Rightarrow 3$

$(2,3) \rightarrow +1([-2 \ 0]\begin{bmatrix} 2 \\ 3 \end{bmatrix} + 5)$

$\Rightarrow -4+5 \Rightarrow \boxed{1} \rightarrow$ support vector

$(3,3) \Rightarrow (-1)[[-2 \ 0]\begin{bmatrix} 3 \\ 3 \end{bmatrix} + 5]$

$\Rightarrow (-1)[-6+5] \Rightarrow (-1)(-1)$

$\Rightarrow \boxed{1} \rightarrow$ support vector

$(4,1) \rightarrow (-1)([-2 \ 0]\begin{bmatrix} 4 \\ 1 \end{bmatrix} + 5)$

$\Rightarrow (-1)[-8+5]$

$\Rightarrow 3$

(2,3) and (3,3) are support vectors

c] $x_{new} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

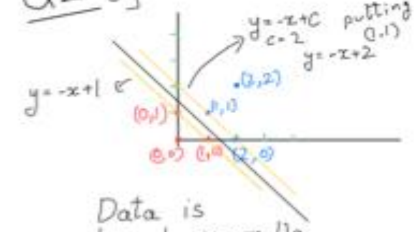
$w^T x_{new} + b = 0$

$\Rightarrow [-2 \ 0]\begin{bmatrix} 1 \\ 3 \end{bmatrix} + 5$

$\Rightarrow 3 > 0$

since $w^T x_{new} + b > 0$, x_{new} will be predicted as $\boxed{+1}$.

Q2 a]



Data is
Linearly seperable

b] Our optimization problem is :→

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

$$\text{s.t.} \quad y_i(w^T x_i + b) \geq 1 \qquad i \in [1, 2, \dots, n]$$

It equivalent
Lagrangean multiplier form is:→

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_i \alpha_i [y_i(w^T x_i + b) - 1]$$

To get the minimum, we diff. w.r.t w, b and α

$$\frac{\partial L}{\partial w} = 0 \quad \frac{\partial L}{\partial b} = 0 \quad \frac{\partial L}{\partial \alpha} = 0$$

Also the α_i's corresponding to the support vectors would be non-zero :→

$0,0 \rightarrow \alpha_1$
$1,0 \rightarrow \alpha_2$
$0,1 \rightarrow \alpha_3$
$1,1 \rightarrow \alpha_4$
$2,2 \rightarrow \alpha_5$
$2,0 \rightarrow \alpha_6$

$$\frac{\partial L}{\partial \alpha_2} = (+1)[w_1 + b - 1] = 0$$
$$\Rightarrow w_1 + b = 1 \quad \text{— (I)}$$

$$\frac{\partial L}{\partial \alpha_3} = (+1)[w_2 + b - 1] = 0$$
$$\Rightarrow w_2 + b = 1 \quad \text{— (II)}$$

$$\frac{\partial L}{\partial \alpha_4} = (-1)[w_1 + w_2 + b] - 1 = 0$$
$$\Rightarrow w_1 + w_2 + b = -1 \quad \text{— (III)}$$

By (I), (II) and (III) :→

$w_2 = -2 \quad w_1 = -2 \quad b = 3$

$$w = \begin{bmatrix} -2 \\ -2 \end{bmatrix} \quad b = 3$$

$(1,0)$; $(0,1)$; $(1,1)$; $(2,0)$
are the support vectors.

Section -B

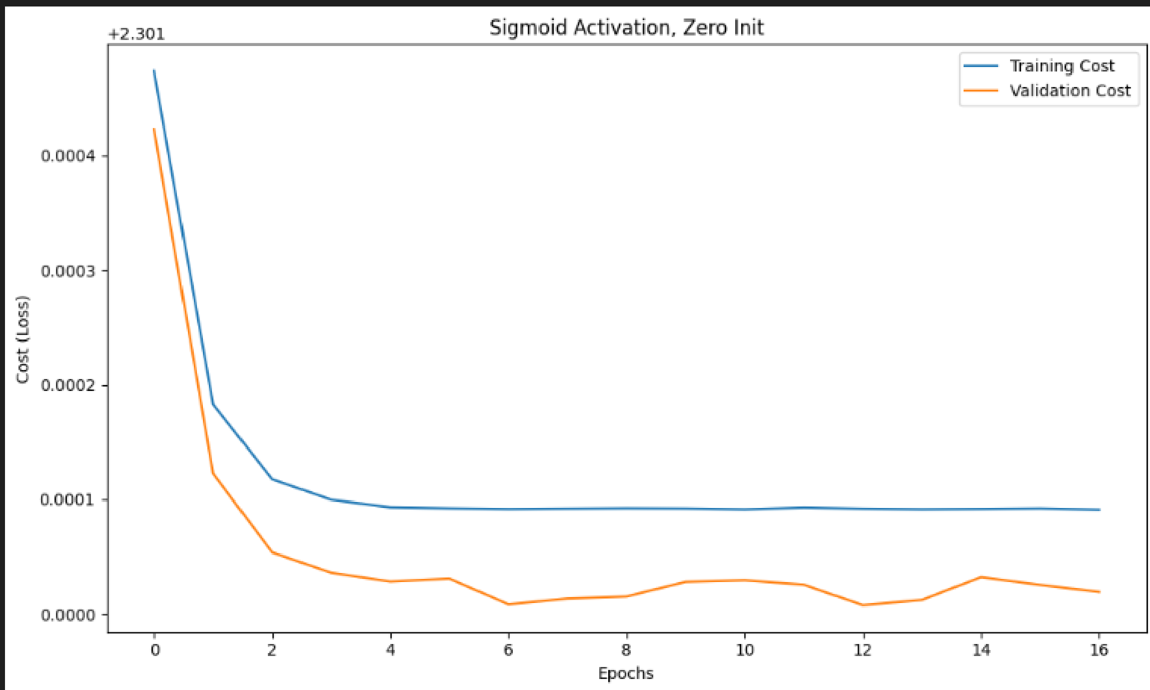
I have kept the patience counter of the fit function as 10 and the threshold as 1e-6. This ensures that if the cost doesn't decrease below the best cost for 10 iteration, the fitting would stop. This is early stopping.

The combination of ReLu and Normal performed the most optimally because it gave the highest accuracy on the test set. But also because the curve for validation and training follow each other very closely, meaning that there is very little overfitting. Also there is a gradual decrease in the cost function for a very low amount of epochs.

The most suboptimal combination is tanh and random. It gives an accuracy of 10.67 on testing set. The training and validation curve mean that the model is unable to capture the pattern. Also the early stopping occurs at 12, meaning that the cost doesn't improve. This might be because the derivative of the tanh function becomes very small even for larger inputs, meaning very low change in weights.

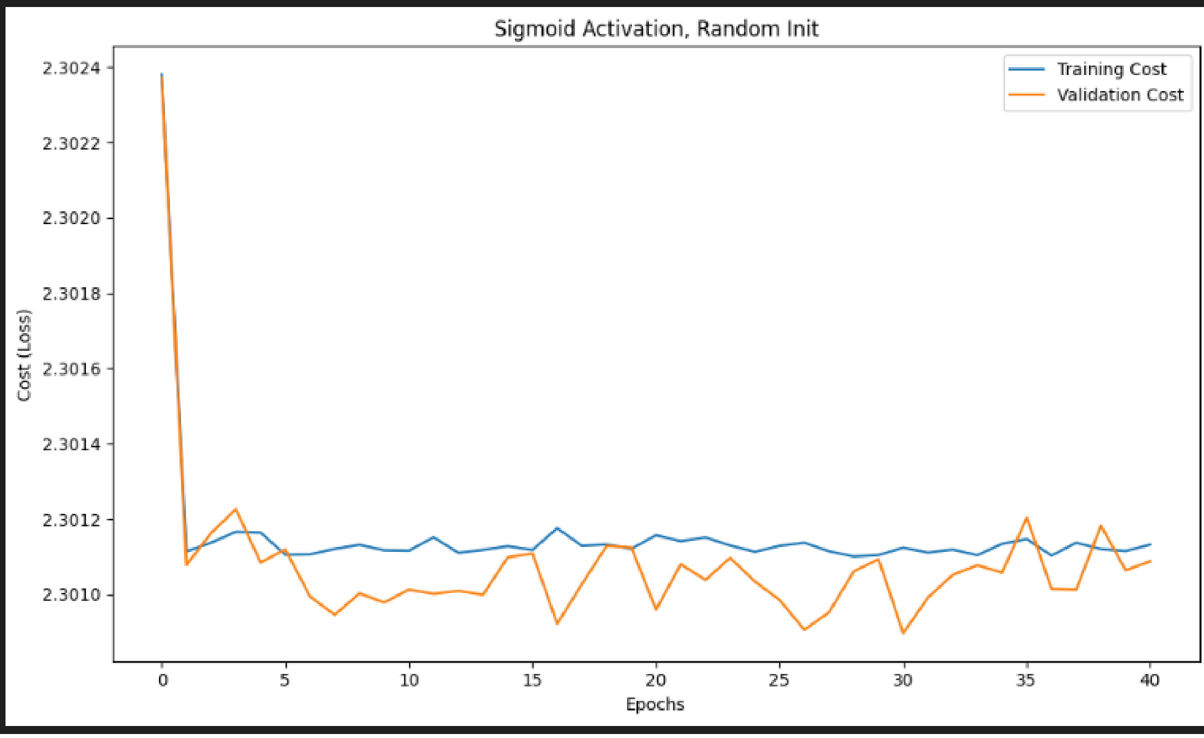
```
Final accuracy for Sigmoid Activation with Zero Init: 10.67%
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



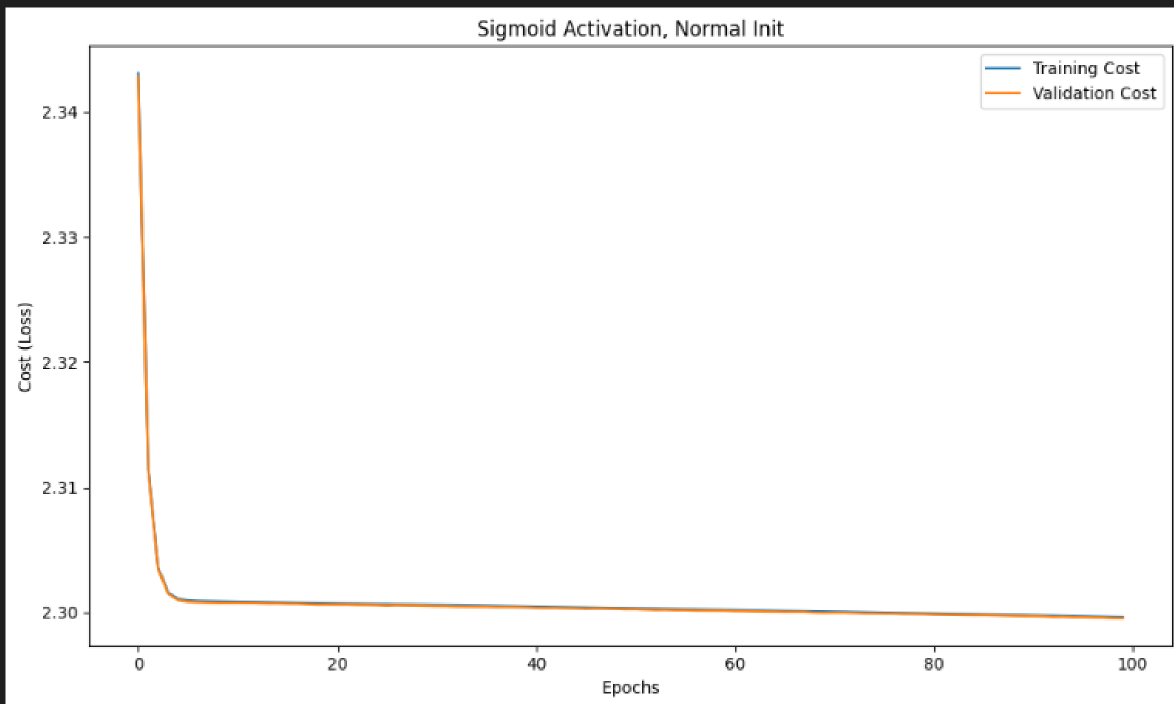
```
Final accuracy for Sigmoid Activation with Random Init: 10.67%
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



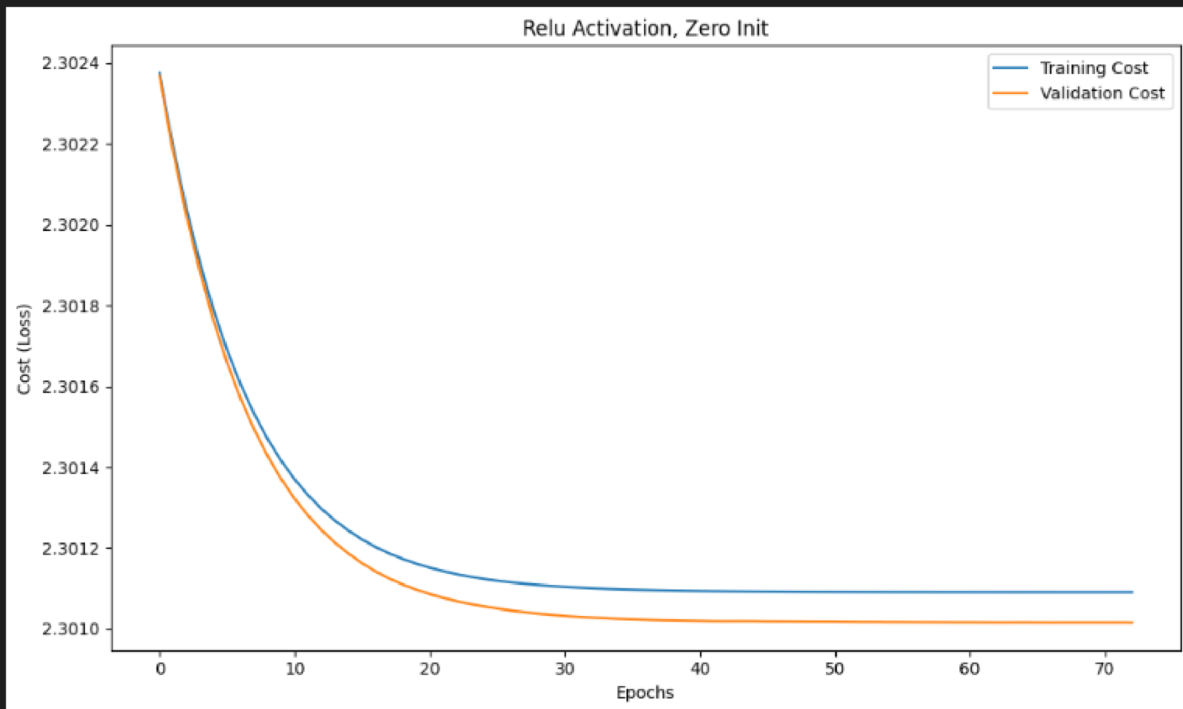
0.0989804 0.09898893 0.09895399 0.09896517 0.09910514 0.09914423]]
Final accuracy for Sigmoid Activation with Normal Init: 10.67%

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



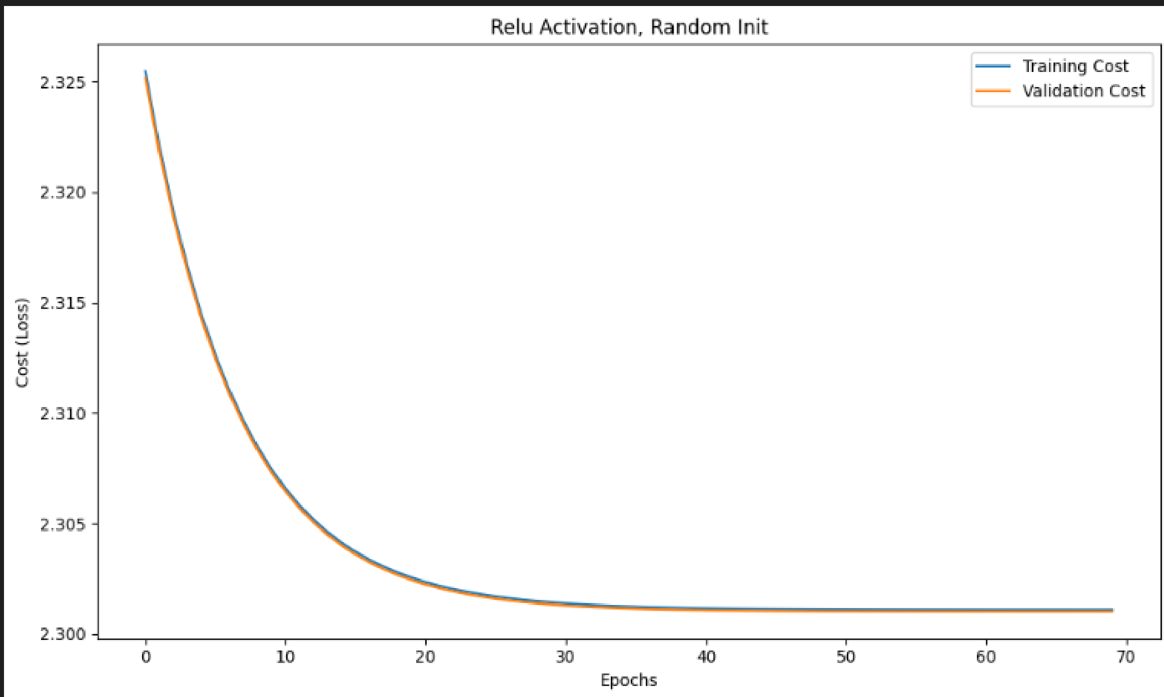
Final accuracy for Relu Activation with Zero Init: 10.67%

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



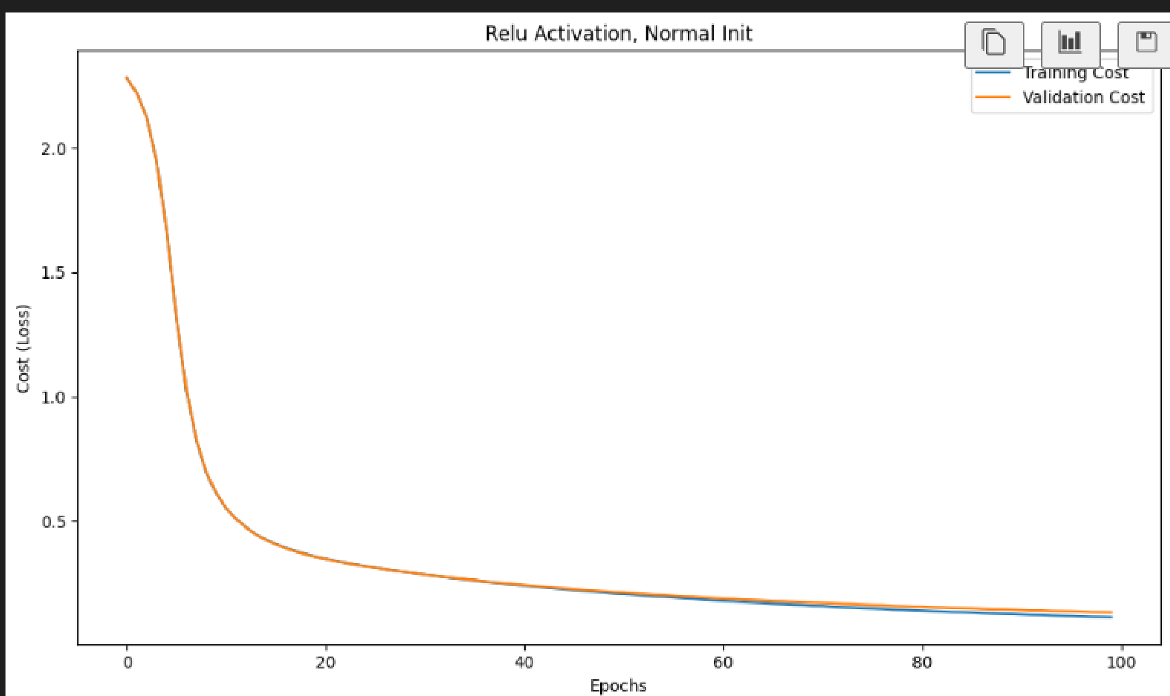
Final accuracy for Relu Activation with Random Init: 10.67%

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



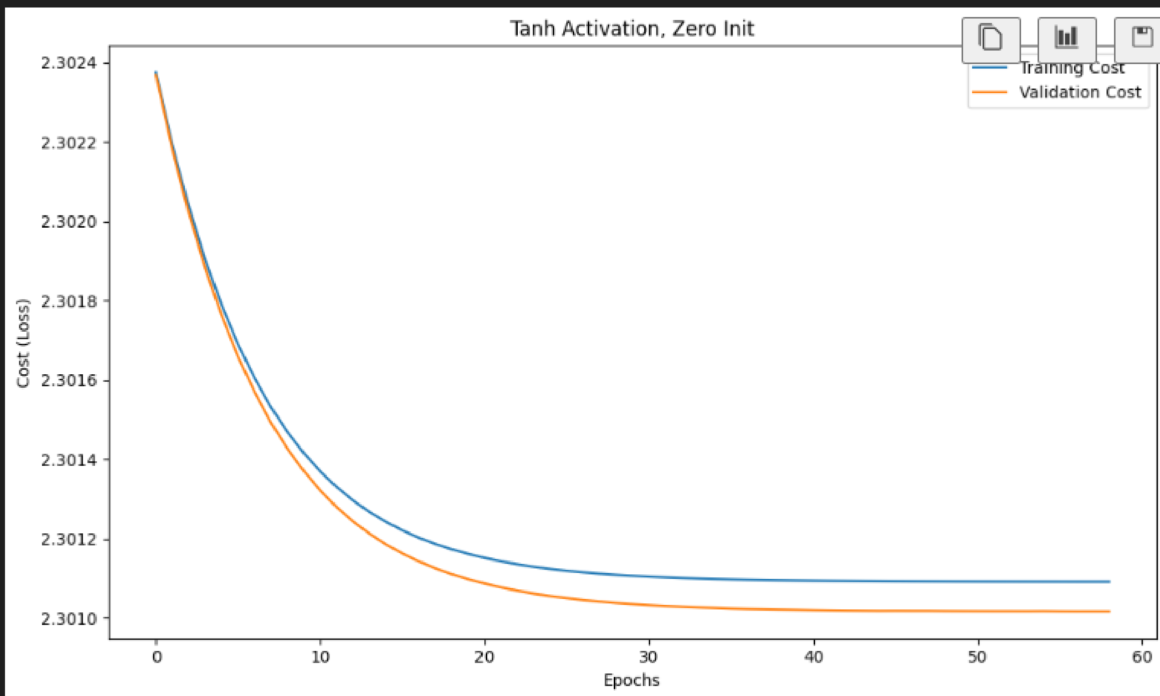
Final accuracy for Relu Activation with Normal Init: 95.87%

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



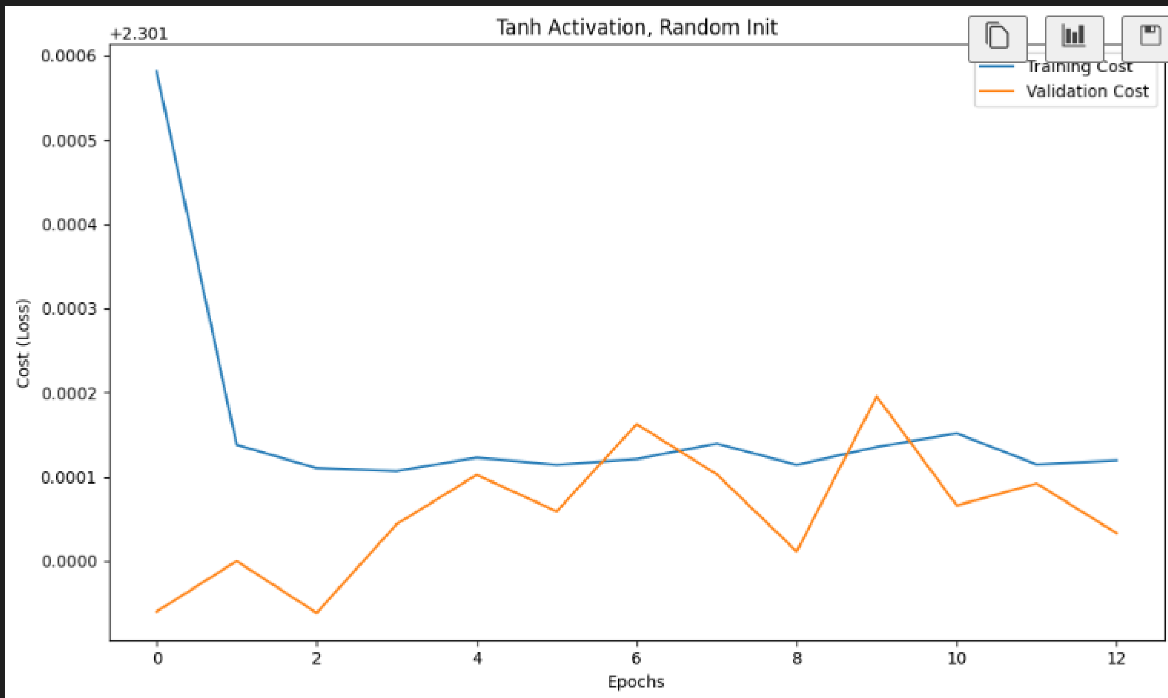
```
Final accuracy for Tanh Activation with Zero Init: 10.67%
```

Output is truncated. View as a [scrollable element]() or open in a [text editor](). Adjust cell output [settings]()...



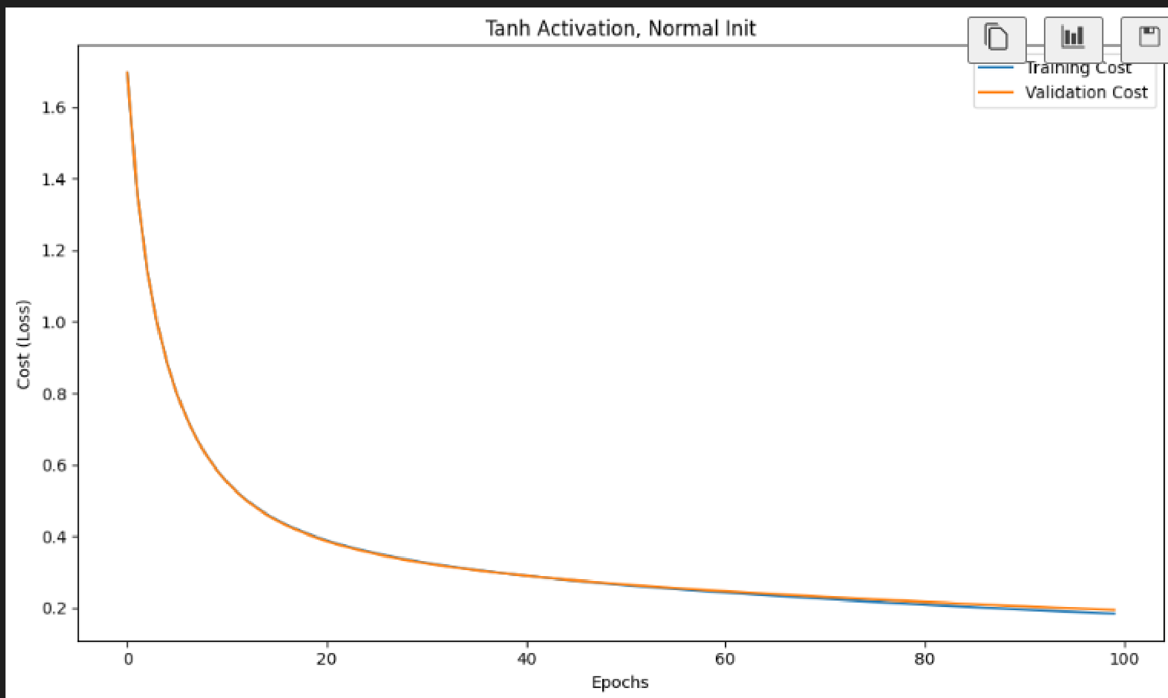
```
Final accuracy for Tanh Activation with Random Init: 10.67%
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



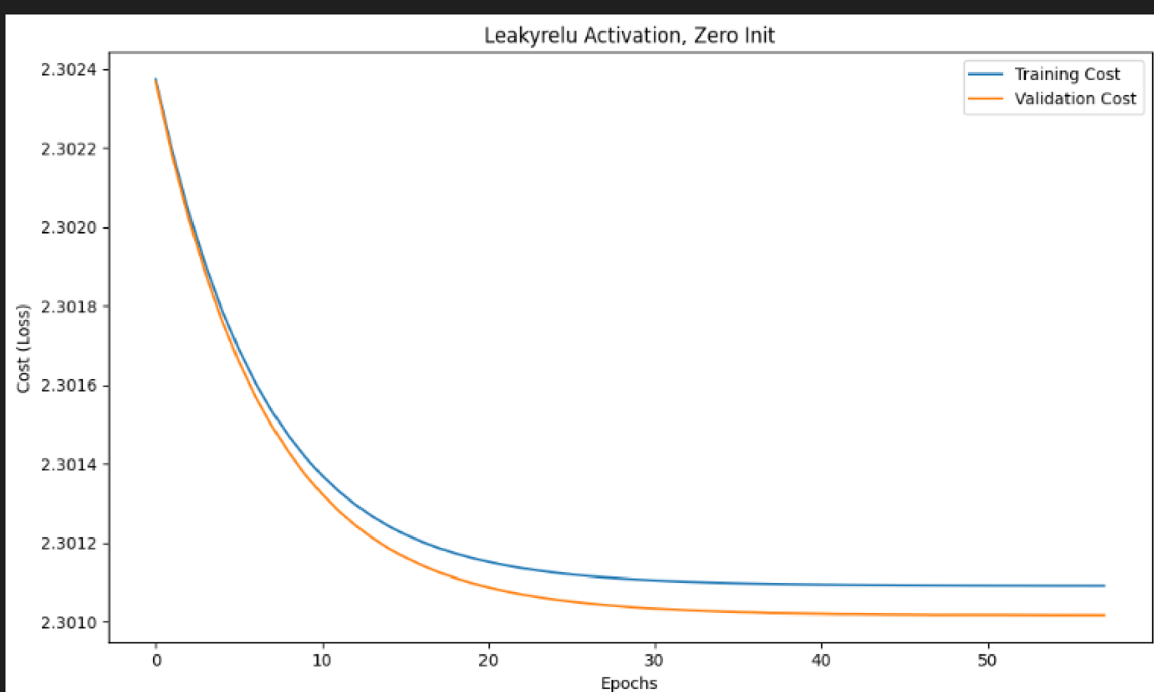
```
Final accuracy for Tanh Activation with Normal Init: 94.28%
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



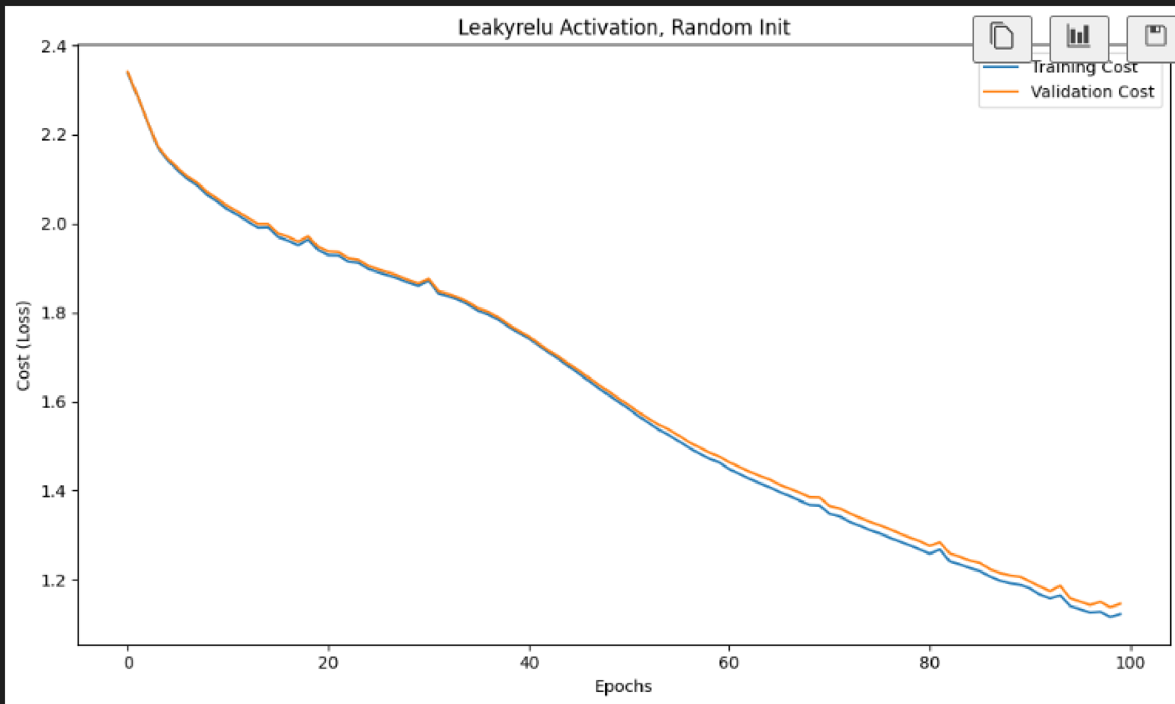

```
Final accuracy for Leakyrelu Activation with Zero Init: 10.67%
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



```
Final accuracy for Leakyrelu Activation with Random Init: 60.80%
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



```
Final accuracy for Leakyrelu Activation with Normal Init: 95.55%
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...