Mohammad Sufyan Azam
2020312

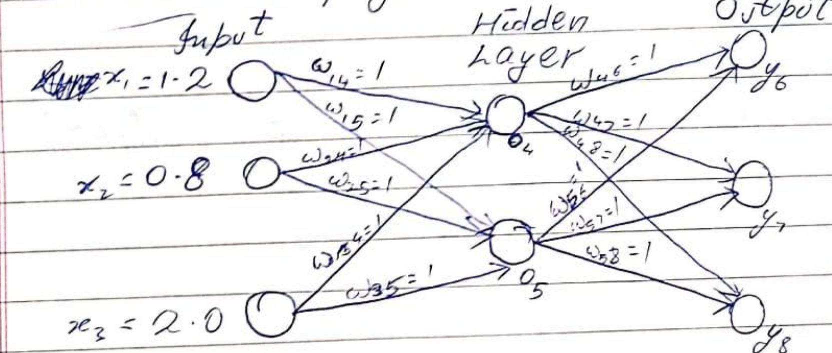
Section A

(a)

change.

(a) Input : [1.2, 0.8, 2.0]
Loss: MSE loss
Activation function: ReLU {max(0, x_i) where $x_i \geq 0$}
Learning rate : $\alpha = 0.01$
Output: [3.0, 2.5, 4.0]

Forward Propagation -



Let the initial weights be (1,1,1,1,1,1) & bias be 0's.

Thus, $o_4' = \omega_{14}x_1 + \omega_{24}x_2 + \omega_{34}x_3 = 4$

$o_4 = RELU(4) = 4$

$o_5' = \omega_{15}x_1 + \omega_{25}x_2 + \omega_{35}x_3 = 4$

$o_5 = RELU(5) = 5$

Mohammad Sufyan Azam
2020312

$y_6 = \cancel{ReLU(}\ w_{46} O_4 + w_{56} O_5 = 9$

$y_7 = w_{48} O_4 + w_{57} O_5 = 9$

$y_8 = w_{48} O_4 + w_{58} O_5 = 9$

Computing the MSE loss –

$$L = \frac{1}{3}\left[(3-9)^2 + (2.5-9)^2 + (4-9)^2 \right]$$

$$= \frac{1}{3}\left[36 + 42.25 + 25 \right]$$

$$\approx 34.42$$

Back Propagation (Updating the weights) –

$$w_{i_{new}} = w_{i_{old}} - \alpha \frac{\delta L}{\delta w_i}$$

Mohammad Sufyan Azam
2020312

DELTA
Date / /
Pg No.

For ~~ou~~ hidden-output Layer ⇒

$$\frac{\delta L}{\delta \omega_{46}} = -2(\hat{y}_{out_1} - y_6)\, o_4 = -2(-6)4 = 48$$

$$\frac{\delta L}{\delta \omega_{47}} = -2(\hat{y}_{out_2} - y_7)\, o_4 = -2(-6.5)4 = 52$$

$$\frac{\delta L}{\delta \omega_{48}} = -2(\hat{y}_{out_3} - y_8)\, o_4 = -2(-5)4 = 40$$

$$\frac{\delta L}{\delta \omega_{56}} = -2(\hat{y}_{out_1} - y_6)\, o_5 = -2(-6)5 = 60$$

$$\frac{\delta L}{\delta \omega_{57}} = -2(\hat{y}_{out_2} - y_7)\, o_5 = -2(-6.5)5 = 65$$

$$\frac{\delta L}{\delta \omega_{58}} = -2(\hat{y}_{out_3} - y_8)\, o_5 = -2(-5)5 = 50$$

$$W_{46}^{new} = W_{46} - \alpha \frac{\delta L}{\delta \omega_{46}} = 1 - 0.01 \times 48 = 0.52$$

$$W_{47}^{new} = 1 - 0.01 \times 52 = 0.48$$

$$W_{48}^{new} = 1 - 0.01 \times 40 = 0.6$$

$$W_{56}^{new} = 1 - 0.01 \times 60 = 0.4$$

$$W_{57}^{new} = 1 - 0.01 \times 65 = 0.35$$

$$W_{58}^{new} = 1 - 0.01 \times 50 = 0.5$$

Mohammad Sufyan Azam
2020312

For input-hidden layer $\Rightarrow$

$$\frac{\delta L}{\delta \omega_{14}} = -2(\hat{y}_{out_1} - y_6)\, W_{46} \cdot x_1 = -2(-6)\,0.52 \times 1.2 = 7.49$$

$$\frac{\delta L}{\delta \omega_{15}} = -2(\hat{y}_{out_2} - y_7)\, W_{47} \cdot x_2 = -2(-6.5)\,0.48 \times 0.8 = 5$$

$$\frac{\delta L}{\delta \omega_{24}} = -2(\hat{y}_{out_3} - y_8)\, \omega_{48} \cdot x_3 = -2(-5)\,0.6 \times 2 = 12$$

$$\frac{\delta L}{\delta \omega_{25}} = -2(\hat{y}_{out_1} - y_6)\, \omega_{56} \cdot x_1 = -2(-6)\,0.4 \times 1.2 = 5.76$$

$$\frac{\delta L}{\delta \omega_{34}} = -2(\hat{y}_{out_2} - y_7)\, \omega_{57} \cdot x_2 = -2(-6.5)\,0.35 \times 0.8 = 3.64$$

$$\frac{\delta L}{\delta \omega_{35}} = -2(\hat{y}_{out_3} - y_8)\, \omega_{58} \cdot x_3 = -2(-5)\,0.5 \times 2 = 10$$

$$\omega_{14}^{new} = \omega_{14} - \alpha \frac{\delta L}{\delta \omega_{14}} = 1 - 0.01 \times 7.49 = 0.93$$

$$\omega_{15}^{new} = 1 - 0.01 \times 5 = 0.95$$

$$\omega_{24}^{new} = 1 - 0.01 \times 12 = 0.88$$

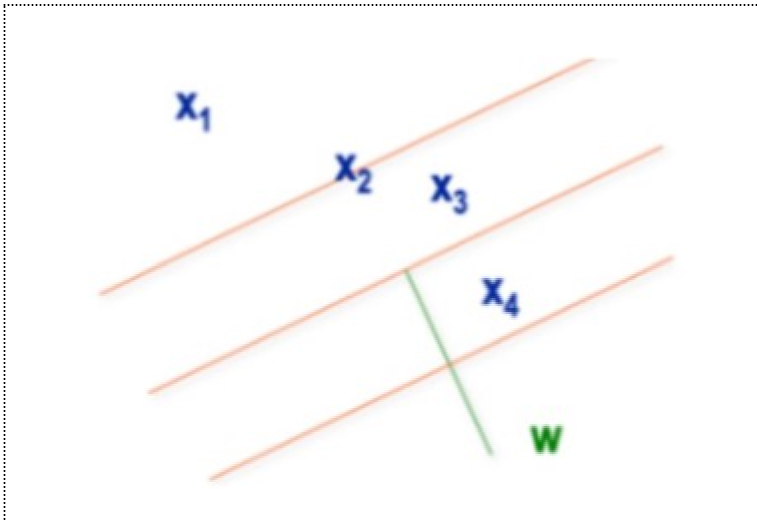
$$\omega_{25}^{new} = 1 - 0.01 \times 5.76 = 0.94$$

$$\omega_{34}^{new} = 1 - 0.01 \times 3.64 = 0.96$$

$$\omega_{35}^{new} = 1 - 0.01 \times 10 = 0.9$$

Mohammad Sufyan Azam
2020312

(b)



(b) (i) Point x1 is on the correct side of the margin. The α_i is typically greater than zero but less than C. $(0 \le \alpha_i < C)$

(ii) Point x2 is on the margin. The α_i parameter ~~lies~~ ~~equal~~ b/w 0 to C. ~~because it~~

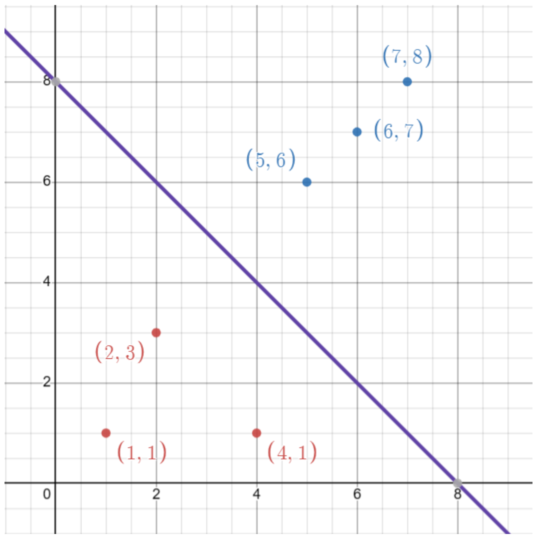
(iii) Point x3 is on the wrong side of margin. The α_i parameter is equal to C.

(iv) Point x4 is on the wrong side of the decision hyperplane. The α_i parameter is typically greater than C.

Mohammad Sufyan Azam
2020312

(c)

(a) The red points denote the positive class while the blue points denote the negative class.



Through plotting the points we can clearly see that the data is linearly separable and one such decision boundary would be x+y-8 = 0;

(b) We can clearly see that the points (2,3), (4,1), and (5,6) will be used to construct the optimal hyperplane. Thus one support vector would pass through (2,3), and (4,1) and the other support vector would pass through the (5,6) point.

Mohammad Sufyan Azam
2020312

(b) ~~Dist~~ Eq'n of support vectors passing through (2,3) & (4,1) →

$$\frac{y-1}{x-4} = \frac{2}{-2} = -1$$

$$\Rightarrow x+y=5$$

Hence, line parallel to it & passing through (5,6) would be - ~~Eq~~,

$$x+y=c$$

$$\Rightarrow 5+6=c$$

$$\Rightarrow x+y=11$$

Hence, support vectors are - $x+y=5$ & $x+y=11$.

Mohammad Sufyan Azam
2020312

Now, the optimal hyperplane would be-

$$x+y=k \quad \text{where } k=\frac{5+11}{2}=8$$

Hence, $x+y=8$ is the optimal ~~hyp~~ ~~hyperplan~~ margin or the decision boundary.

(c) The support vectors as previously identified are - (2,3), (4,1) & (5,6) while their equations are - $x+y=5$ & $x+y=11$.

~~(d) Let a_1 be a point in $x+y=8$ line.~~
~~Therefore, $a_1=(a, 8-a)$~~
~~Calculating its perpendicular distance with $x+y=11$, we~~

(d) Distance would be $d_1=\frac{|c_2-c_1|}{\sqrt{1+m^2}}=\frac{|11-8|}{\sqrt{1+(-1)^2}}$

$\Rightarrow d_1=\frac{3\sqrt{2}}{2}$ b/w $x+y=8$ & $x+y=11$

Distance b/w $x+y=8$ & $x+y=5 \Rightarrow d_2=\frac{|5-8|}{\sqrt{1+(-1)^2}}=\frac{3}{\sqrt{2}}$

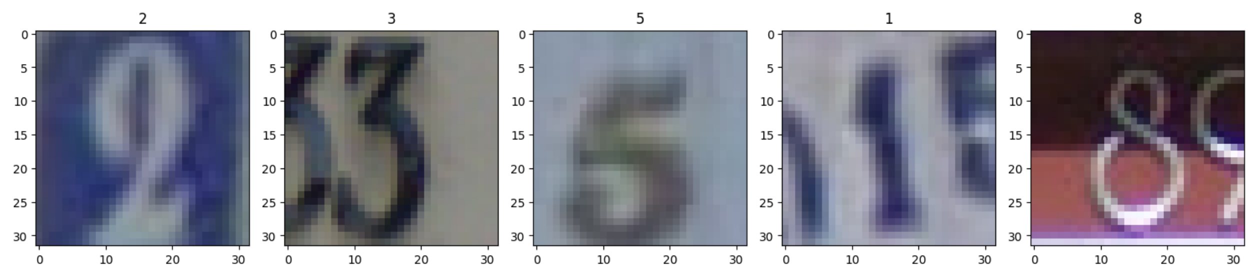
$\Rightarrow d_2=\frac{3\sqrt{2}}{2}$

(e) The optimal margin would not change if any point other than the support vectors would be removed. If we remove any of the support vectors, then the optimal margin would change, with the exception of one of (2,3) or (4,1) being removed. If both are removed then the optimal margin would change.

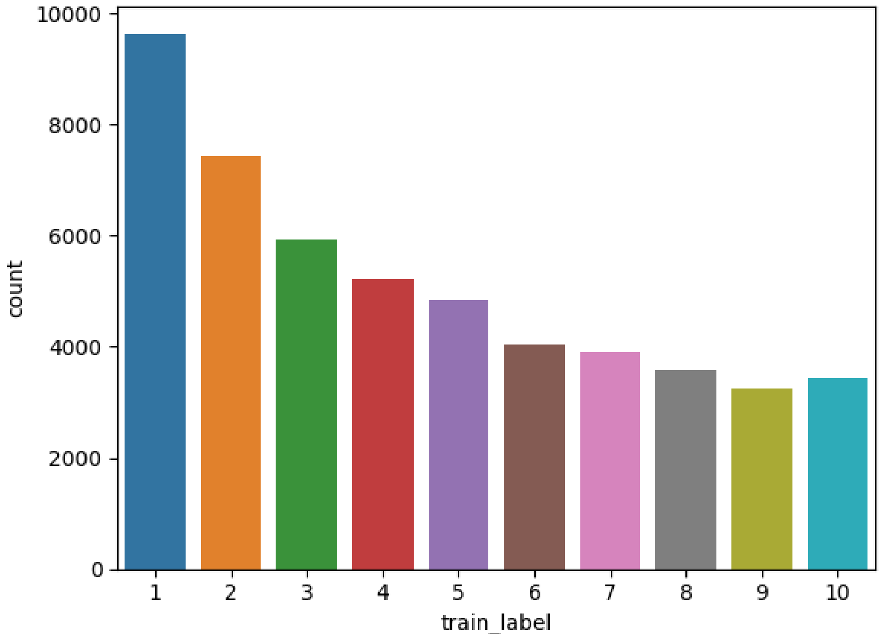
Section B

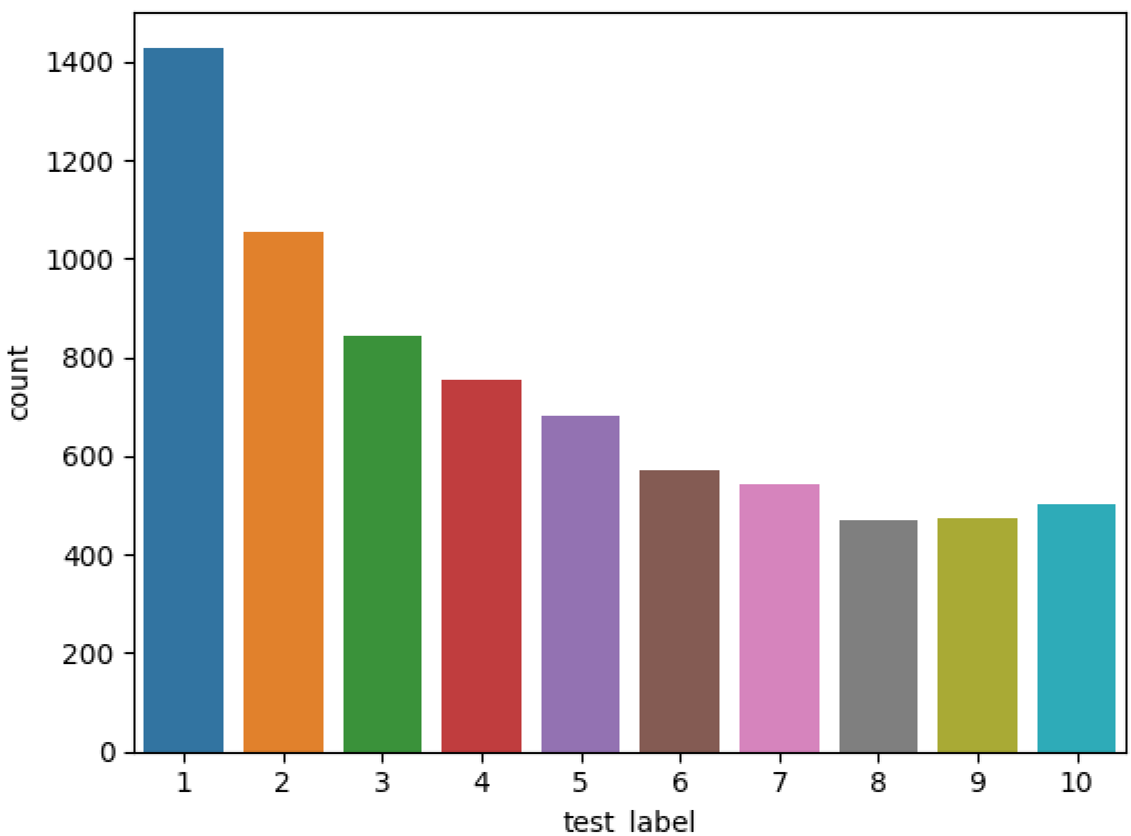
Section C

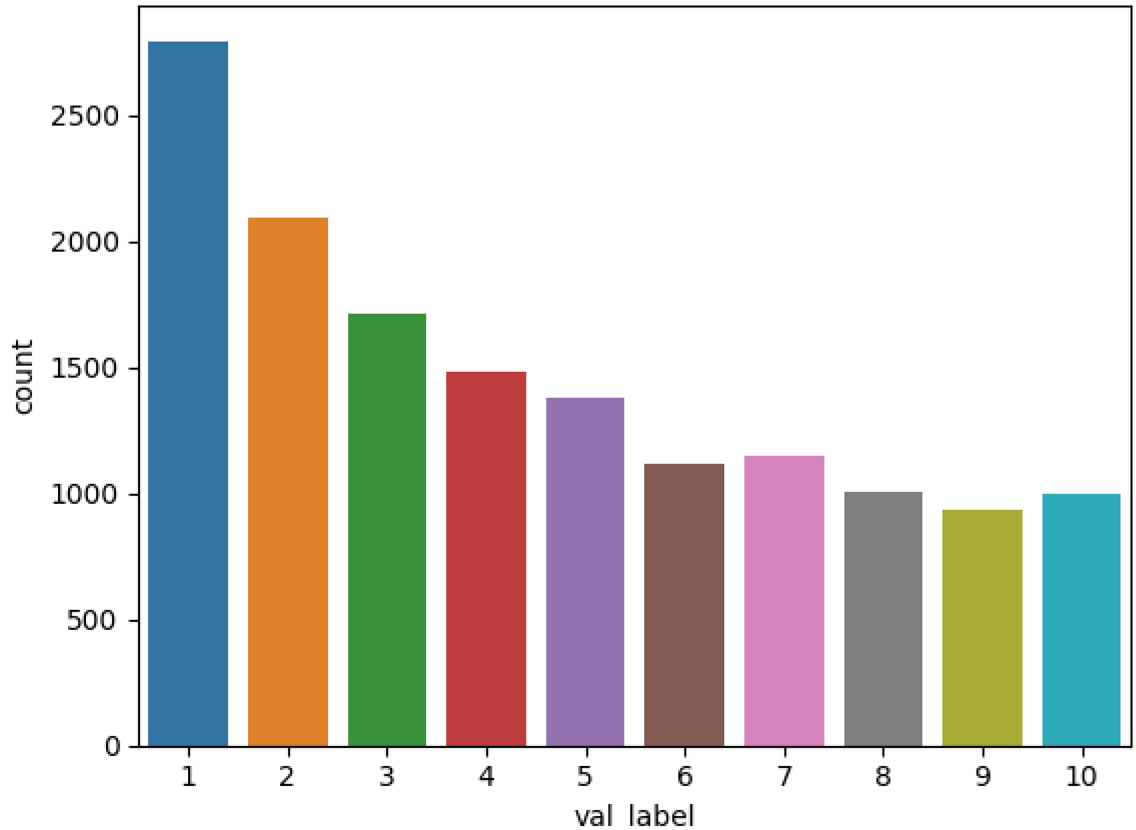
5 Unique Images -



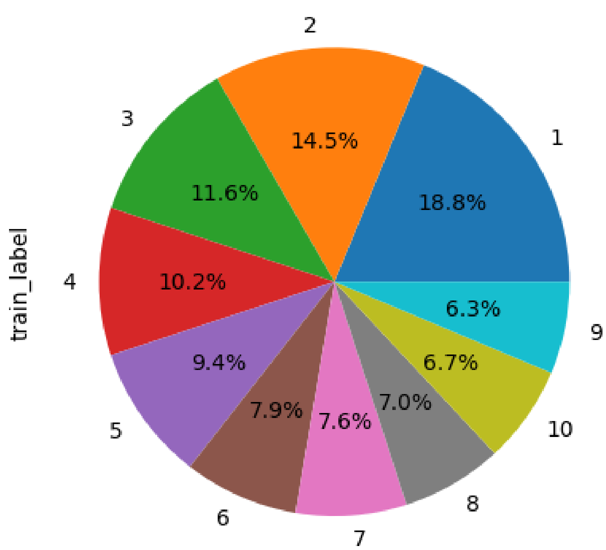
Distribution of class labels -


10000
8000
6000
4000
2000
0
count
1
2
3
4
5
6
7
8
9
10
train_label

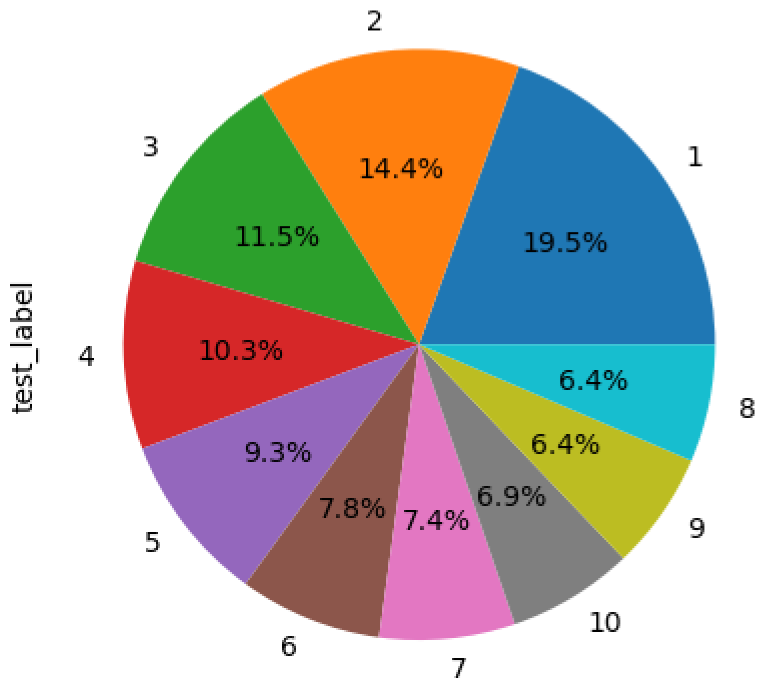

1400
1200
1000
800
600
400
200
0
count
1
2
3
4
5
6
7
8
9
10
test_label

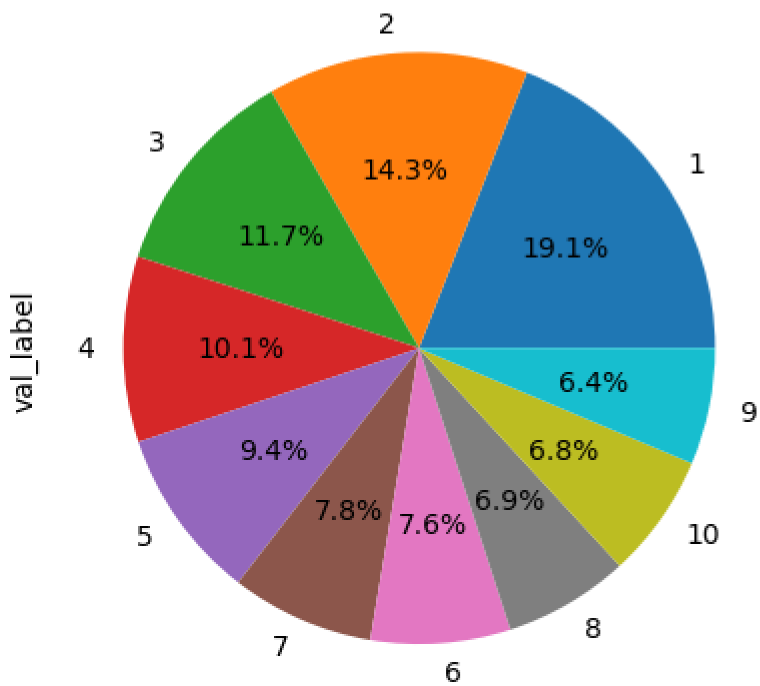


Pie Chart for the class labels distribution -



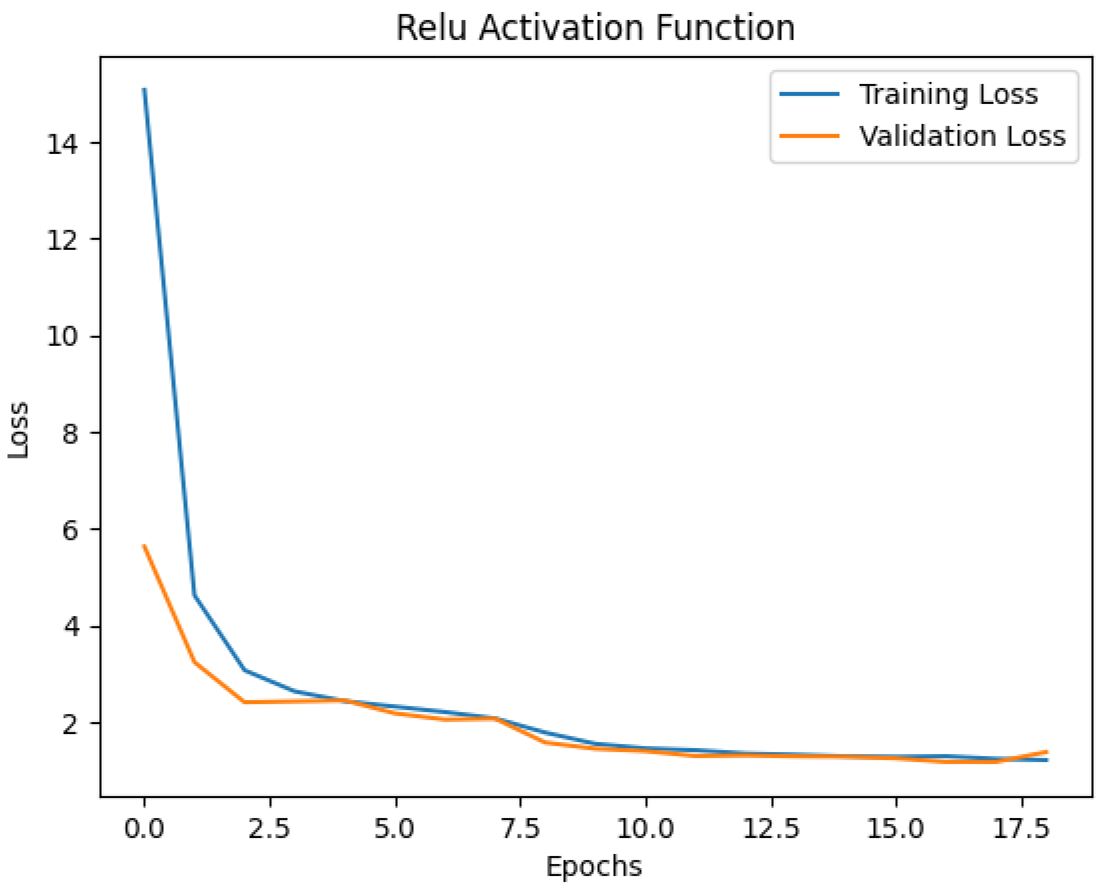
Mohammad Sufyan Azam
2020312



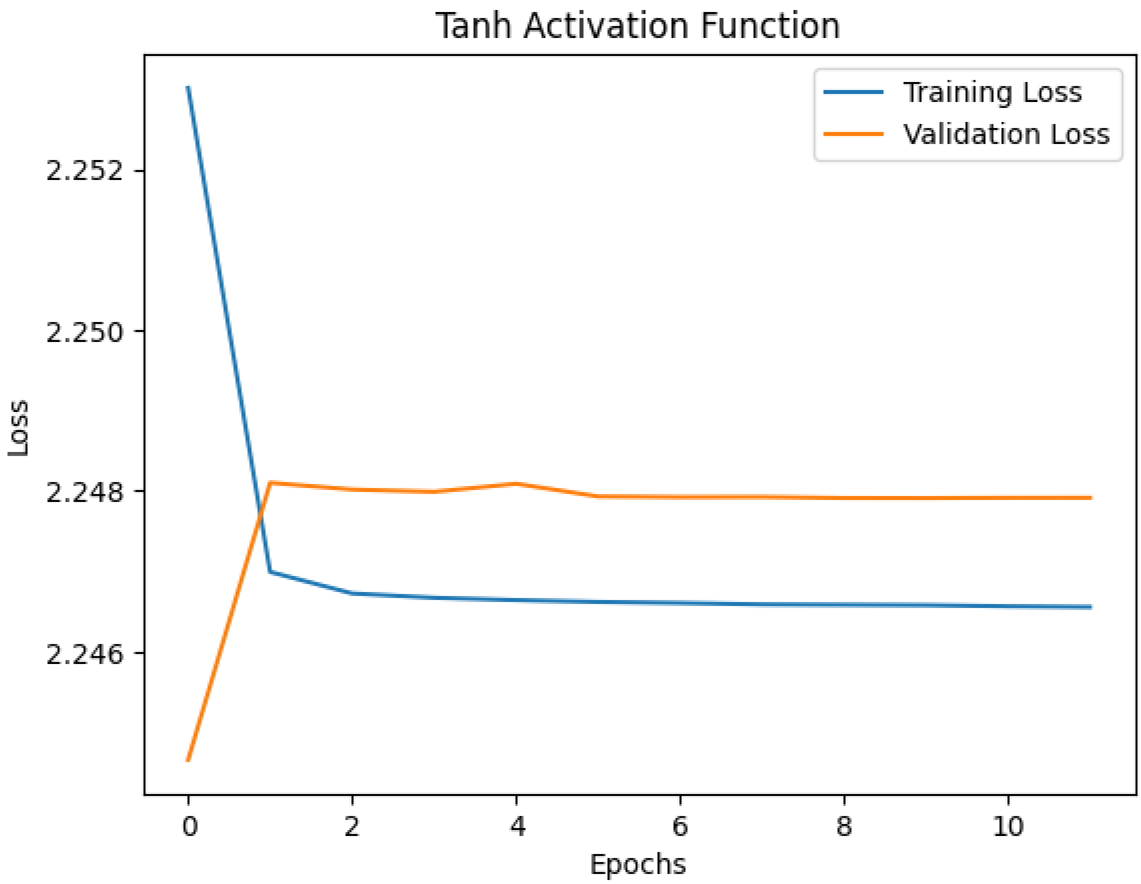


Mohammad Sufyan Azam
2020312

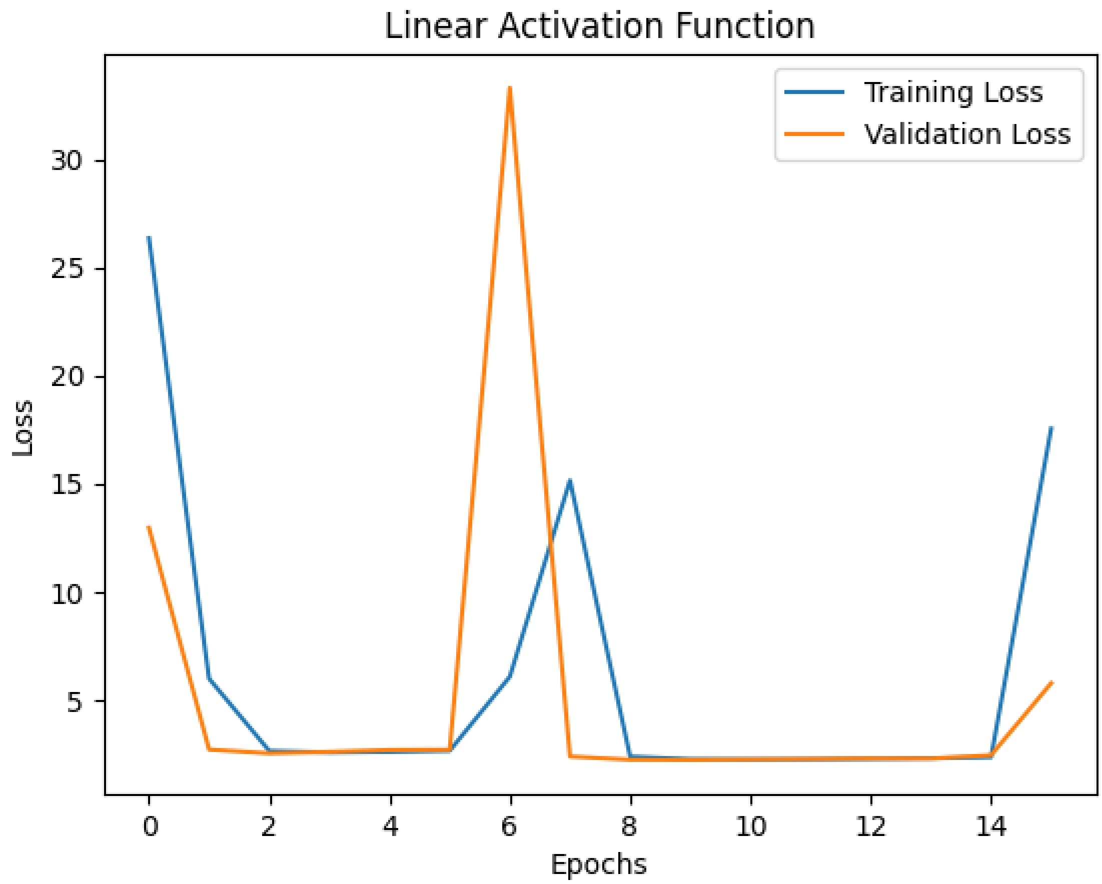
Activation Function Graphs for training loss and validation loss vs epochs -

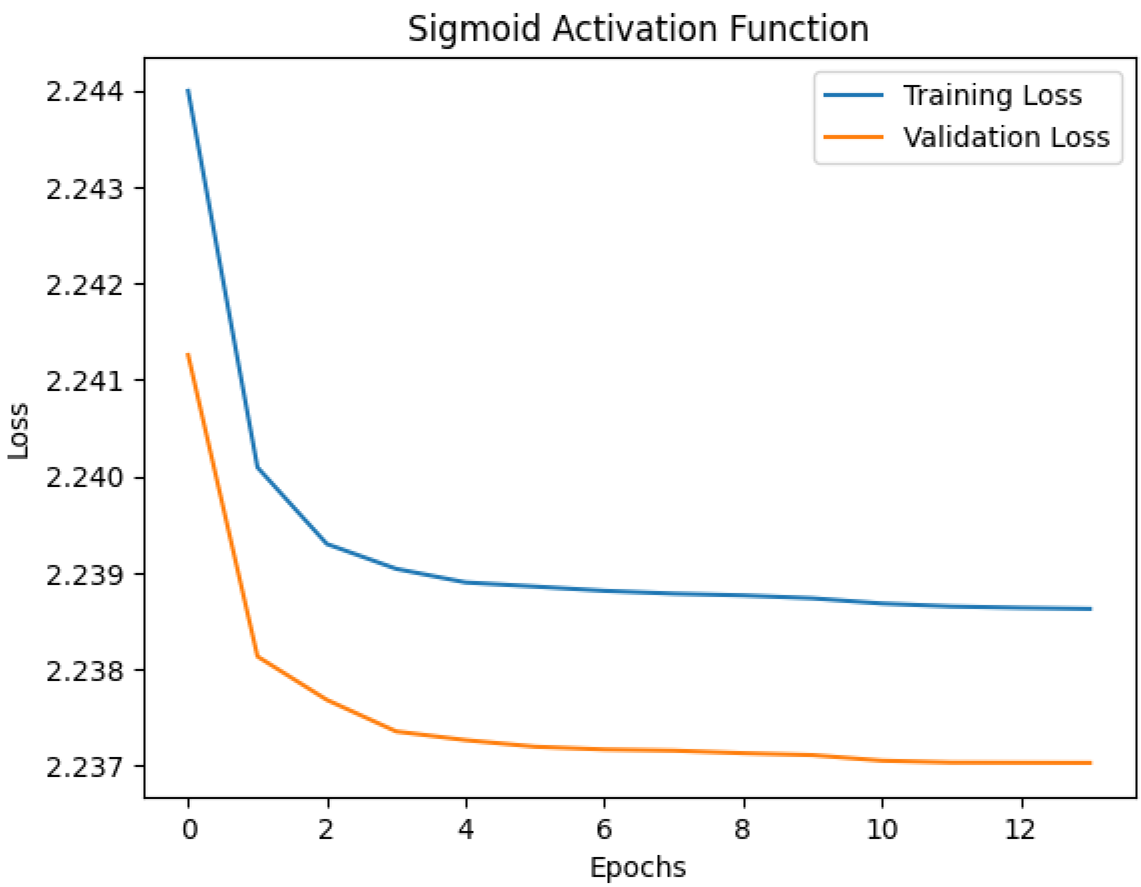


Mohammad Sufyan Azam
2020312



Mohammad Sufyan Azam
2020312





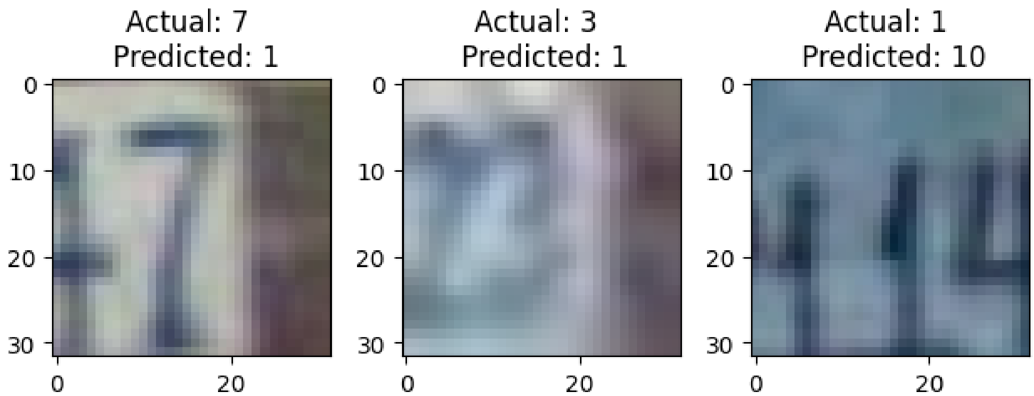
We get a decent accuracy after experimenting with different parameters for the MLPClassifier ~ 70%.

```
 20%|██        | 40/200 [17:57<1:11:49, 26.94s/it]
Stopping early ...

Accuracy with hidden layer size (256, 128) on training set with activation function relu: 0.6924651924651924
```

Mohammad Sufyan Azam
2020312

Misclassifications -



Misclassifications occur due to several reasons such as:

- The model might not have learned the distinguishing features of the classes well enough during training.
- The misclassified image might contain noise or other elements that confuse the model.
- The model might be overfitting to the training data, causing it to perform poorly on unseen data.