Assignment - 3

Nueral Networks

Prateek Mishra (2020EE10527)

Shubh Goel (2020EE10672)

Bharat Kumar (2020EE10587)

ELL 409

Date of Submission - 13th November, 2022

Contents

1 Network Architecture

- The network architecture opted for training neural network was : '3 Hidden Layers' with '32 neurons' each and a 'output layer' with 'single neuron'.
- The Activation Function used for hidden layers was 'ReLu' and 'Sigmoid' for output layer.
- Initial Learning Rate Chosen = 1
- The weight matrix is initialised using the default initializer : 'glorot_uniform'
- Stopping Criterion : 'val_acc' with *patience* = 25, i.e., the training stops if validation accuracy doesn't improve for 25 consecutive epochs.
- momentum = 0.5
- Learning Rate Update Schedule = Exponential Decay with *decay rate* = 0.01.

2 Optimum Initial Learning Rate

Initial Learning Rate	Validation Accuracy
1	47.49
0.1	87.21
0.01	88.58
0.001	80.37
0.0001	52.51

- For learning rate update schedule : exponential decay with *decay rate* = 0.01, the **optimum value** of initial learning rate = 0.01.

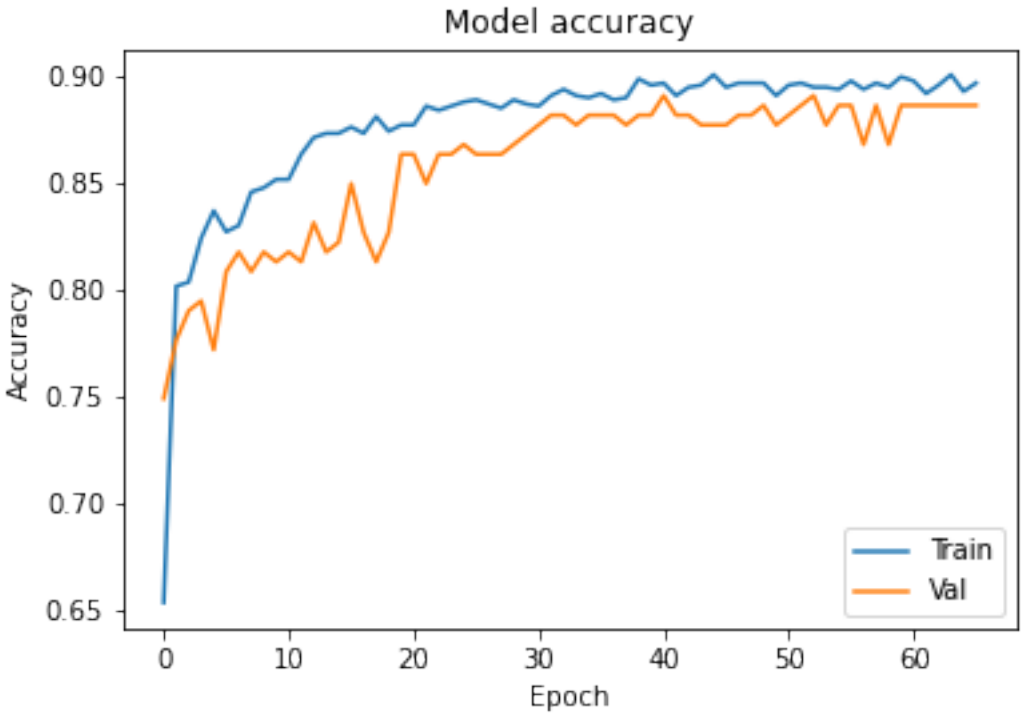

Figure 1: Plot of Model Accuracy on training and validation set with each epoch

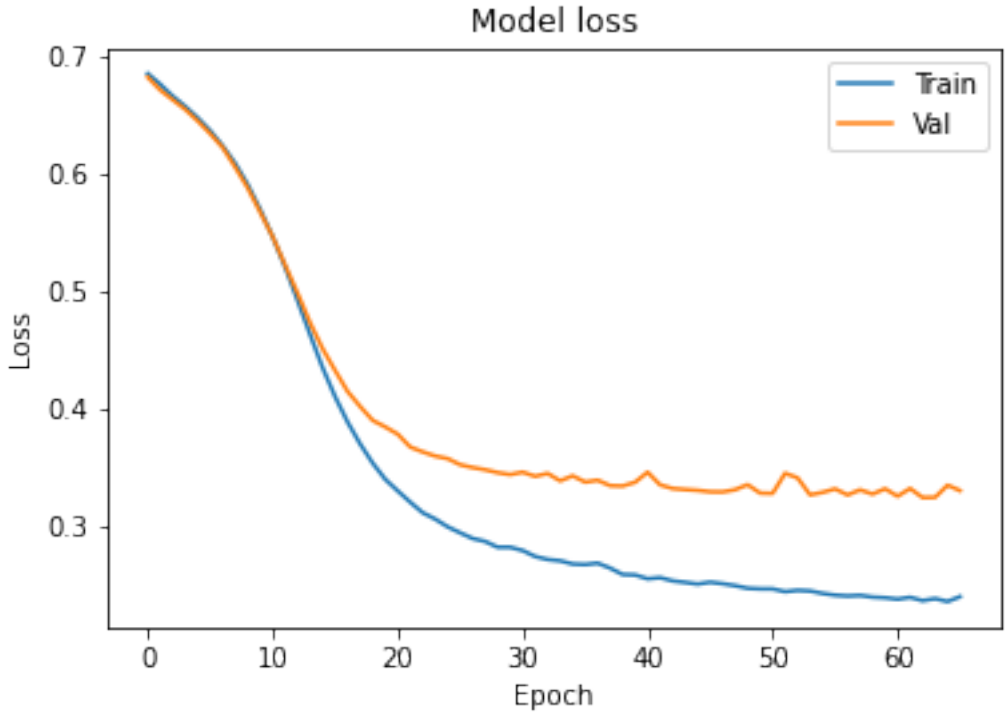

Figure 2: Plot of Model Loss on training and validation set with each epoch

3 Optimum Learning Rate Schedule

We will use these two following learning rate update schedule for our analysis :

- Exponential Decay : lr_new = lr * exp {- decay_rate}
- Time based decay : lr_new = lr / (1 + decay_rate * epoch)

Learning Rate Schedule	Decay Rate	Validation Accuracy
Exponential	0.01	88.58
Exponential	0.005	88.58
Exponential	0.1	84.93
Time based	0.01	80.37
Time based	0.001	87.21
Time based	0.0005	88.13

- The **optimum** learning rate update scheduler : Exponential Decay with decay_rate = 0.005.

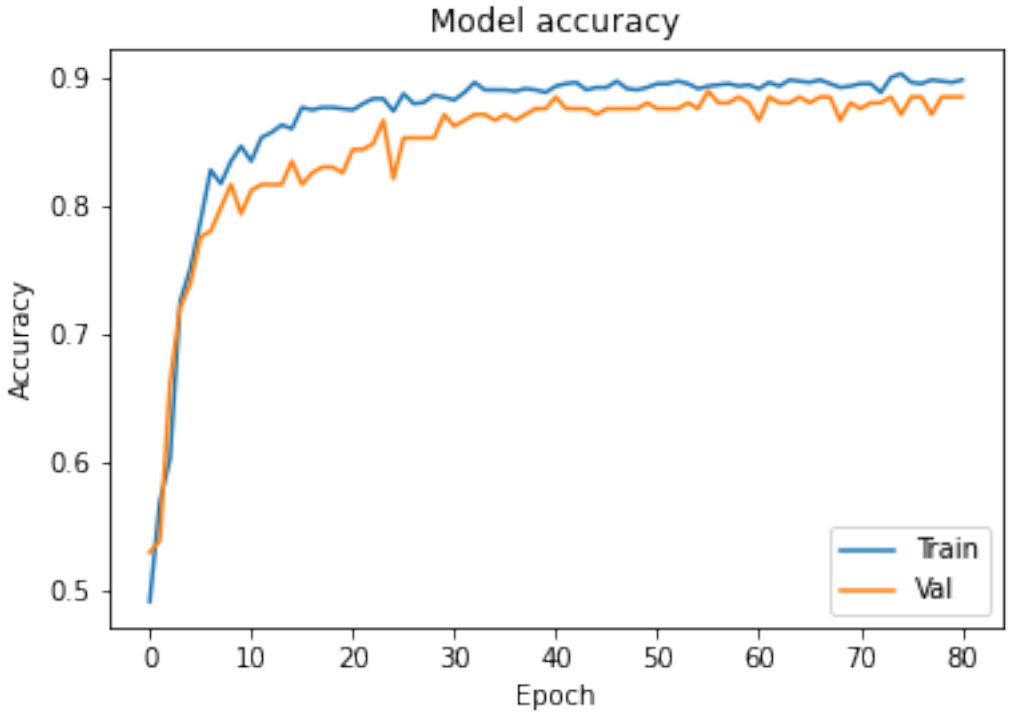

Figure 3: Plot of Model Accuracy on training and validation set with each epoch

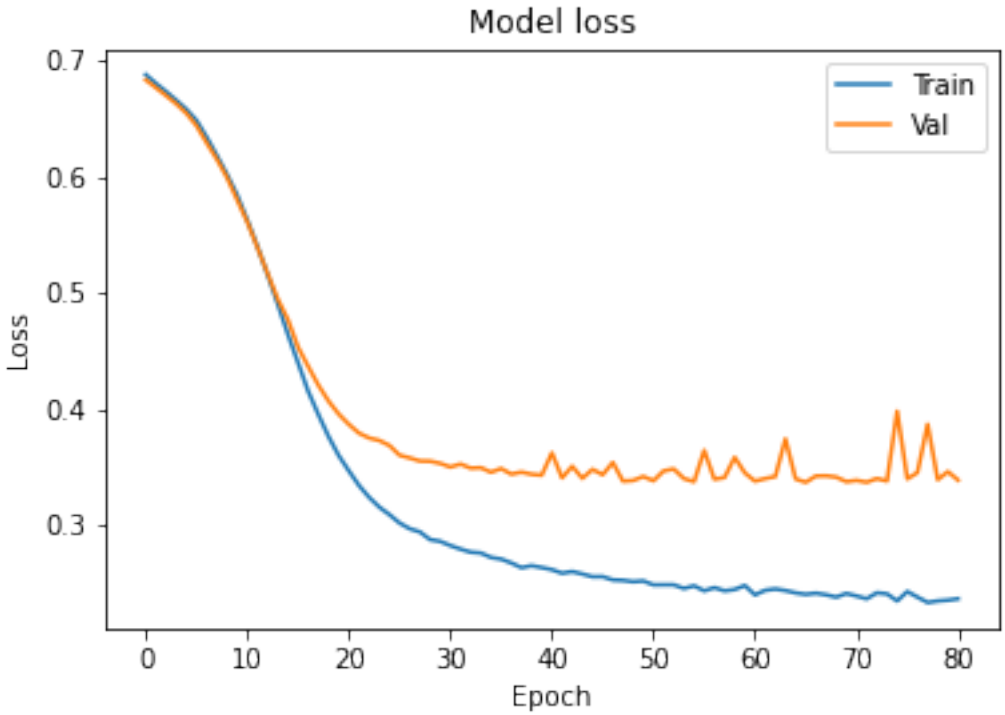

Figure 4: Plot of Model Loss on training and validation set with each epoch

4 Effect of Dropout in Network Architecture

- Dropout Layer was added after the 2nd layer.
- It did not improve the accuracy.

Dropout Rate	Validation Accuracy
0.5	88.13
0.4	86.76

5 Effect of L1 and L2 Regularization

- Regularization is used to prevent over-fitting and thus improve model accuracy on unseen data.

5.1 L1 Regularization

Lambda (λ)	Validation Accuracy
0.0001	89.04
0.001	89.50
0.01	89.50
0.1	47.49
1	47.49

- We do not observe any improvement in accuracy using L1 regularization.

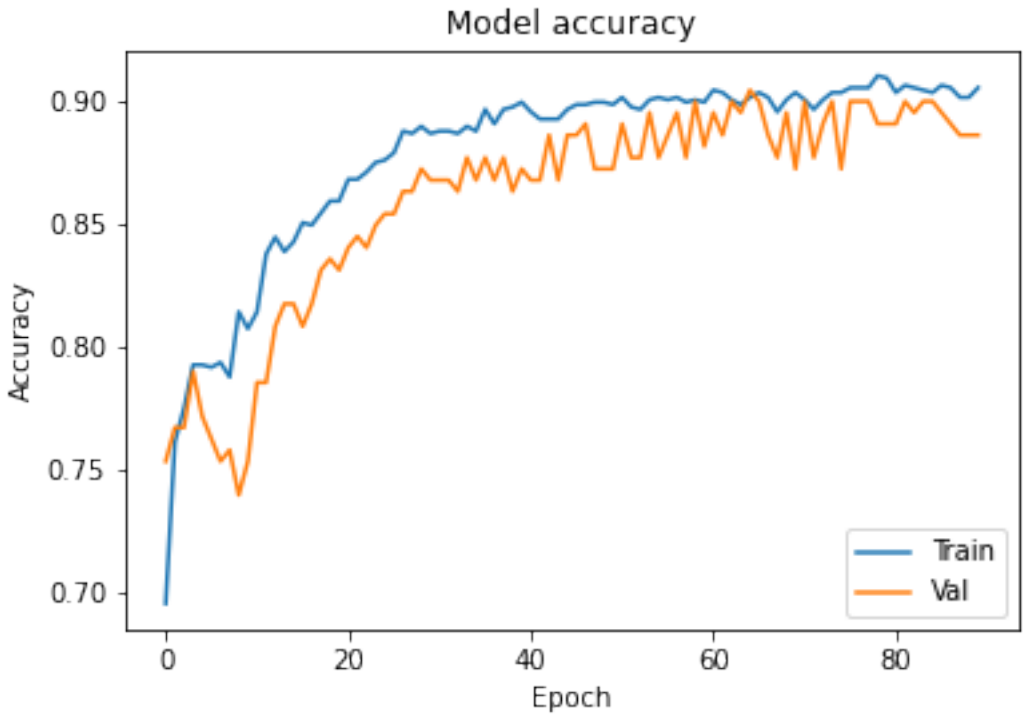

Figure 5: Plot of Model Accuracy on training and validation set with each epoch

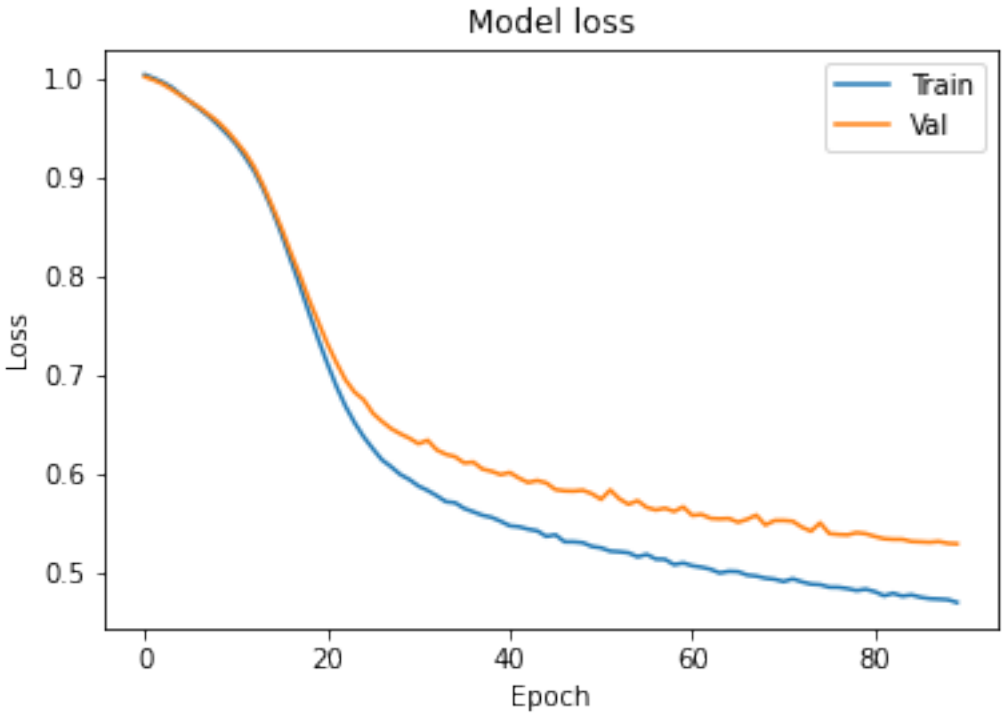

Figure 6: Plot of Model Loss on training and validation set with each epoch

5.2 L2 Regularization

Lambda (λ)	Validation Accuracy
0.0001	89.04
0.001	90.87
0.01	90.87
0.1	51.14
1	51.14

- We observe that the model's performance on validation and training set increases on using L2 regularization with $\lambda = 0.01$.

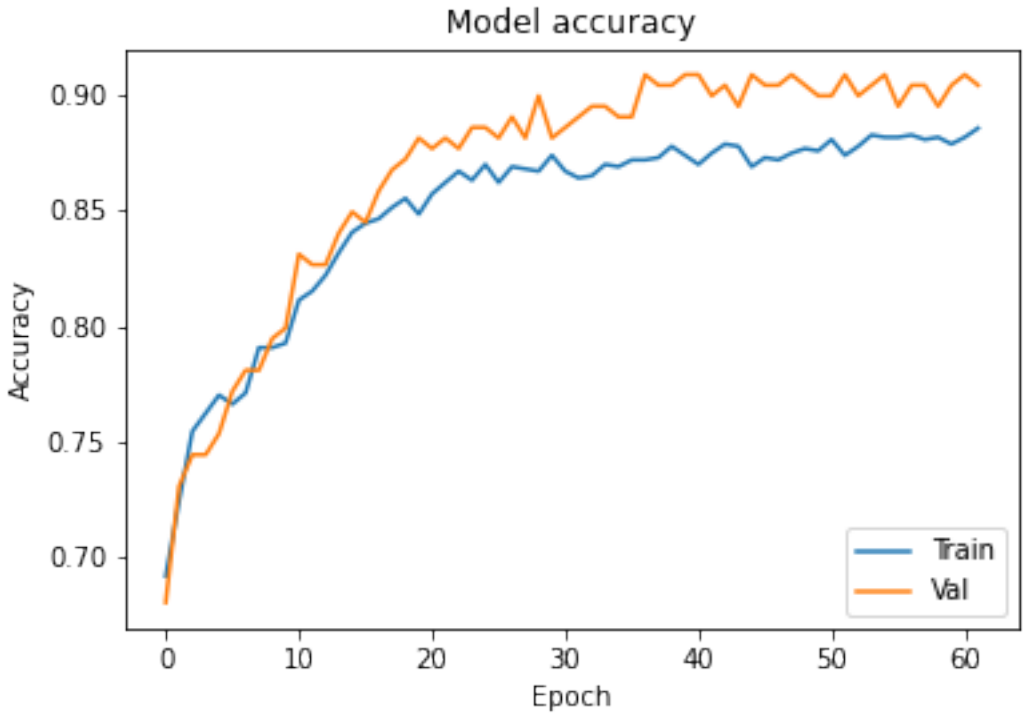

Figure 7: Plot of Model Accuracy with each epoch for $\lambda = 0.01$

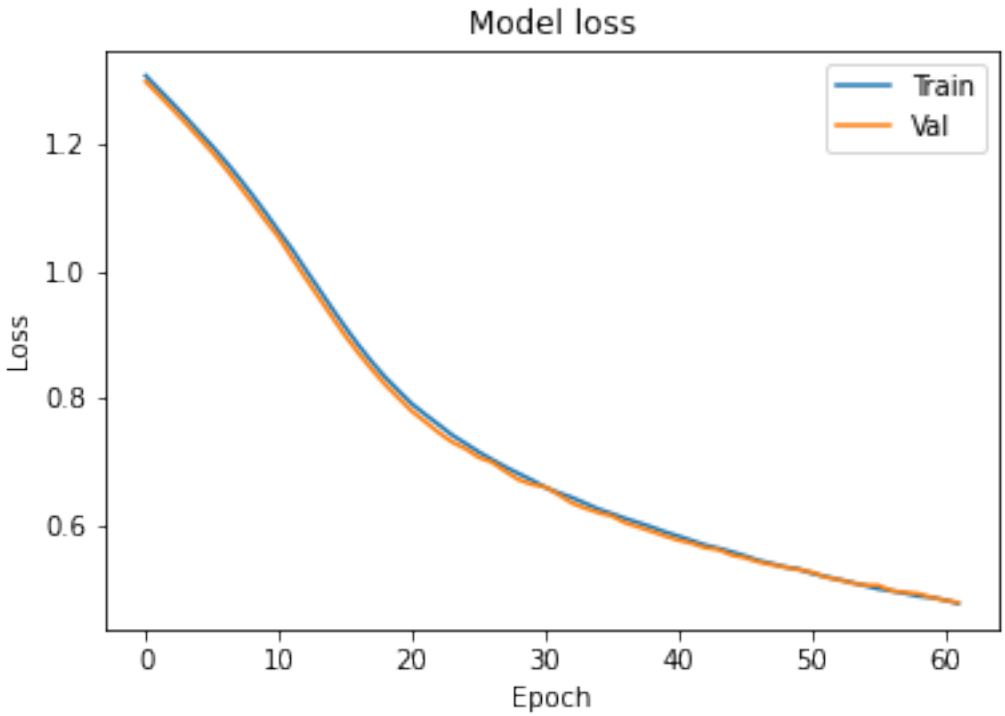

Figure 8: Plot of Model Loss with each epoch for $\lambda = 0.01$

6 Optimum Topology of the Network Architecture

Network Topology	Validation Accuracy
32	88.58
32,32	90.41
32,32,32	90.41
32,32,32,32	91.31
16,32,32,16	91.78

- We see that we get the best accuracy of 91.78 % using 4 Hidden Layers with 16, 32, 32, 16 neurons.

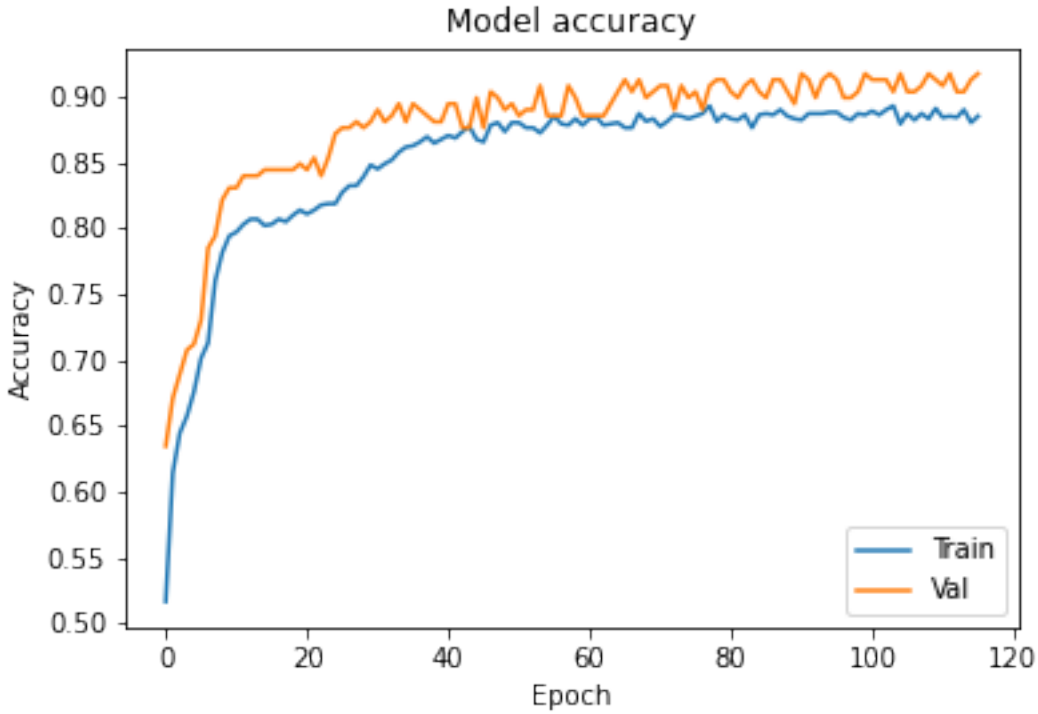

Figure 9: Plot of Model Accuracy with each epoch for topology = 16,32,32,16

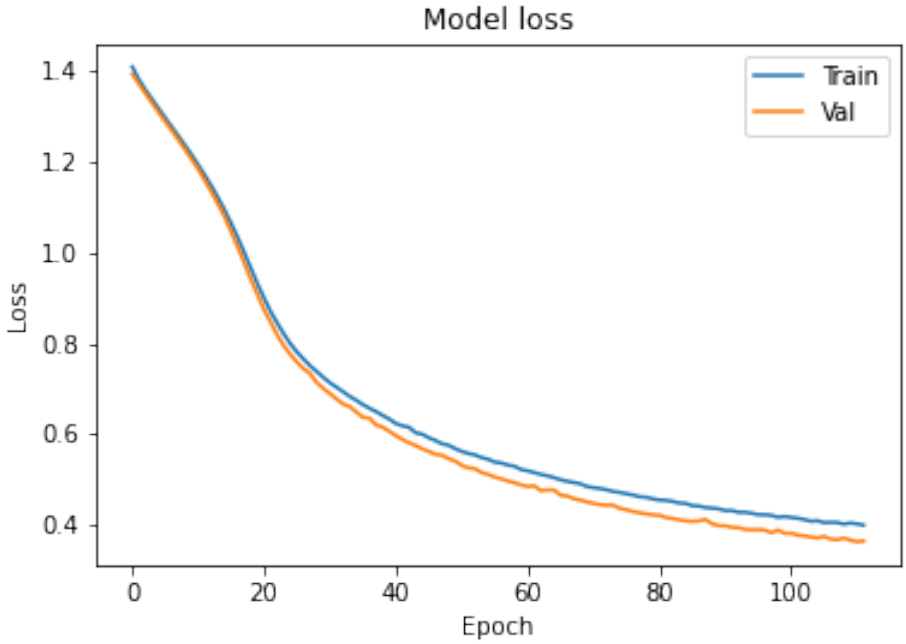

Figure 10: Plot of Model Loss with each epoch for topology = 16,32,32,16

7 *tanh* Activation Function

- Using *tanh* activation function instead of ReLu, we obtained an accuracy of 87.21 % on validation set.
- Hence, accuracy doesn't improve. We'll use 'ReLu' Activation Function only for our optimal model.

8 Optimum Momentum Parameter

Momentum Parameter	Validation Accuracy
0.5	91.78
0.6	91.32
0.7	90.87
0.8	90.87
0.9	90.87

- We observe that we get highest accuracy of 91.78 % on using *momentum* = 0.5.

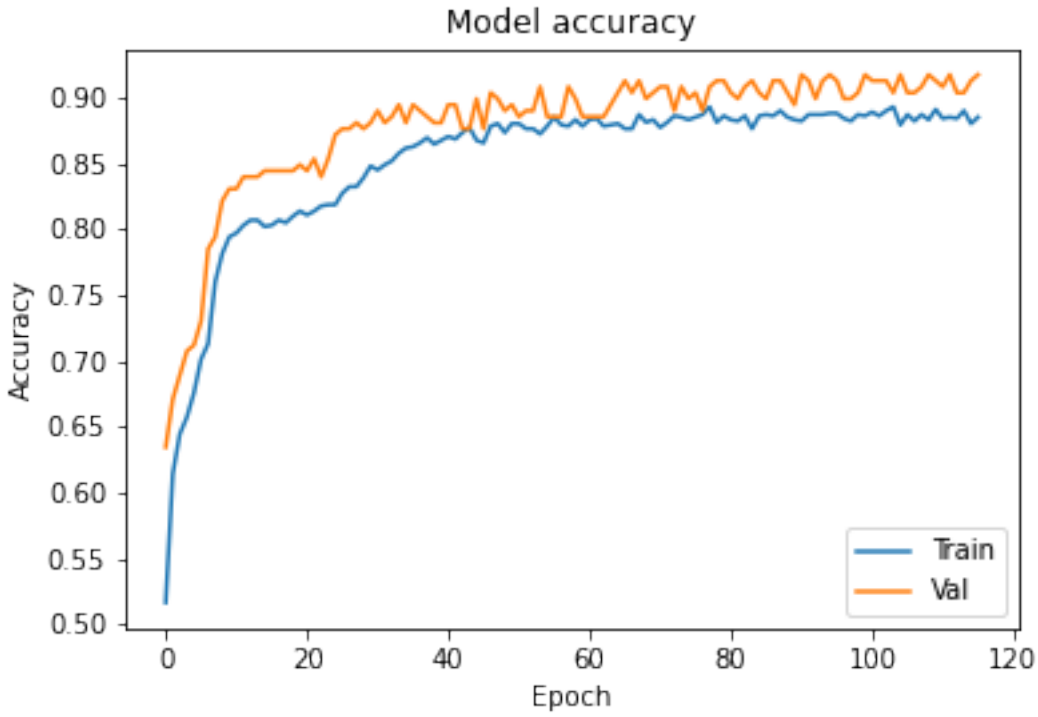

Figure 11: Plot of Model Accuracy with each epoch for momentum = 0.5

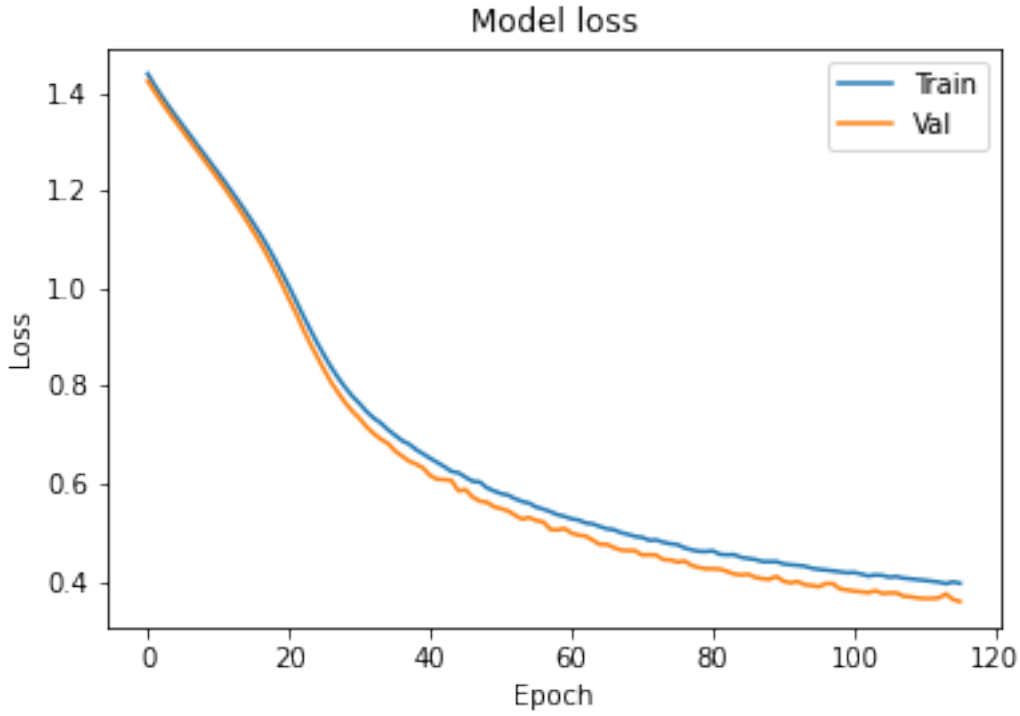

Figure 12: Plot of Model Loss with each epoch for momentum = 0.5

9 Most Optimal Neural Network

- Parameters for our most optimal model :
  - Network Topology : 4 Hidden Layers with 16, 32, 32, 16 neurons each and output layer with 1 neuron.
  - Activation function = ReLu for hidden layers and 'Sigmoid' activation function for output layer.
  - Initial learning rate = 0.01
  - Learning rate update schedule : Exponential decay with $decay\_rate = 0.005$
  - L2 regularization is used with $\lambda = 0.01$
  - *Momentum* parameter = 0.5
- Highest Test Accuracy achieved = 91.32 %

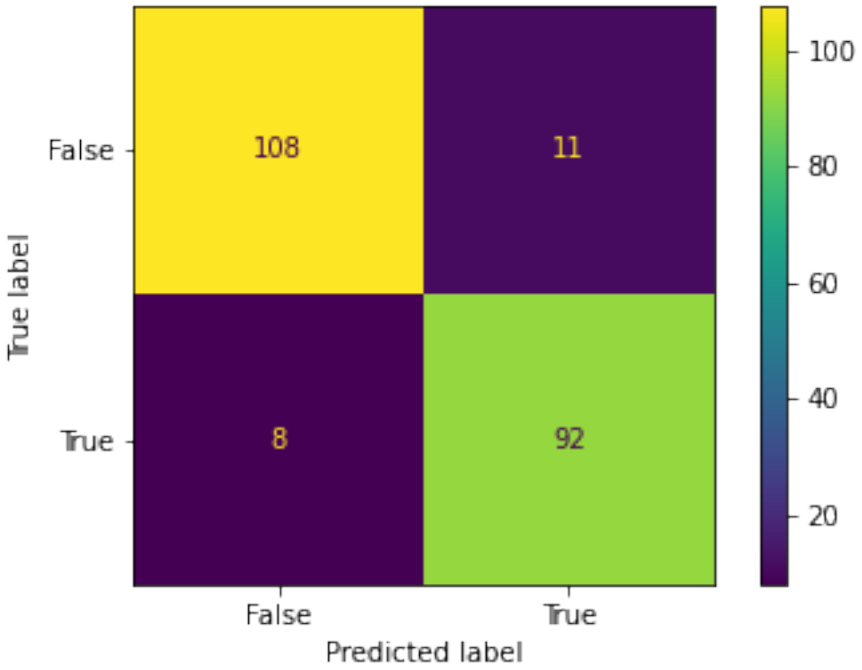

Figure 13: Confusion Matrix for our most optimal neural network architecture