Deep Learning Spring 2020
Assignment 1 Report

Ali Khalid
MSDS21001

April 24, 2022

1 Task # 1

1.1 Computational graph

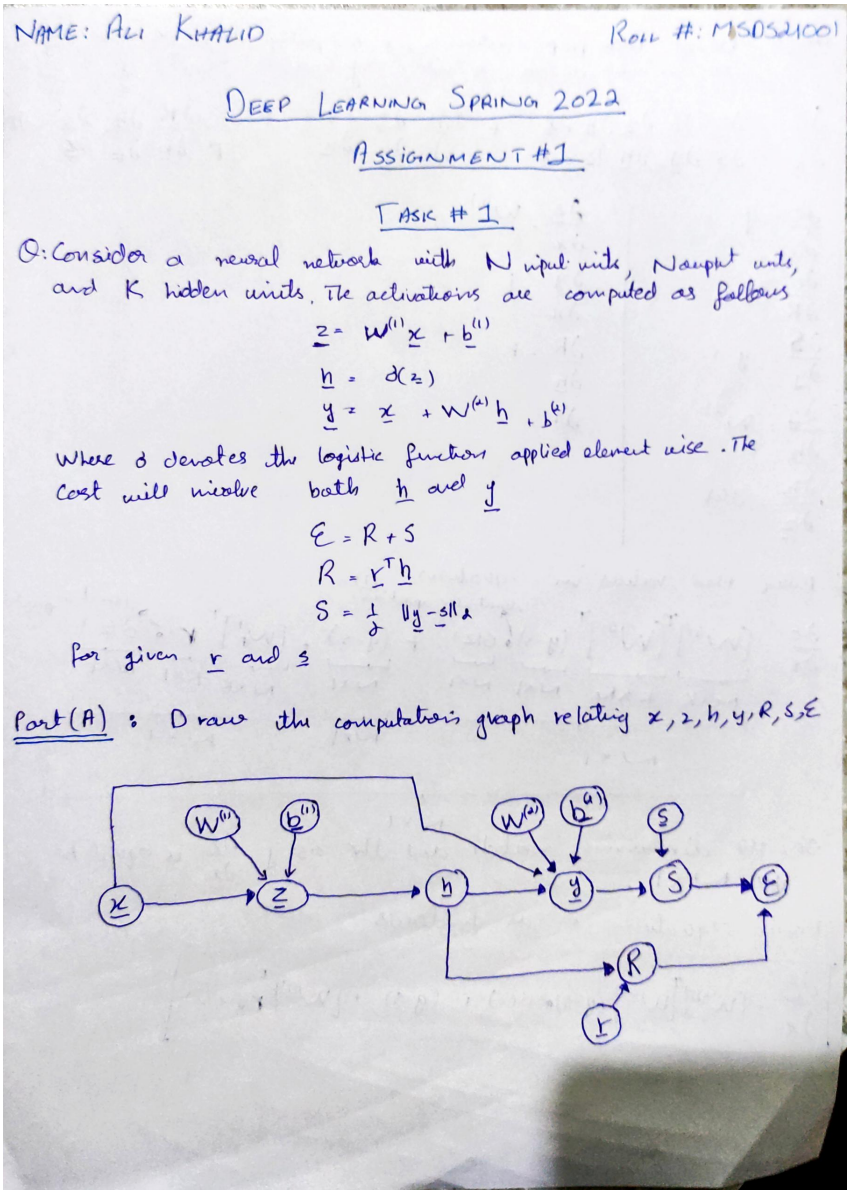

Figure 1: Computational graph

1.2 Mathematical Derivation

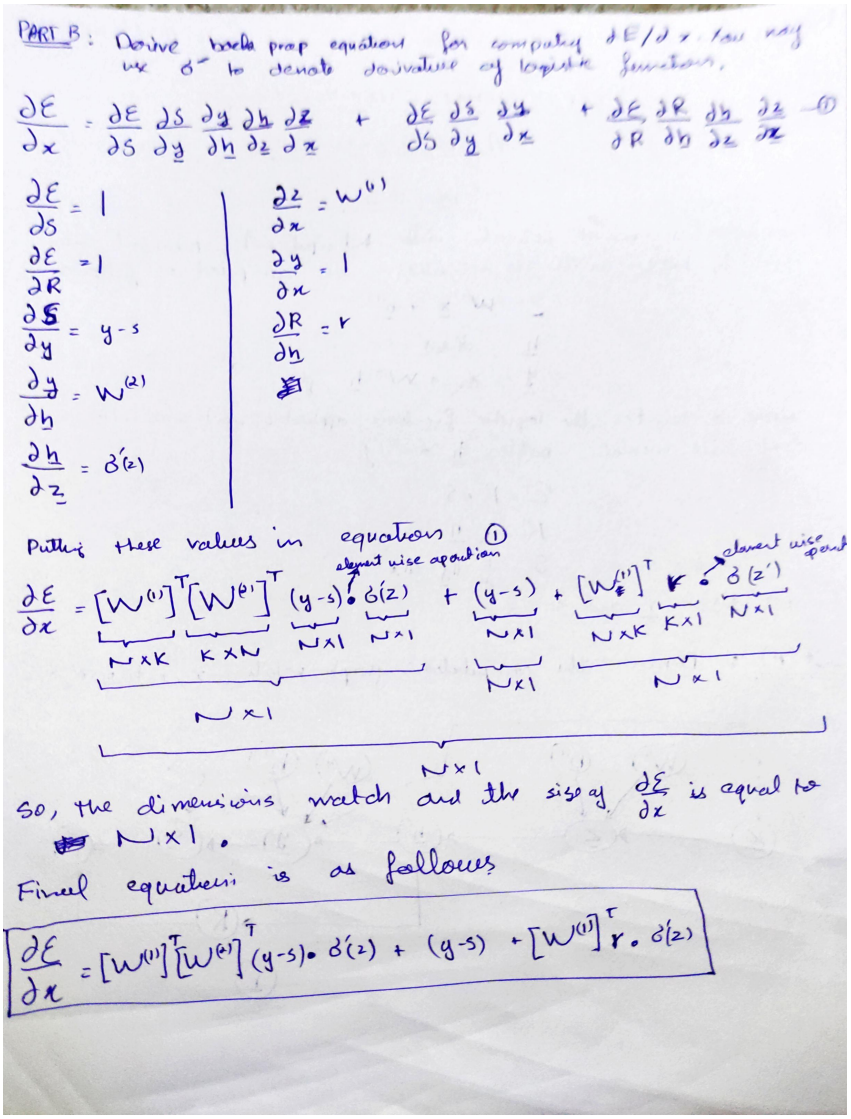

Figure 2: Mathematical Derivation

2 Task # 2

2.1 Loss and accuracy curves

2.1.1 sigmoid

Epochs = 10000
Learning rate = 0.1
Loss Curve

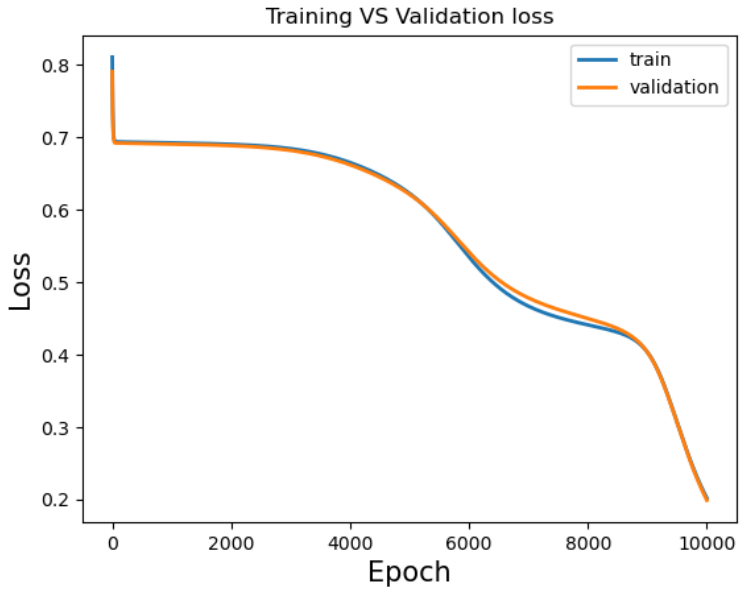

Figure 3: sigmoid loss curve for training and validation data

Accuracy Curve

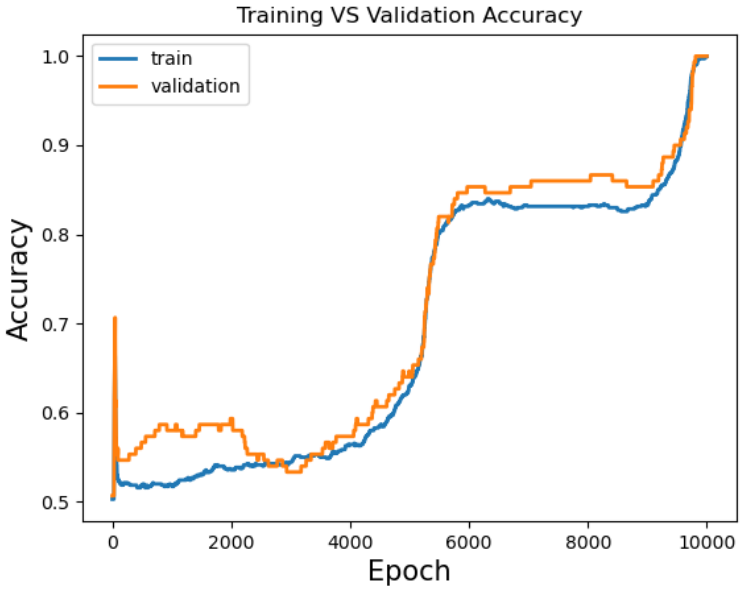

Figure 4: sigmoid accuracy curve for training and validation data

2.1.2 tanh

Epochs = 4000
Learning rate = 0.1
Loss Curve

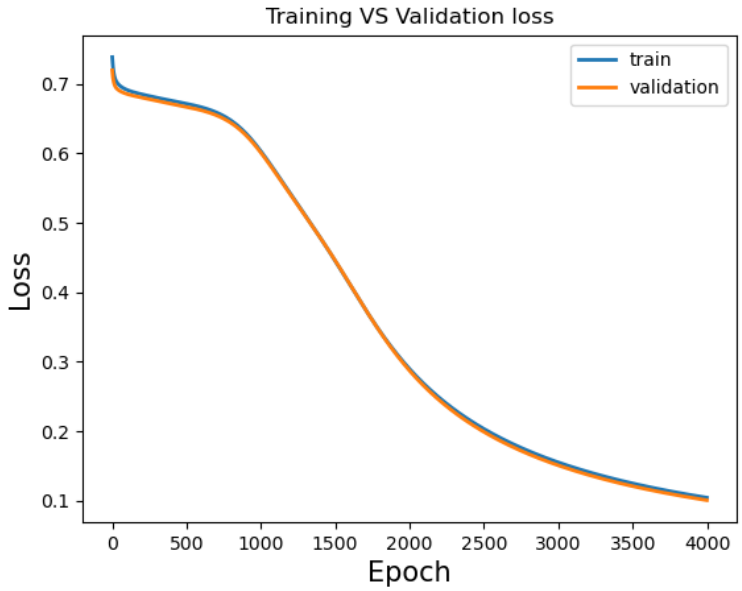

Figure 5: tanh loss curve for training and validation data

Accuracy Curve

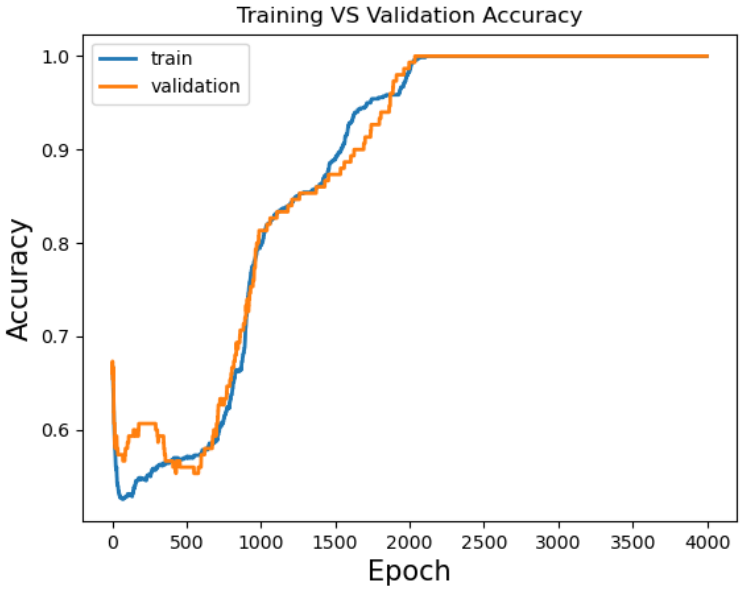

Figure 6: tanh accuracy curve for training and validation data

2.1.3 relu

Epochs = 2000
Learning rate = 0.1

Loss Curve

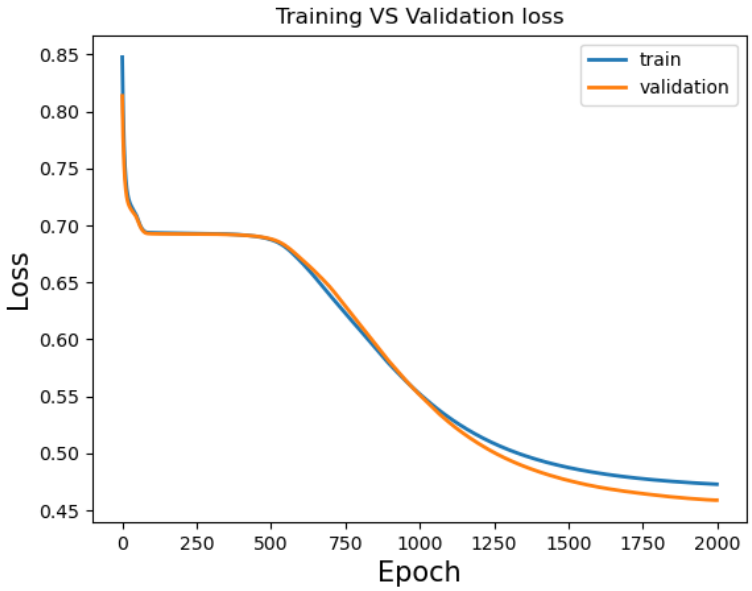

Figure 7: relu loss curve for training and validation data

Accuracy Curve

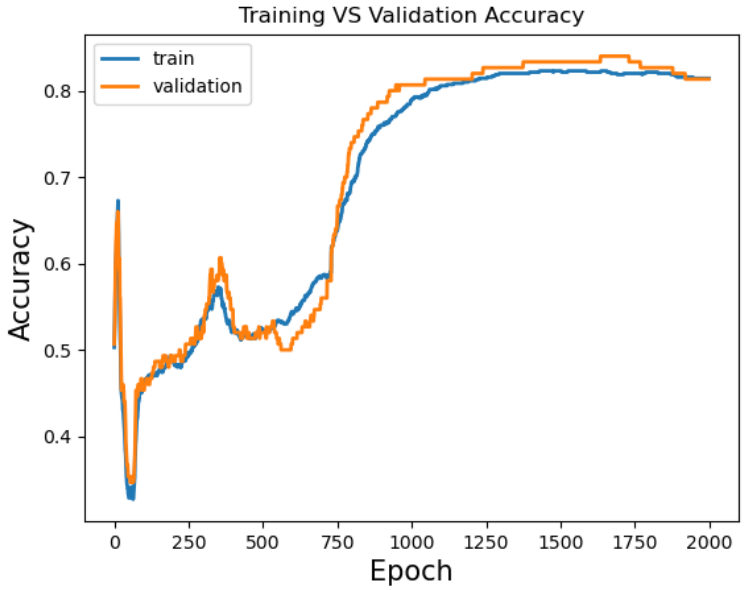

Figure 8: relu accuracy curve for training and validation data

2.2 Test Accuracy

2.2.1 sigmoid

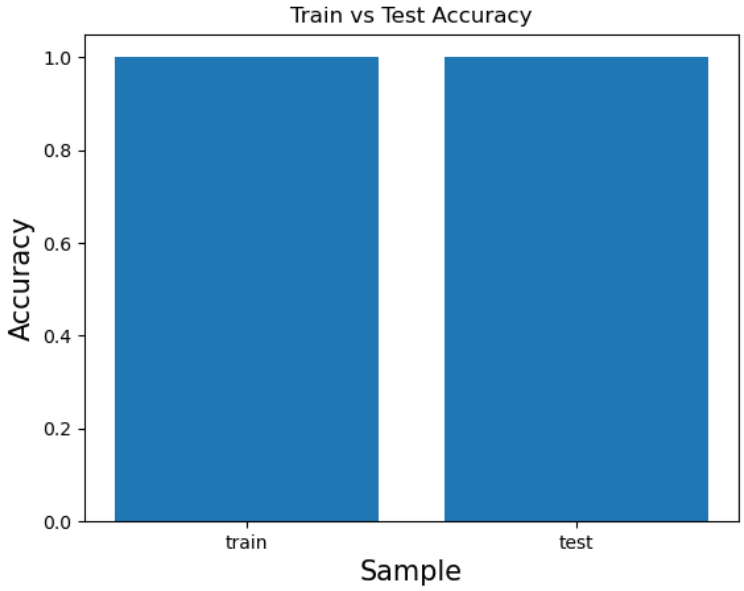

Figure 9: Training vs Test accuracy for sigmoid

2.2.2 tanh

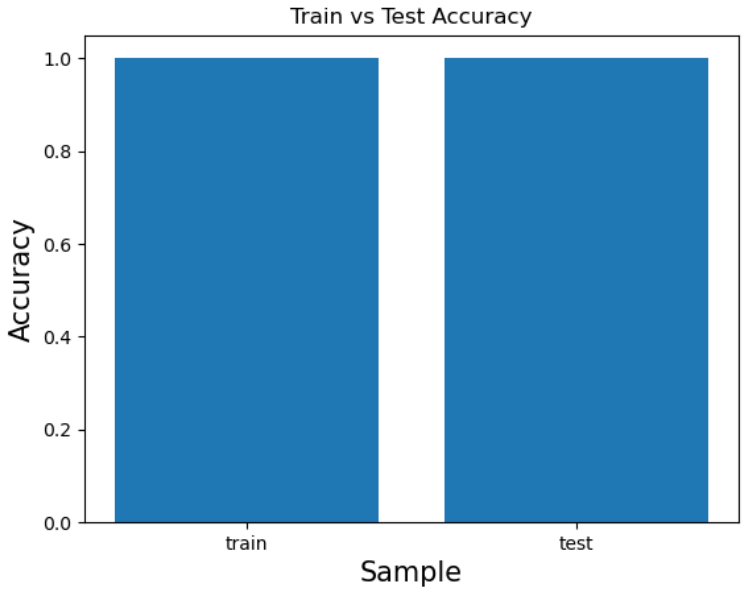

Figure 10: Training vs Test accuracy for tanh

2.2.3 relu

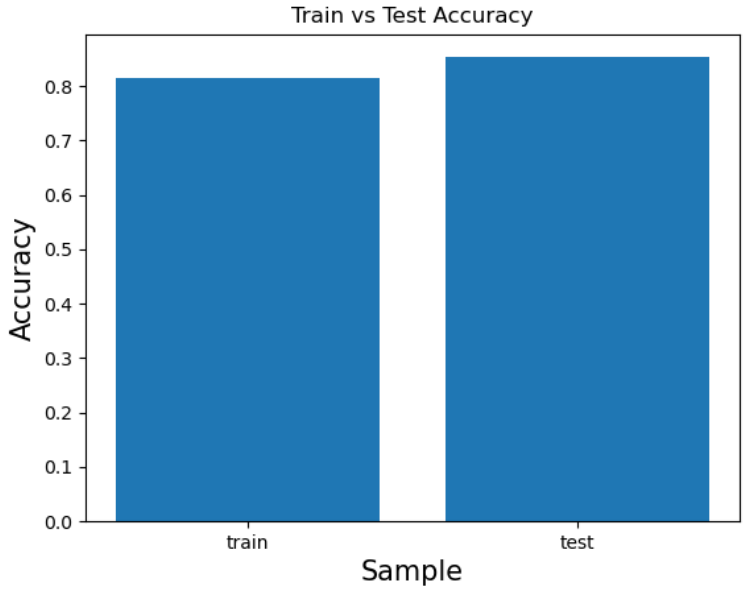

Figure 11: Training vs Test accuracy for relu

2.3 Analysis(in different experiments)

2.3.1 sigmoid

The training and validation loss were decreasing very slowly after 1000 epochs, giving the notion that model has reached the minimum loss. But, on further increasing the epochs (keeping learning rate same) , the loss decrease drastically. Even after 10000 epochs the loss continues to decrease for both validation and test set. As the accuracy reaches the maximum value after at around 10000 epochs, the experiment was terminated and model was saved.

The change in learning rate, has significant impact on loss and accuracy. With small learning rate, the loss decreases very slowly and hence requires more number of epochs. But, with small value of learning rate, the curves (loss, accuracy) becomes more smothers.

2.3.2 tanh

The training and validation loss were decreasing very slowly initially but after 1000 epochs the loss decreases drastically. Using tanh instead of sigmoid, I achieved mthe same accuracy with less number of epochs. The epochs decreases from 10000 to 4000.

The change in learning rate, has significant impact on loss and accuracy. With small learning rate, the loss decreases very slowly and hence requires more number of epochs. But, with small value of learning rate, the curves (loss, accuracy) becomes more smothers. .

2.3.3 relu

Relu has better convergence than tanh and sigmoid. It reaches the maximum accuracy uin only 2000 epochs, where sigmoid and tanh took 10000 and 4000 epochs, respectively. One thing to note is that, using relu I wasnt able to reach 100 percent accuracy. Maybe with more experimentation, a person can find the perfect combination of epochs and learning rate to reach 100 percent accuracy. The initialization of weights also perform a key role in determine the accuracy of model.

3 Task # 3

3.1 Loss and accuracy curves with and without mean subtraction

Loss and accuracy curves are shown for epochs = 2000 and learning rate= 0.1

3.1.1 with mean subtraction

Loss curves

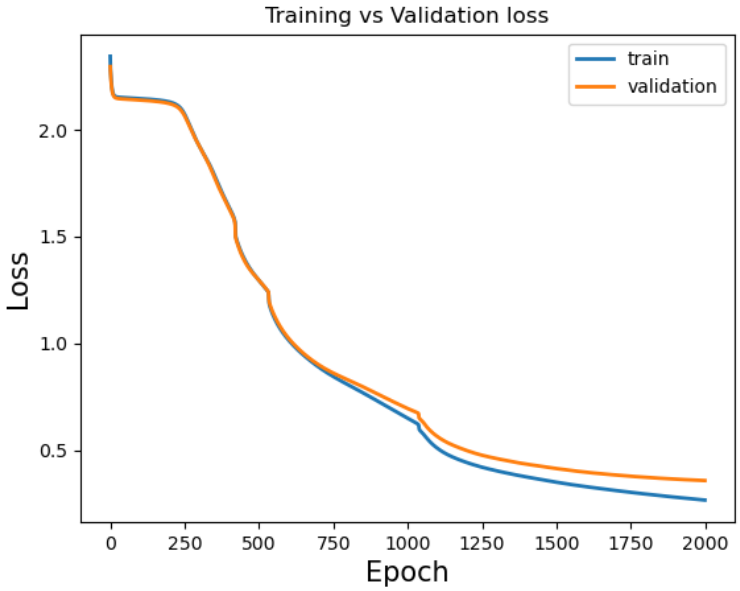

Figure 12: Training vs validation loss with mean subtraction

Accuracy curves

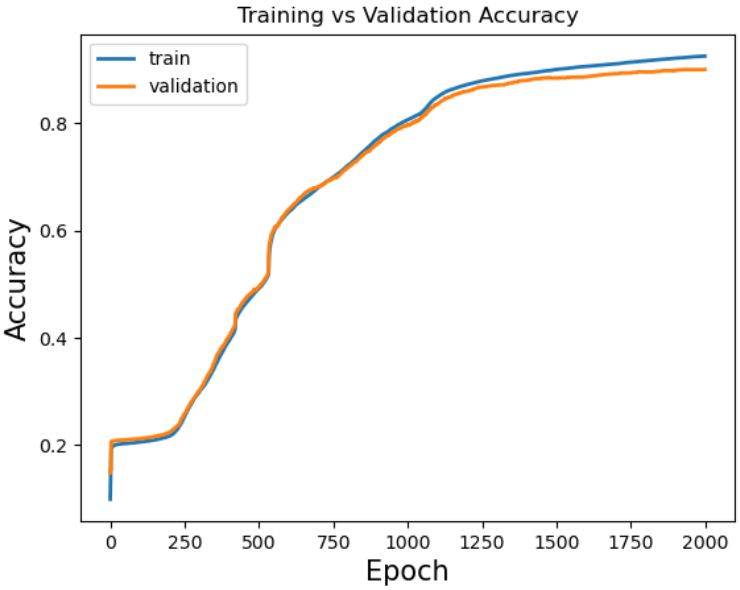

Figure 13: Training vs validation accuracy with mean subtraction

3.2 Without mean subtraction

Without subtracting the mean from images, the loss decrease in the beginning and then it remains constant for rest of the epochs. With different learning rate things may change, but for comparison learning rate was kept same for both cases (with mean subtraction and without mean subtraction).

Loss curves

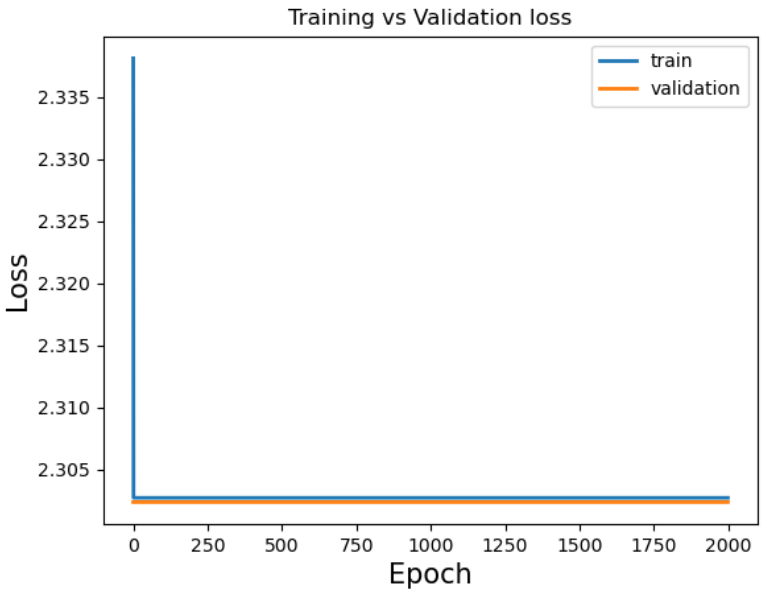

Figure 14: Training vs validation loss without mean subtraction

Accuracy curves

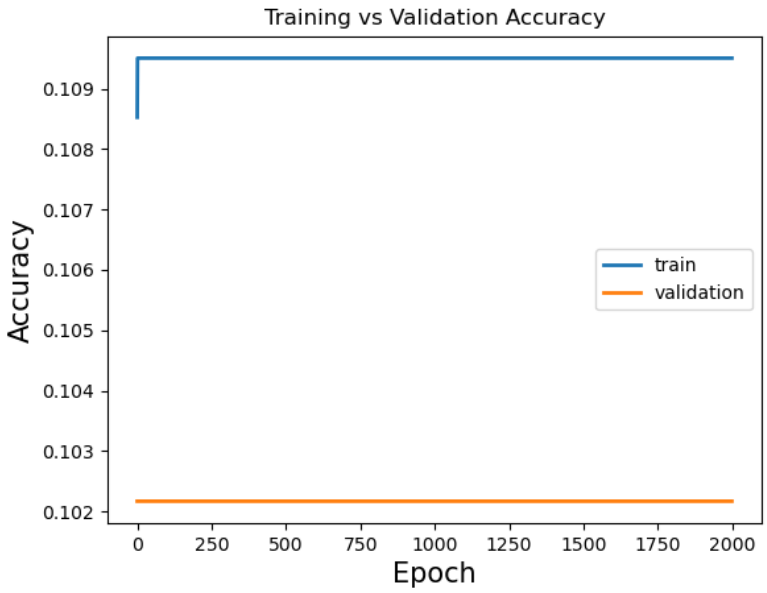

Figure 15: Training vs validation accuracy without mean subtraction

3.3 Train and Test Accuracy

3.3.1 With Mean Subtraction

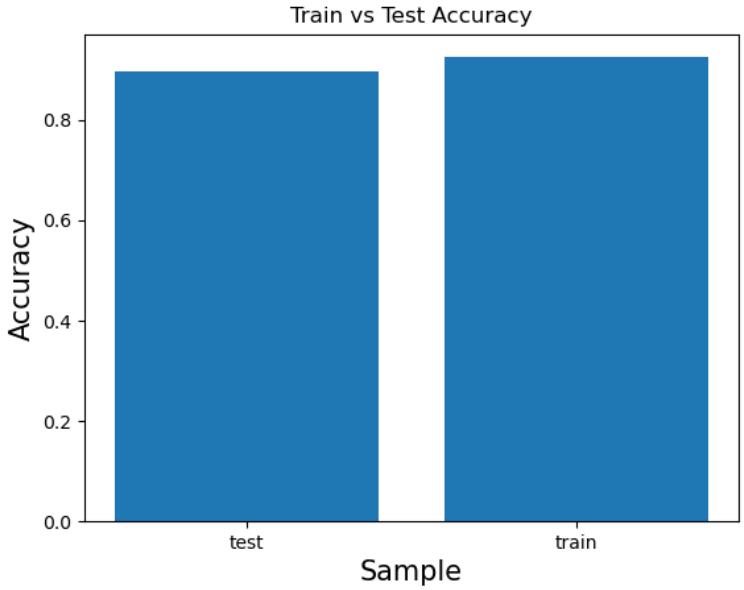

Figure 16: Train and test accuracy with mean subtraction

3.3.2 Without Mean Subtraction

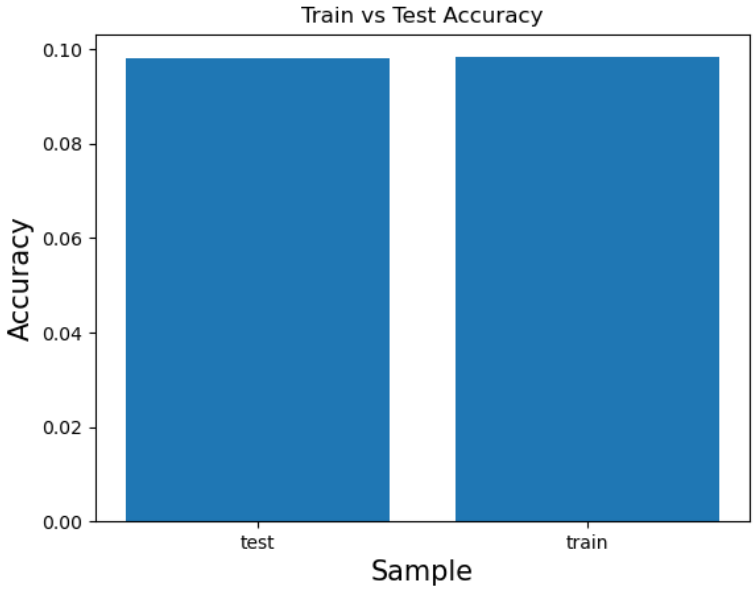

Figure 17: Train and test accuracy without mean subtraction

3.4 Confusion Matrix for Training, Validation and Test set

Confusion matrix are shown for epochs =2000 and learning rate =0.01 using mean subtraction.

3.4.1 Training Data

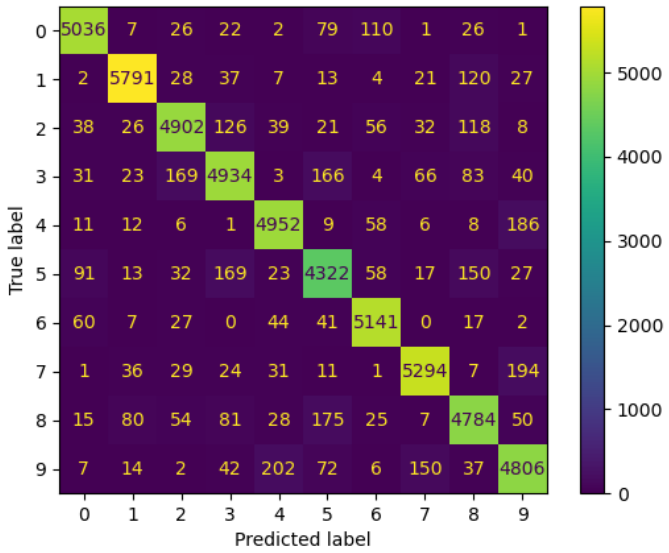

Figure 18: Confusion matrix for train data

3.4.2 Validation Data

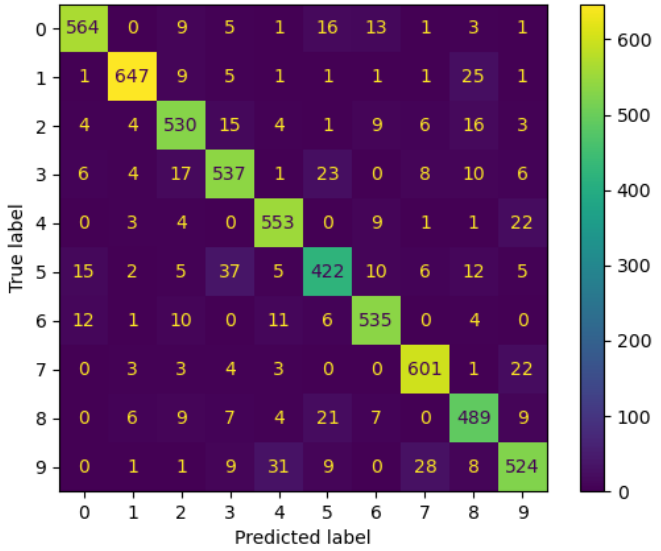

Figure 19: Confusion matrix for validation data

3.4.3 Testing Data

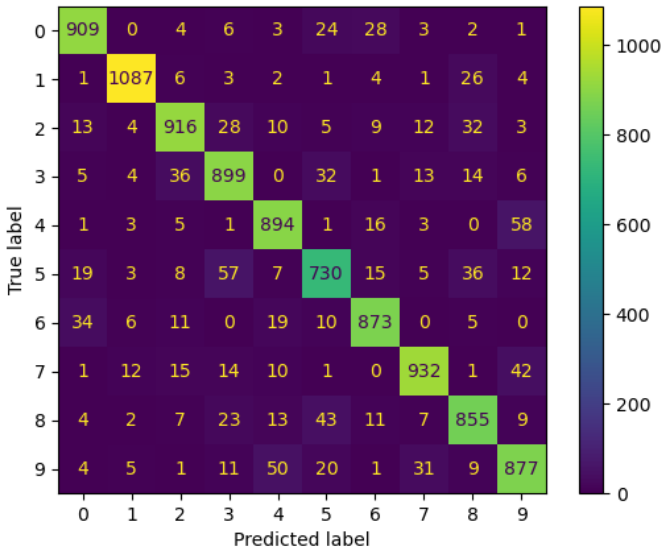

Figure 20: Confusion matrix for test data

3.5 t_SNE plot

Confusion matrix are shown for epochs =1000 and learning rate =0.01 using mean subtraction.

3.5.1 Network with 2 hidden layers

t_SNE plot of hidden layer 1

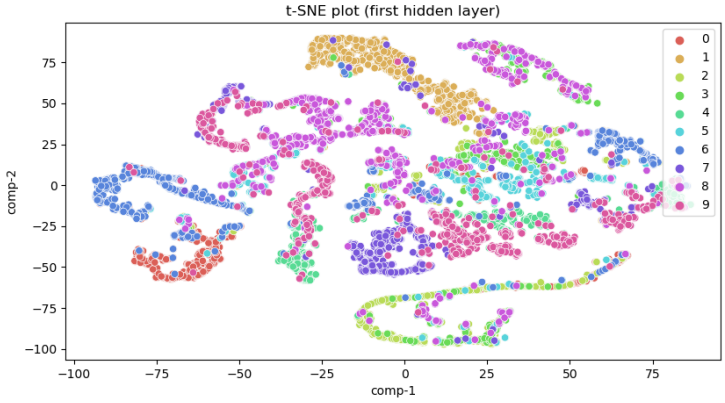

Figure 21: t_SNE plot using activation map of first hidden layer

t_SNE plot of hidden layer 2

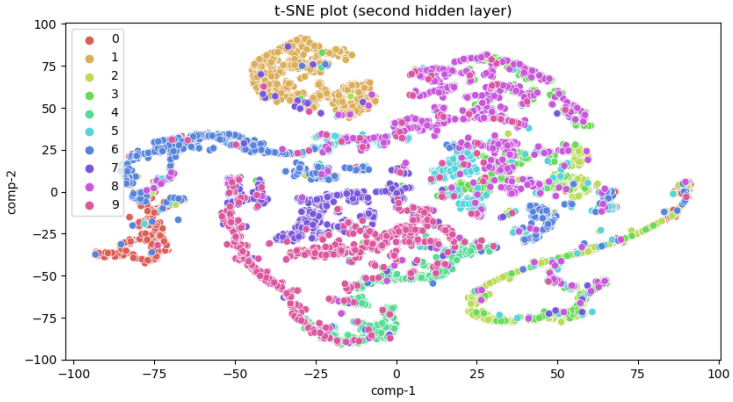

Figure 22: t_SNE plot using activation map of second hidden layer

3.5.2 Network with 3 hidden layers

The t_SNE visualization of 3 hidden layers of network is shown below. The network is trained for only 1000 epochs with learning rate 0.01. The visualizations could have been better if network was trained for more number of epochs. **t_SNE plot of hidden layer 1**

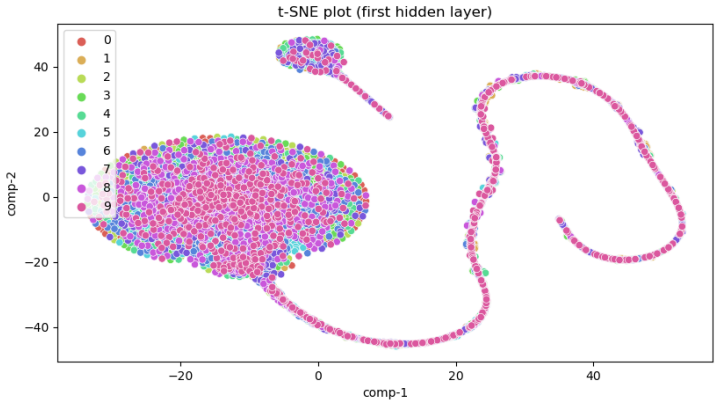

Figure 23: t_SNE plot using activation map of first hidden layer

t_SNE plot of hidden layer 2

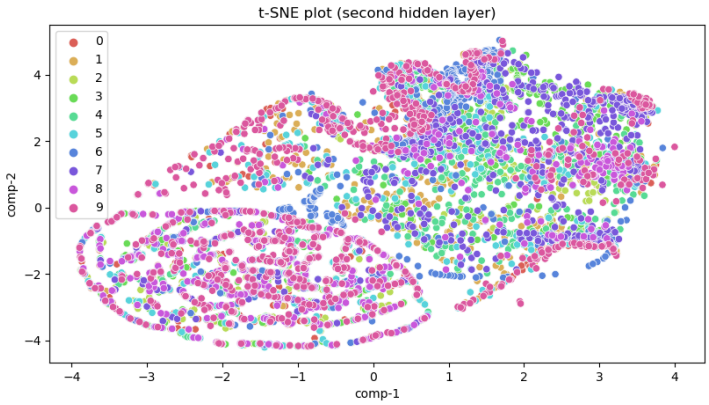

Figure 24: t_SNE plot using activation map of second hidden layer

t_SNE plot of hidden layer 3

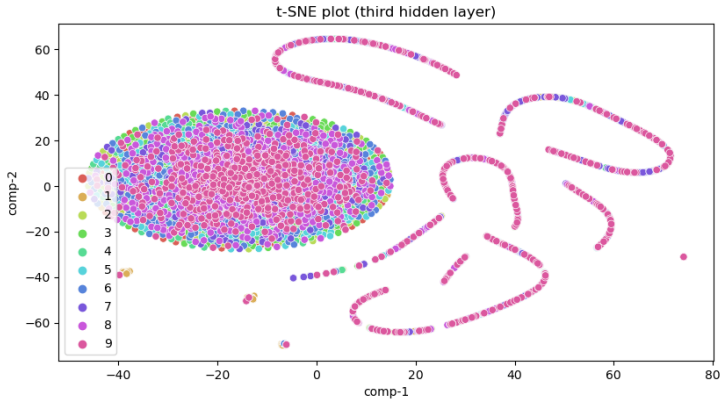

Figure 25: t_SNE plot using activation map of third hidden layer

3.6 Analysis(in different experiments)

3.6.1 Changing Epochs

Below plots show the change in training loss, validation loss loss, training accuracy, validation accuracy and testing accuracy with change in epochs. It also shows the output of t_SNE visualization for second hidden layer. Learning rate is 0.1 for all experiments.

The general trend observed is that with increase in epochs, the loss decreases and accuracy improves as can be shown from the plots below. On the other hand decreasing learning rate makes convergence slow because by reducing learning rate, the change in weights also reduce and hence the process of learning is slowed down.

Epochs = 100

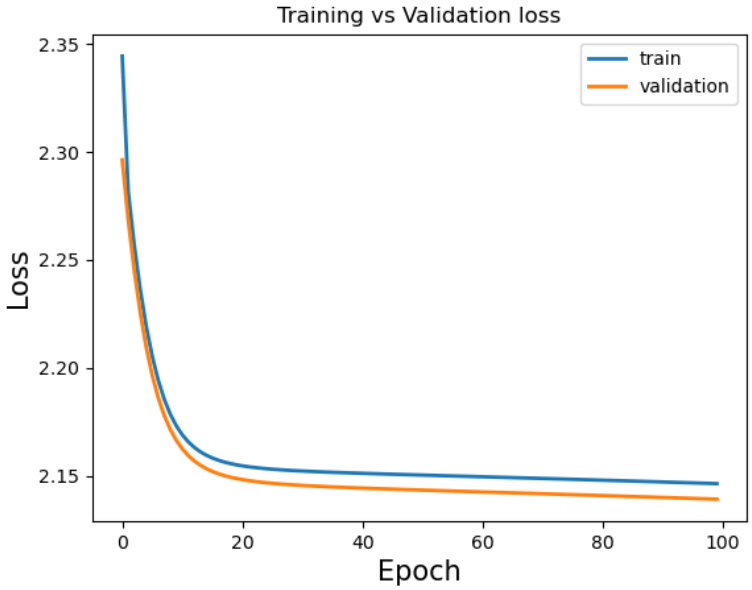

Figure 26: Training and validation loss for epoch = 100

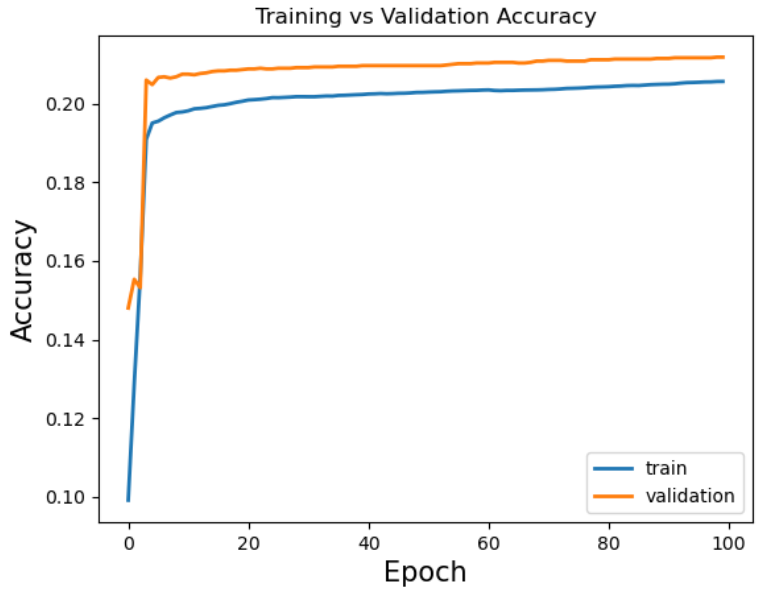

Figure 27: Training and validation accuracy for epoch = 100

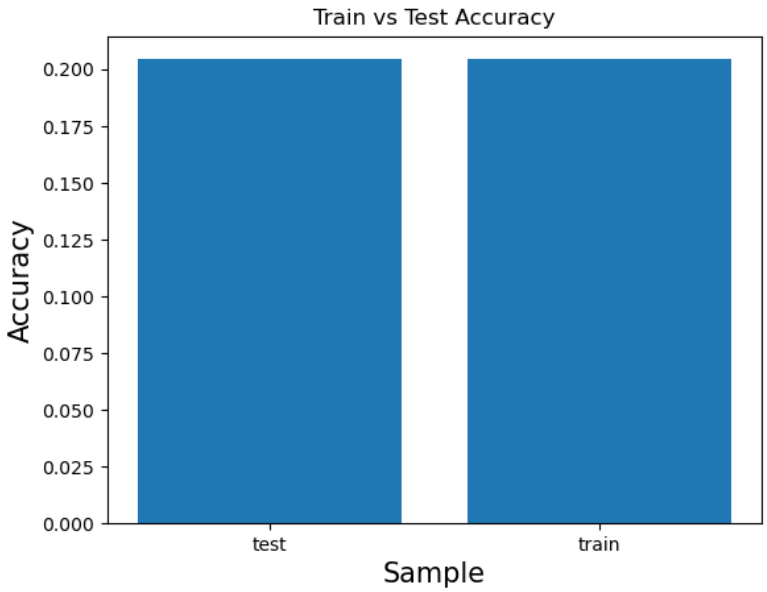

Figure 28: Training vs Testing Accuracy for epoch = 100

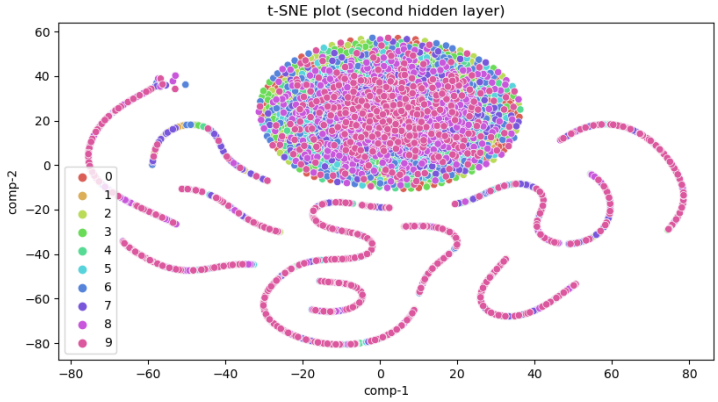

Figure 29: t_SNE plot using activation map of second hidden layer for epoch = 100

Epochs = 500

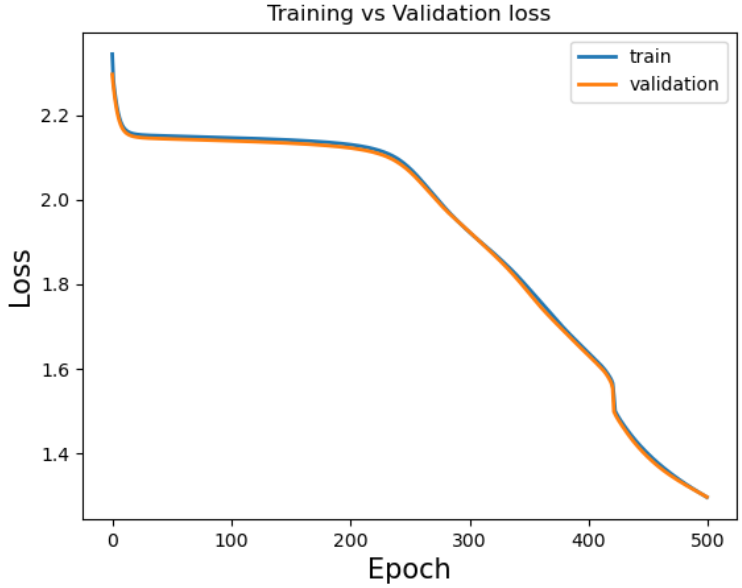

Figure 30: Training and validation loss for epoch = 500

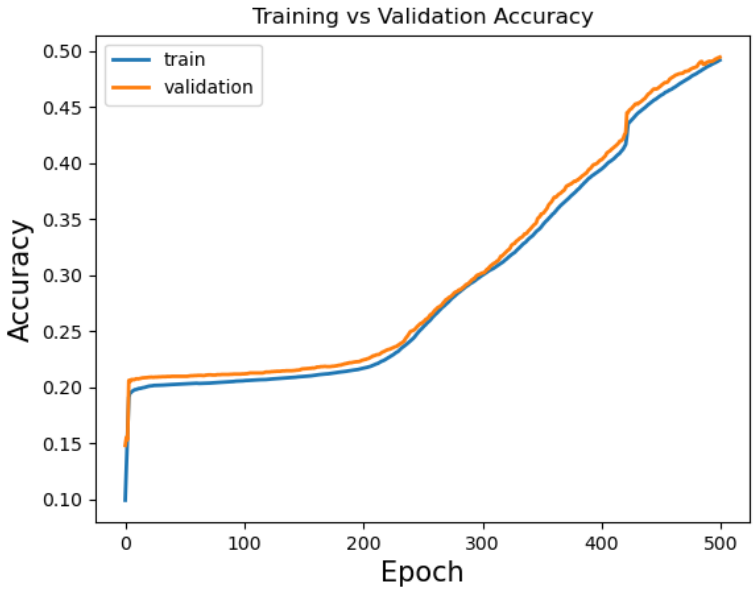

Figure 31: Training and validation accuracy for epoch = 500

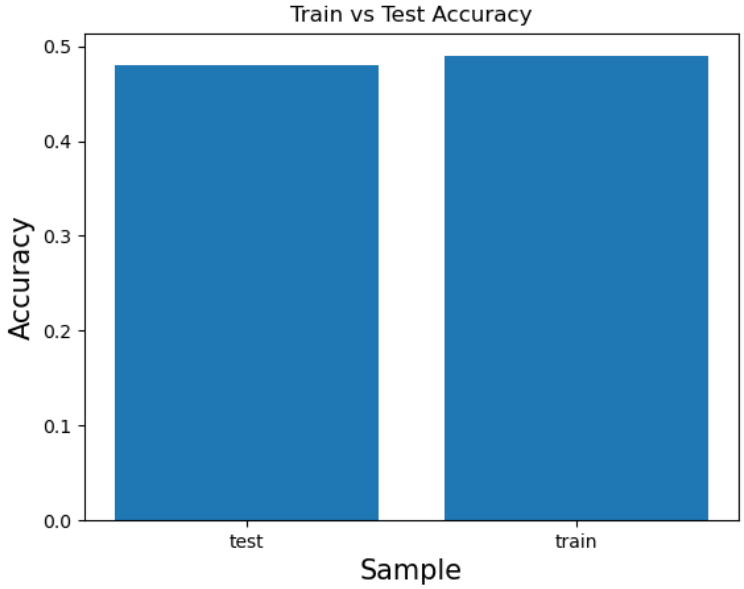

Figure 32: Training vs Testing Accuracy for epoch = 500

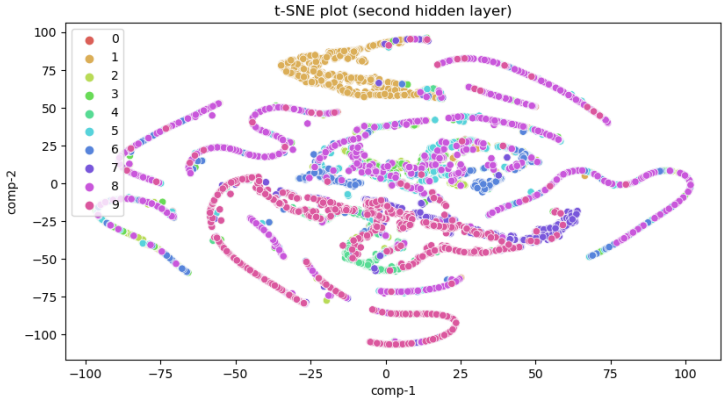

Figure 33: t_SNE plot using activation map of second hidden layer for epoch = 500

Epochs = 1000

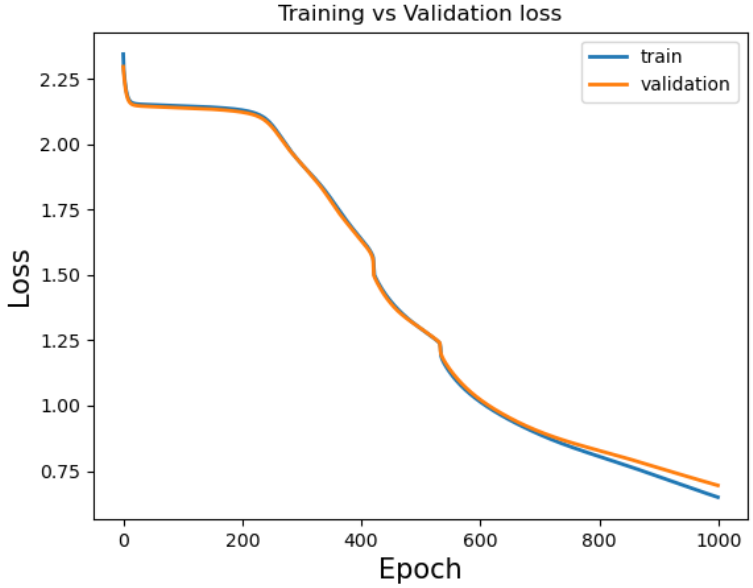

Figure 34: Training and validation loss for epoch = 1000

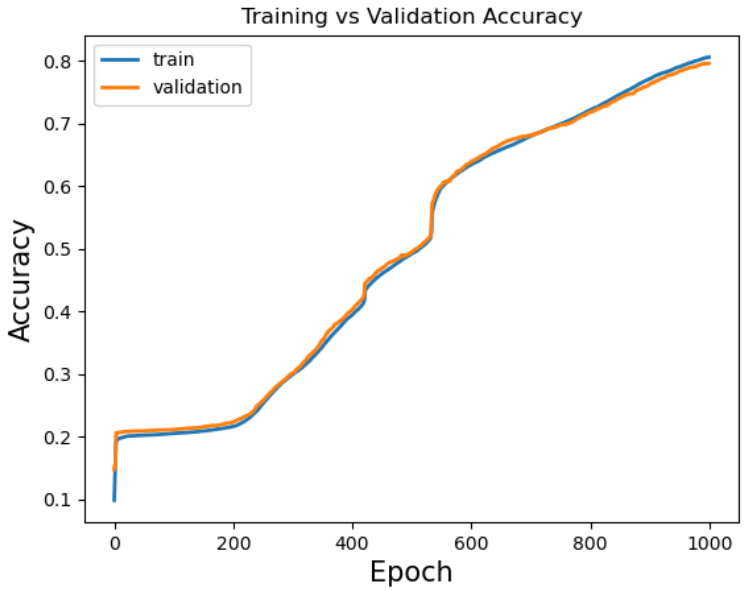

Figure 35: Training and validation accuracy for epoch = 1000

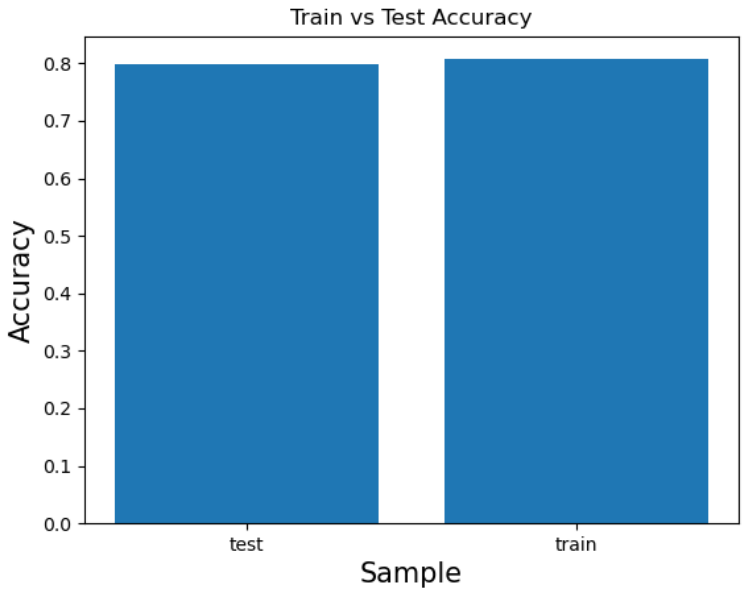

Figure 36: Training vs Testing Accuracy for epoch = 1000

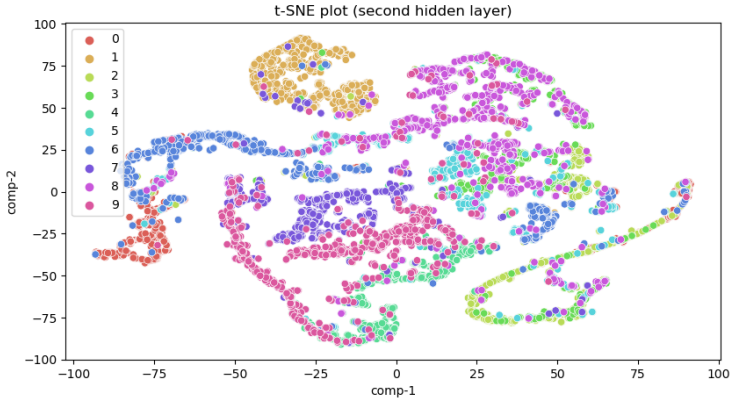

Figure 37: t_SNE plot using activation map of second hidden layer for epoch = 1000

Epochs = 2000

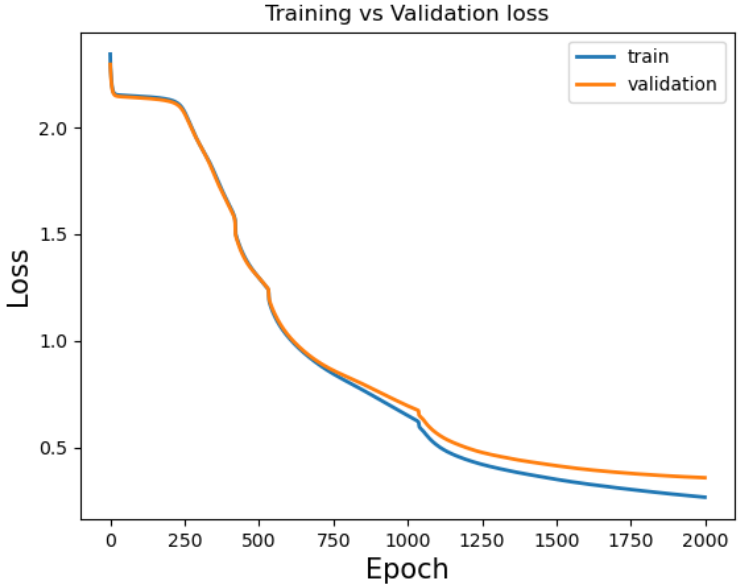

Figure 38: Training and validation loss for epoch = 2000

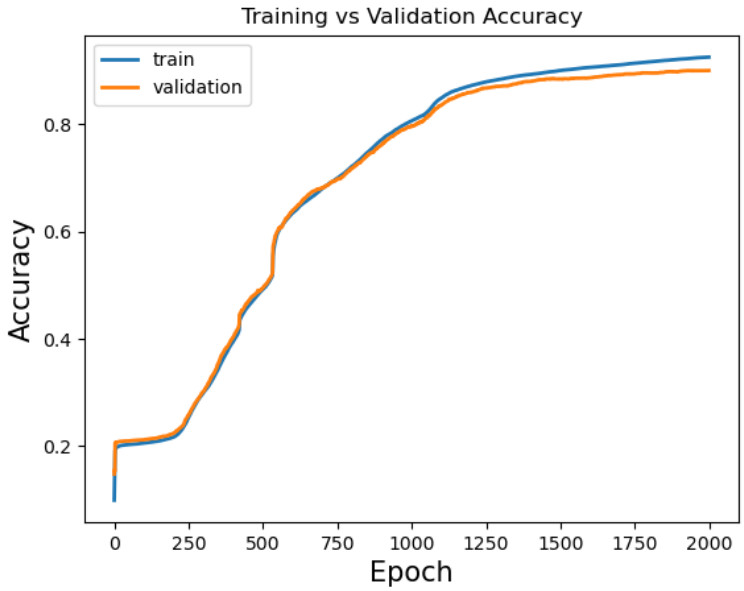

Figure 39: Training and validation accuracy for epoch = 2000

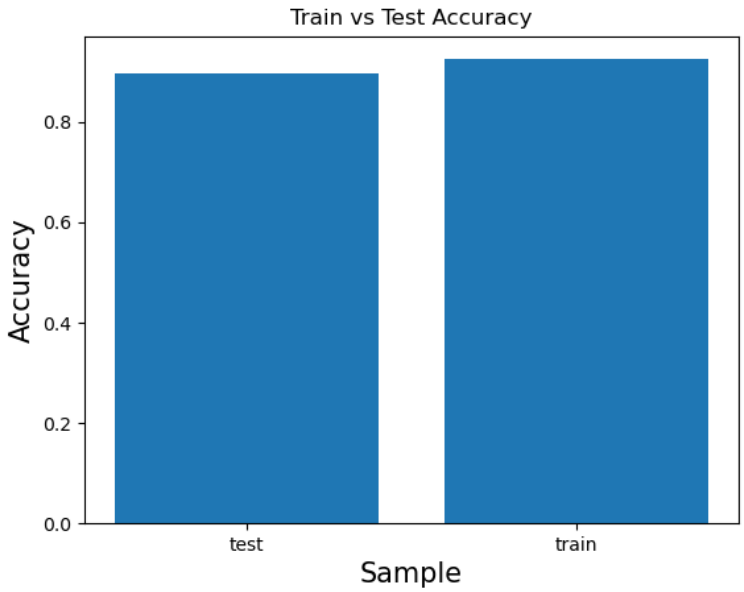

Figure 40: Training vs Testing Accuracy for epoch = 2000

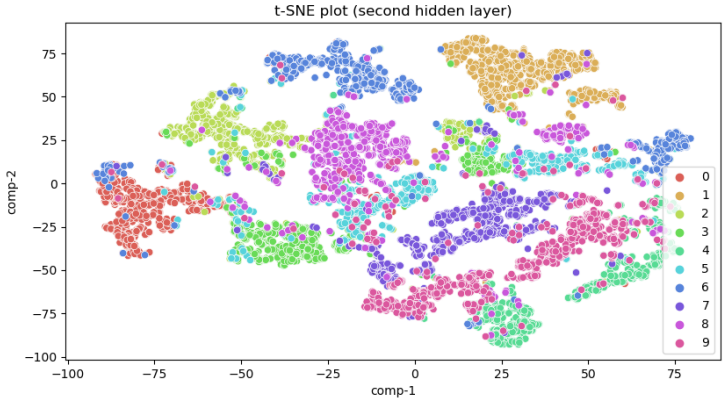

Figure 41: t_SNE plot using activation map of second hidden layer for epoch = 2000

3.6.2 Changing Learning Rate

Below plots show the change in training loss and validation loss for different learning rate and same number of epochs i.e. 100

Learning rate = 0.1

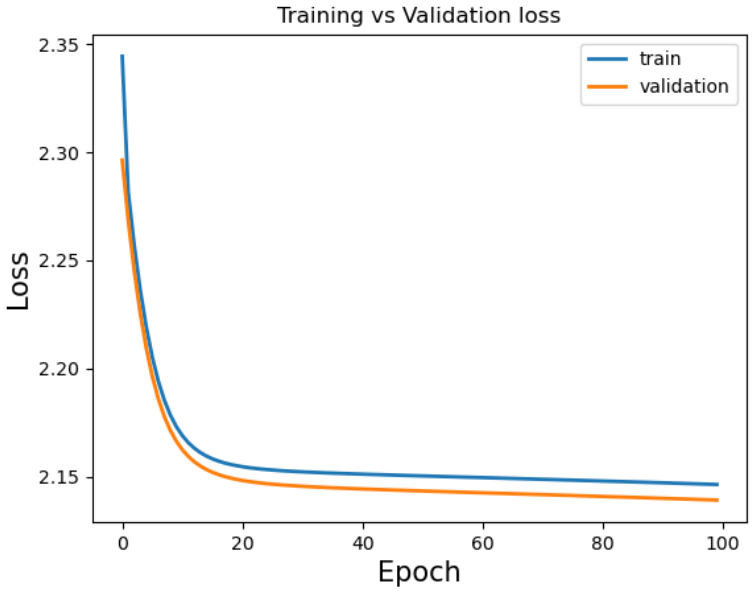

Figure 42: Training and validation loss for lr = 0.1

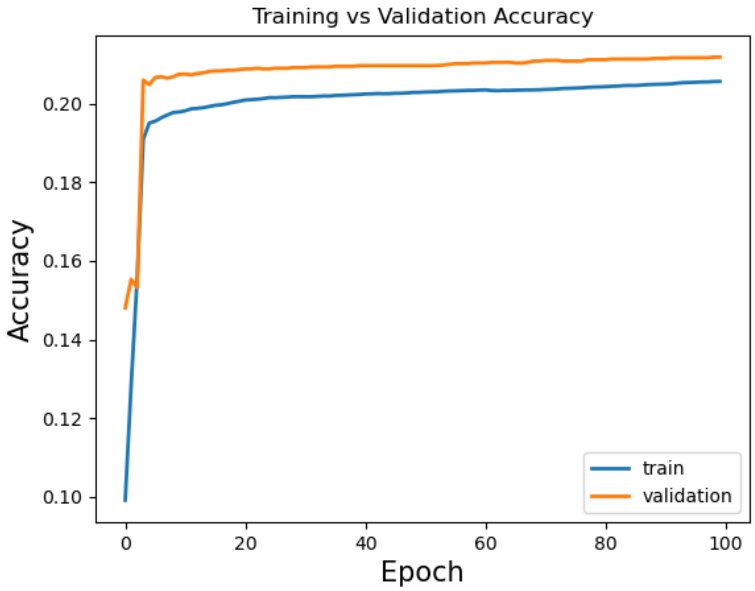

Figure 43: Training and validation accuracy for lr = 0.1

Learning rate = 0.01

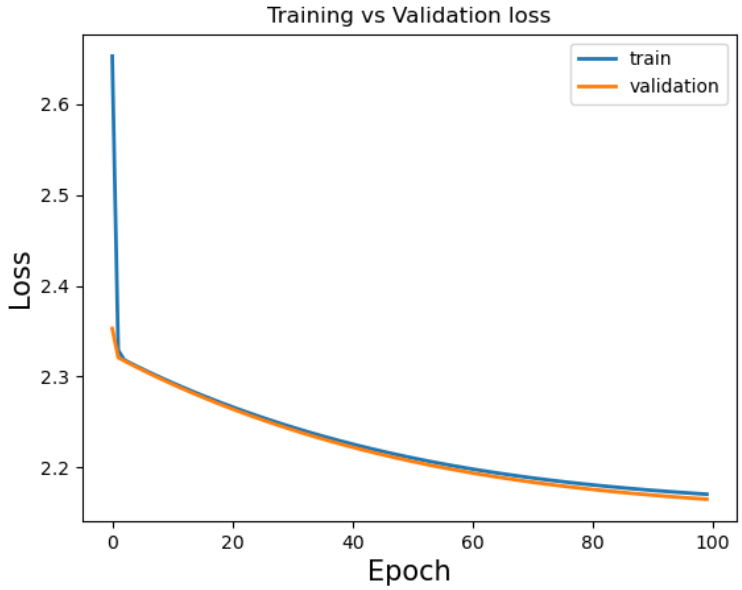

Figure 44: Training and validation loss for lr = 0.01

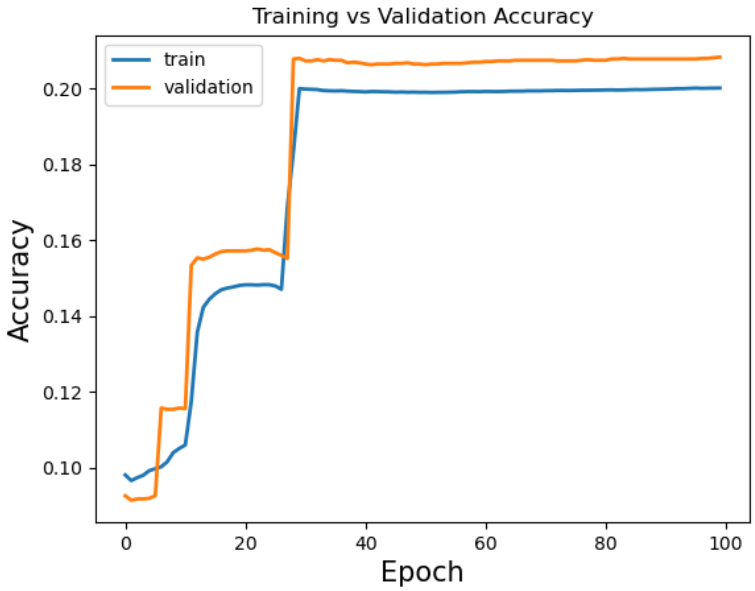

Figure 45: Training and validation accuracy for lr = 0.01

Learning rate = 0.001

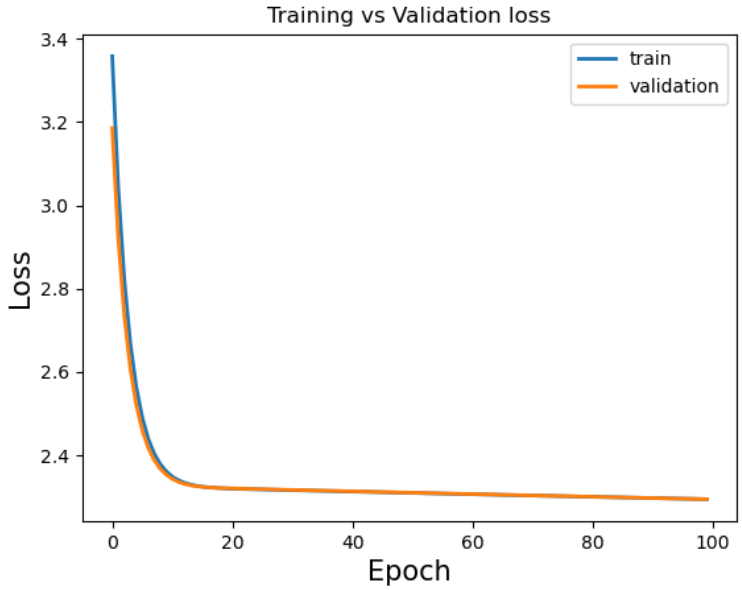

Figure 46: Training and validation loss for lr = 0.001

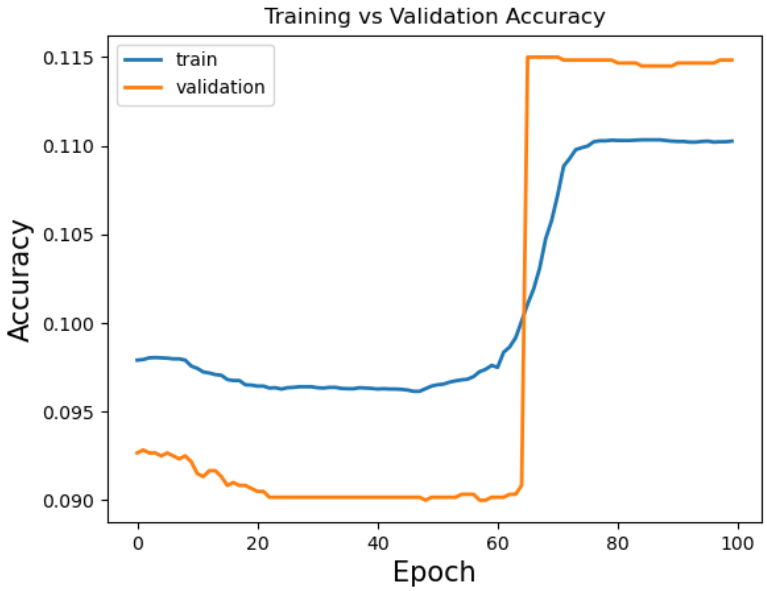

Figure 47: Training and validation accuracy for lr = 0.001

3.7 Conclusion

Following conclusions can be extracted from all the experiments that I have done so far:

1- With the increase in epochs, loss either decreases or stays the same. But it is not always correct, if learning rate is high, loss can increase with increasing epochs.

2- Learning rate has a significant impact on the learning process. Having small learning rate slows down the convergence but is more stable and vice versa.

3- Deeper networks are can result in better performance but are harder to train.