CS3631 – Deep Neural Networks
Assignment – Backpropagation
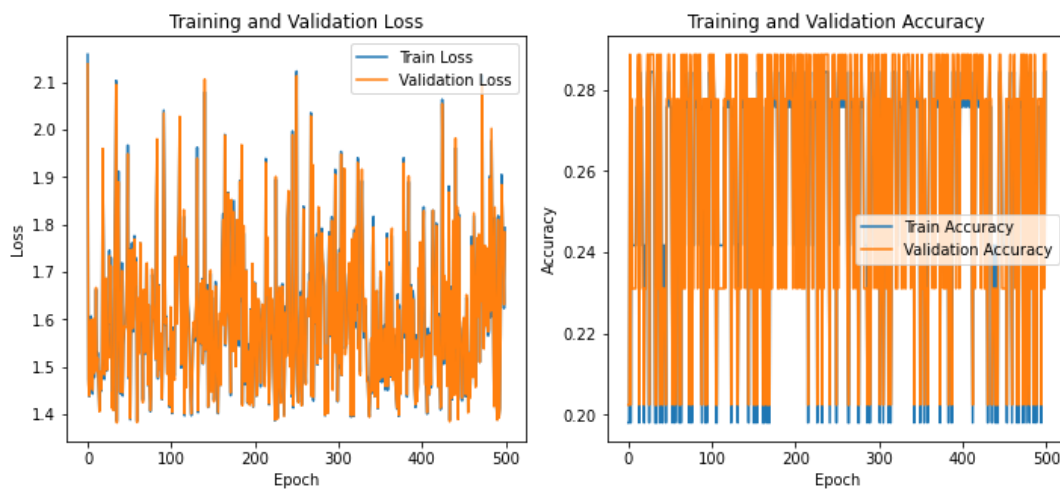By: Sajeev Kugarajah (210554M)

Methodology

The architecture was created according to the instructions in the assignment pdf. And for the training, momentum and rmsprop optimizations were implemented with the following hyperparameters.
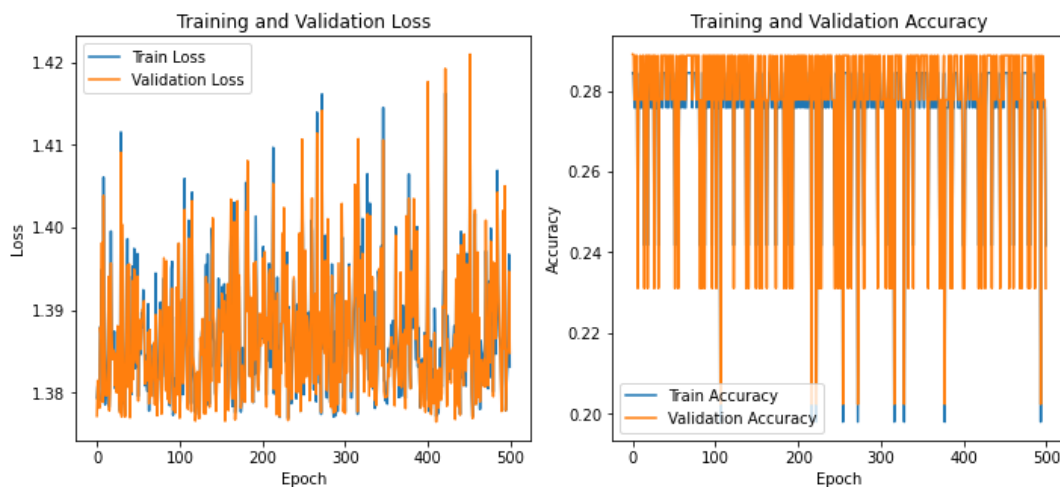
- Momentum factor (Beta) = 0.9
- RMSProp decay factor (Rho) = 0.9
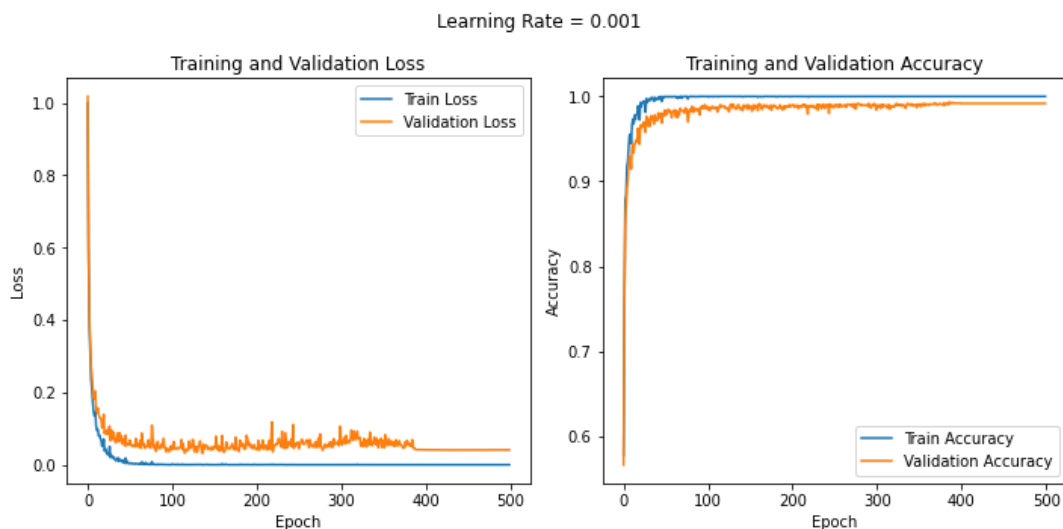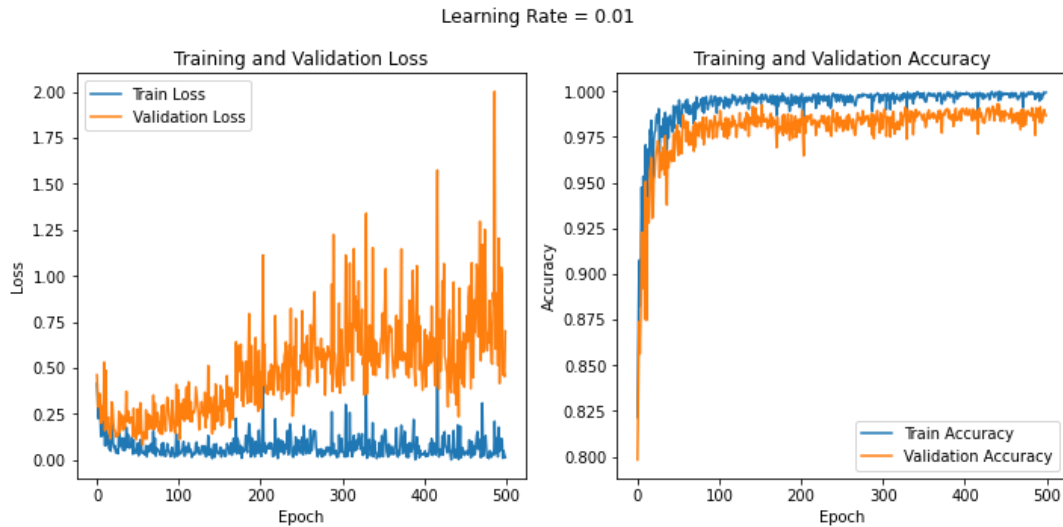- RMSProp epsilon = 1e-8 #for avoiding division by zero

And then, the neural network has trained for 500 epochs with learning rates (1, 0.1, 0.01, 0.001) and the train-test losses and train-test accuracies were calculated and been plotted for each learning rate.

Learning Rate = 0.01

Training and Validation Loss / Training and Validation Accuracy

Learning Rate = 0.001

Training and Validation Loss / Training and Validation Accuracy

Discussion

As expected for high and moderate learning rates 1 and 0.1 the neural networks losses and accuracies tend to oscillate rapidly and have not converged because of instability caused by high learning rates.

And for learning rate 0.01 losses tends to oscillate rapidly and with respect to iterations the validation loss becomes higher than train loss and the accuracies increased and reaching 1 but still have a slight oscillation, which expresses an instability in the training.

And finally, for the learning rate 0.001, the losses tend oscillate a bit initially and then converges to almost 0 with iterations. Likewise, the accuracies initially oscillate a bit but eventually converges to almost 1 with iterations.

From the experiment, we could conclude that for the given architecture 0.001 is the optimal learning rate.