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.

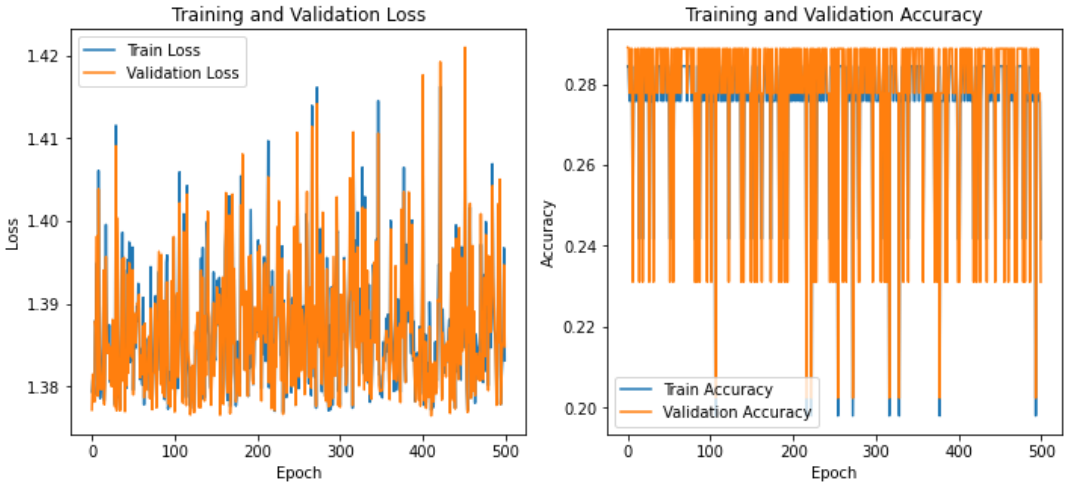
- Momentum factor (Beta) = 0.9
- RMSProp decay factor (Rho) = 0.9
- RMSProp epsilon = 1e-8 #for avoiding division by zero

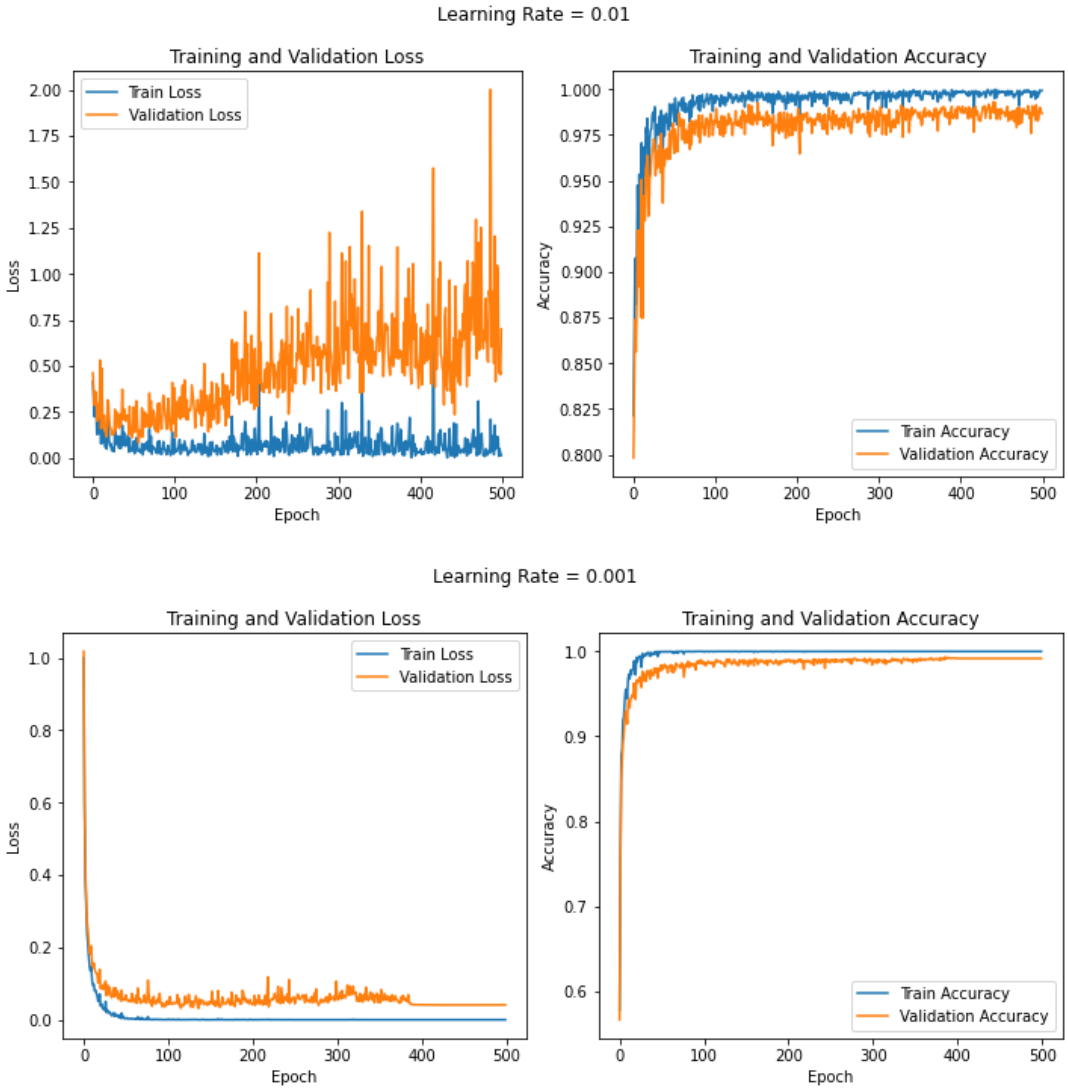
And then, the neural network has trained for 500 epochs with learning rates (1, 0.1, 0.01, 0.001) and the train-test losses and train-test accuracies were calculated and been plotted for each learning rate.

Plots

Learning Rate = 1

Learning Rate = 0.1





Discussion

For high and moderate learning rates of 1.0 and 0.1, the neural network exhibits significant oscillations in both loss and accuracy metrics, indicating a failure to converge. This behavior is attributed to instability introduced by the elevated learning rates.

With a learning rate of 0.01, the network's loss values exhibit rapid oscillations. Additionally, the validation loss surpasses the training loss over iterations, while accuracy metrics approach 1.0 but still display minor oscillations, indicating residual instability in the training process.

Conversely, with a learning rate of 0.001, the loss function initially shows minor oscillations but eventually converges towards zero as training progresses. Similarly, accuracy values begin with slight fluctuations but converge towards 1.0 with continued iterations.

Based on the experimental results, it can be concluded that a learning rate of 0.001 is optimal for this network architecture, as it provides the most stable and convergent training performance.