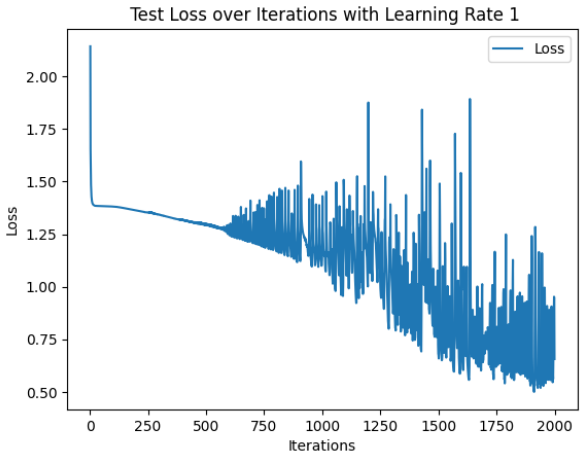
CS3630 Assignment 1

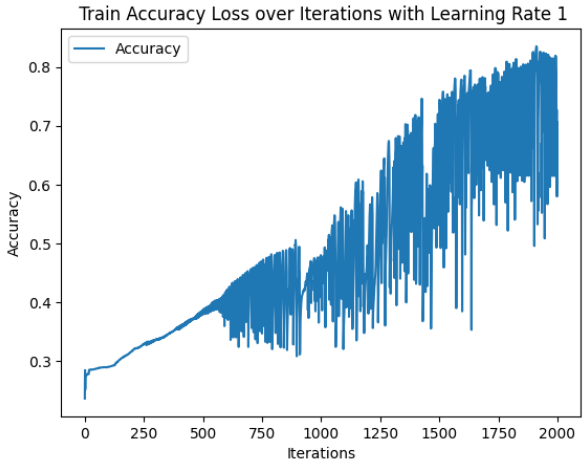
D.M.N.D. Dissanayake
210144G

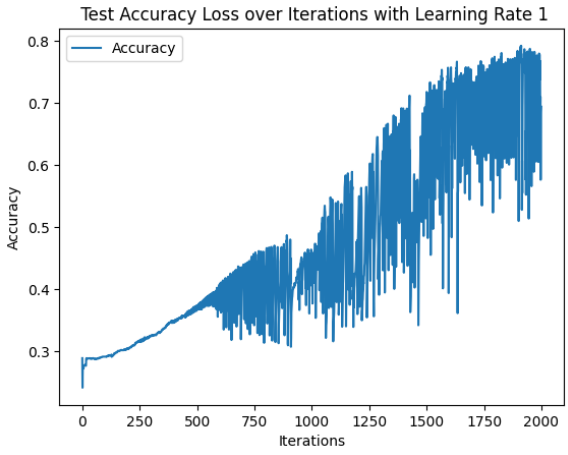
Git Hub link for the assignment - NisithDivantha

For the below plots as initial weights and biases, I have used the weights and biases set in b-100-40-4.csv & w-100-40-4.

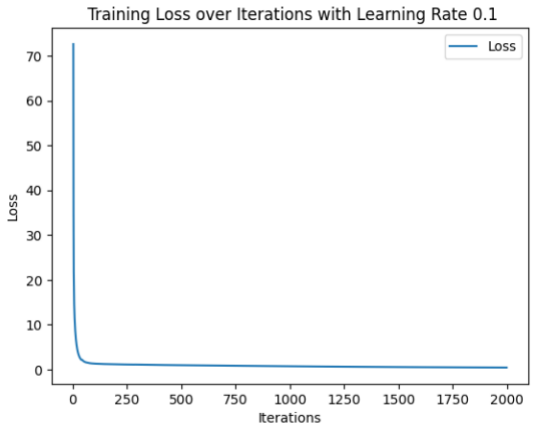
- **Plots for Learning Rate = 1**

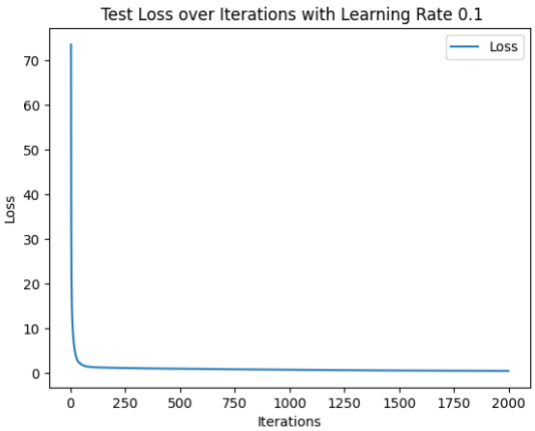


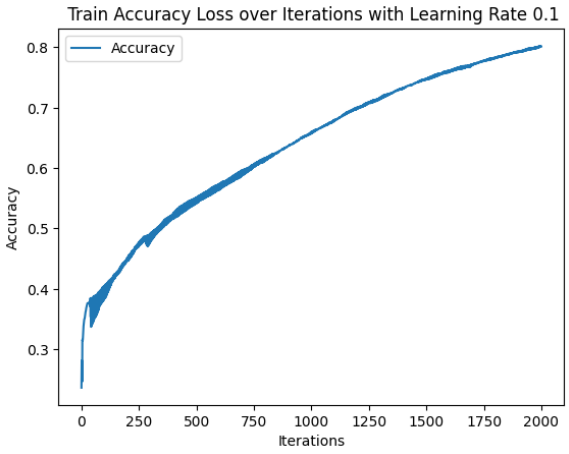


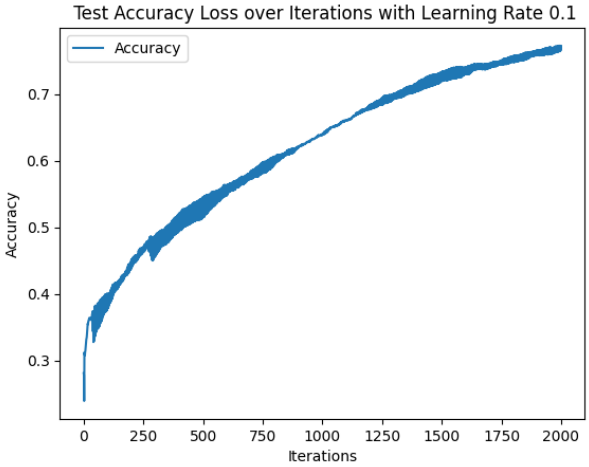


- **Plots for Learning Rate = 0.1**

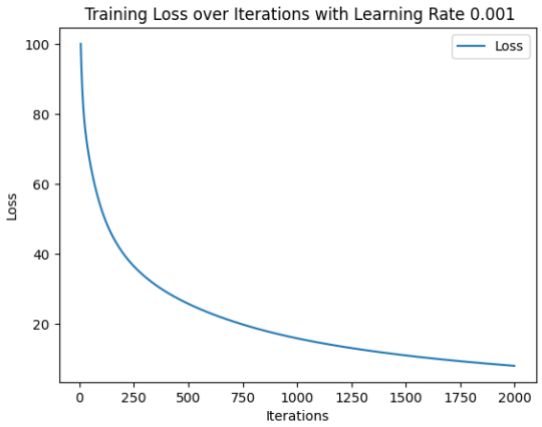


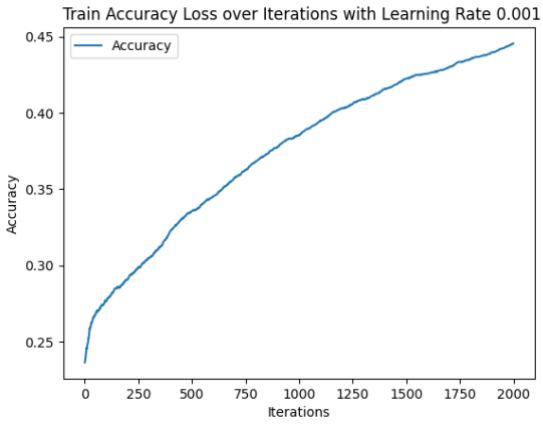


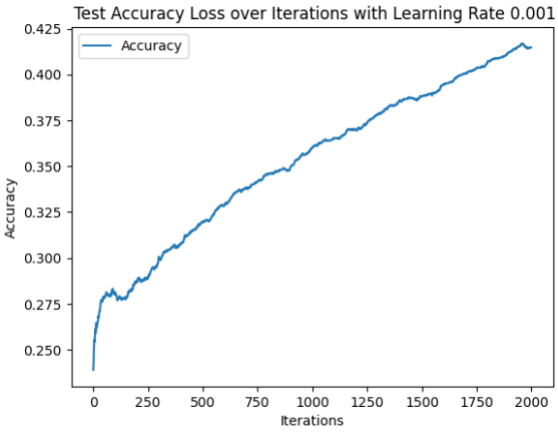




- **Plots for Learning Rate = 0.001**







- When we use a high learning rate like 1 there's a possibility of overshooting a local or global minima and wandering around for a long time. Accuracy also fluctuates due to this reason.
- When the learning rate is too low it takes a lot of time to converge. In the loss plots for learning rate = 0.001, we can see even after 2000 iterations loss is around 10. The reason for this is at each iteration steps of gradient descent are very small. Accuracy is increasing very slowly.

Influence of the learning rate on test accuracy

- If we use a high learning rate our test accuracy would be unstable. It will fluctuate but as an average, it will increase with the number of iterations.
- If we use a learning rate which too small then accuracy will increase very slowly.
- So we should use a moderate value as our learning rate to reach better and efficient accuracy.