INDIAN INSTITUTE OF TECHNOLOGY, MADRAS

Programming Assignment 1

Subhrajit Makur(CS17m046) and Amar Vashishth(CS17m052)

February 17, 2018

1 VARYING HIDDEN LAYER CONFIGURATIONS

1.1 PLOTS OF TRAINING LOSS

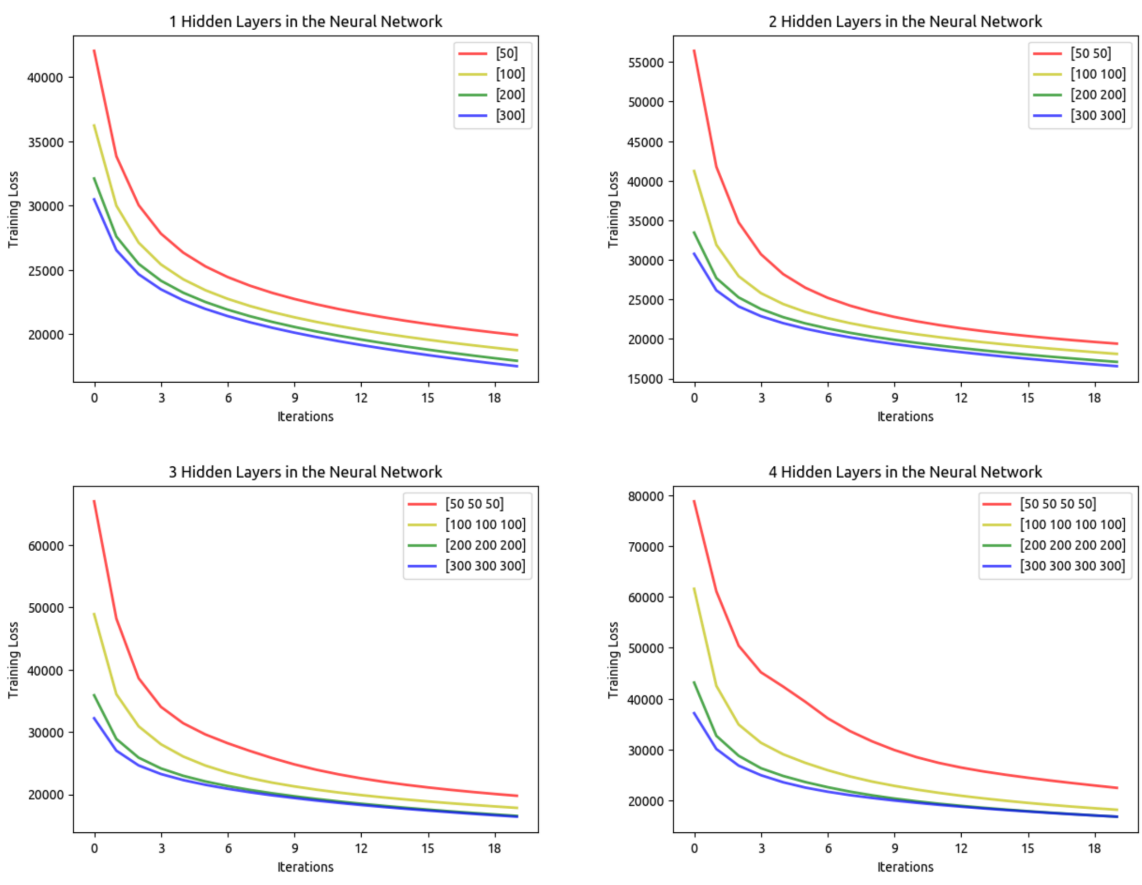

Figure 1.1: Varying Hidden layers in Neural Networks

This Training loss Vs Iteration plot compares different scenarios of Neural Networks.

1.2 PLOTS OF VALIDATION LOSS

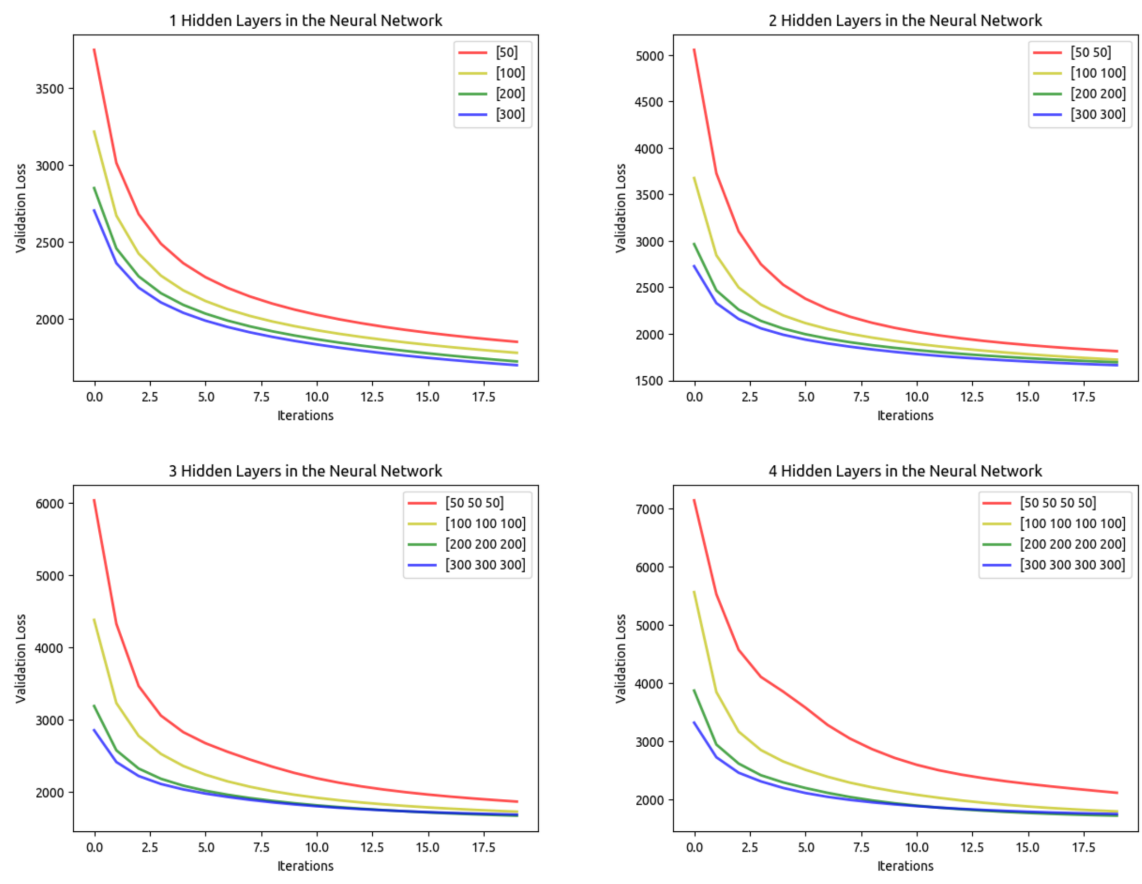

Figure 1.2: Varying Hidden layers in Neural Networks

This Validation loss Vs Iteration plot compares different scenarios of Neural Networks.

2 OPTIMIZER PLOTS

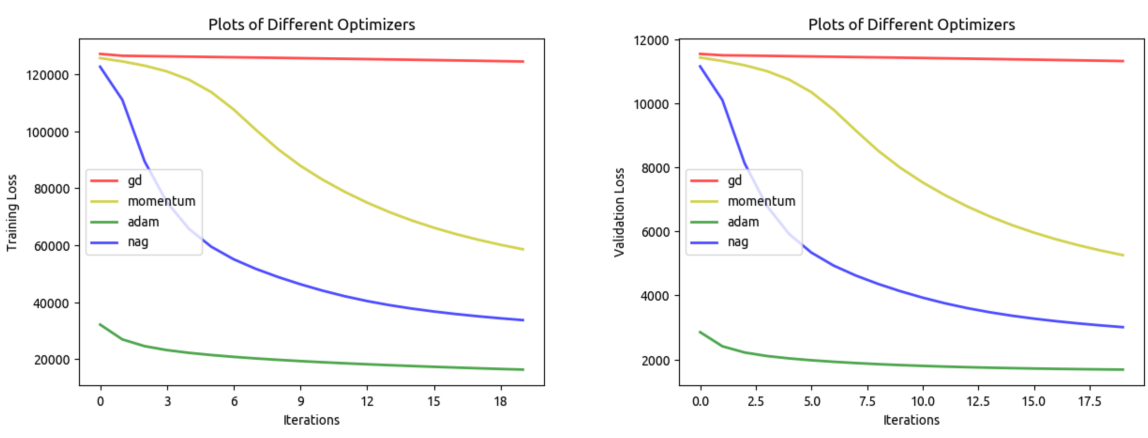

Figure 2.1: Loss when using Different Optimizers

3 ACTIVATION PLOTS

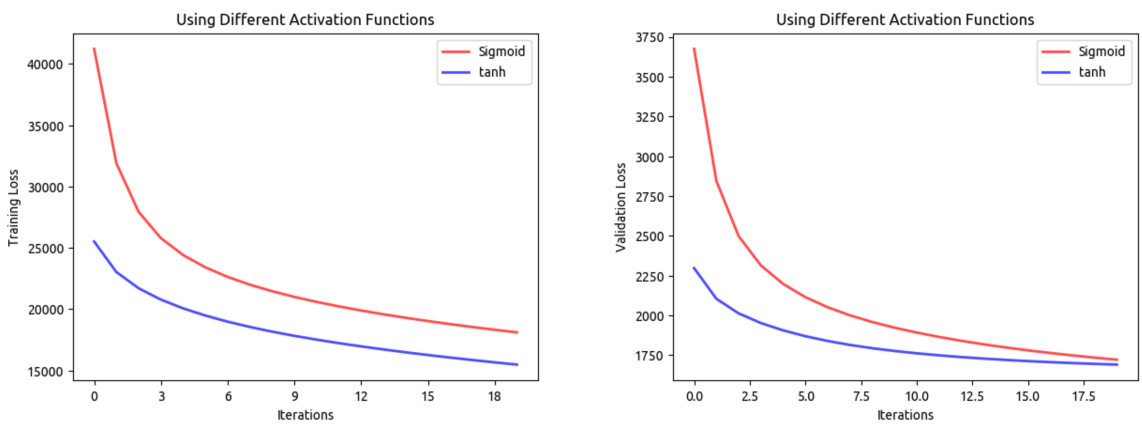

Figure 3.1: Loss when using Different Activation Functions

4 LOSS FUNCTION PLOTS

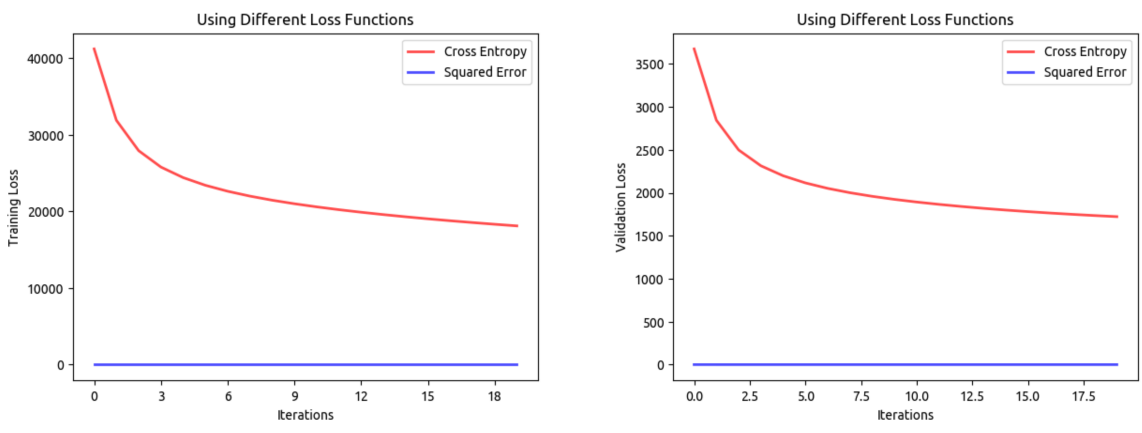

Figure 4.1: Loss when using Different Loss Functions

5 BATCH PLOTS

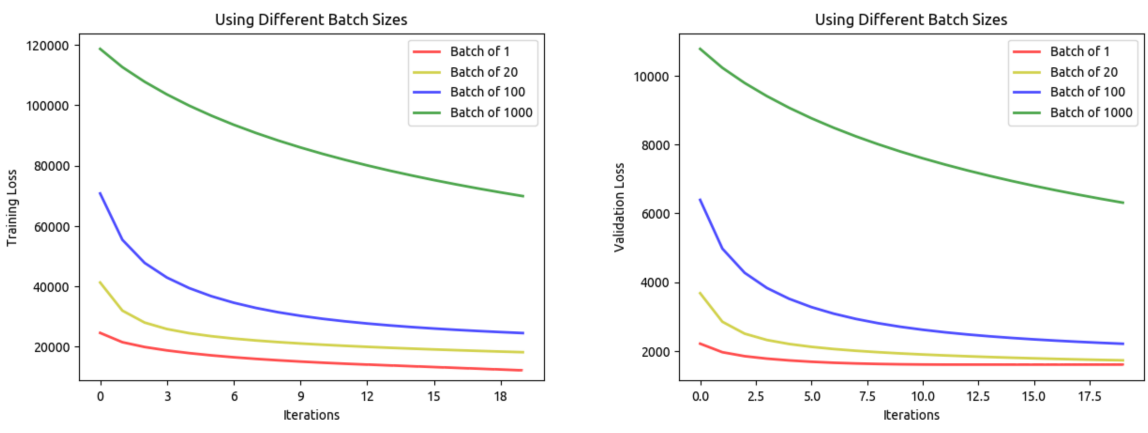

Figure 5.1: Loss when using Different Batch Sizes

6 INFERENCE

- On increasing the number of neurons in the network we obtain a lower loss value and hence it converges faster in comparison with a neural network with lesser number of layers.
- On increasing the number of layers keeping the neurons same we obtain a higher loss value initially but the loss converges to a lower value faster.
- We found **adam** to be the best optimizer, loss with this is very low initially and it converges to a lower value very fast.
- Nesterov's accelerated gradient descent converges the fastest in comparison with other used optimizers. Although, the starting loss value with this optimizer is same as other optimizers except **adam**.
- Increasing the batch size helps. It accelerates loss convergence.
- Performance of SGD is improved if we shuffle the data.