Multi-layer Neural Network

Spoorti Joshi
PID: A53283648
University of California San Diego
sjoshi@ucsd.edu

Khushboo Agrawal
PID:A53271205
University of California San Diego
khagrawa@ucsd.edu

Abstract

In this project, we aim to perform hand-wriiten digit classification using back-propagation in multi-layer neural network. The model is trained using mini batch stochastic gradient descent for three activation functions namely, sigmoid, tanh, and ReLU. We experimented the model for different model parameters such as learning rate, regularization, momentum and hidden layer architecture. The best accuracy obtained on the test set is 96.24 percent using ReLU as the activation function, learning rate 0.0001, l2 penalty as 0.001 and batch size of 1000.

1 Training the Classifier

1.1 Training Procedure

- For this part of the problem, we have used mini-batch stochastic gradient descent to train a classifier which maps each input to the one-hot encoded labels.
- Batch size of 1000 with learning rate of 0.0001 is considered to implement the training.
- Multi layered neural network is trained using back propagation to make predictions. One hidden layer with 50 units is used.
- Three different activation functions are used to train the model and the validation set is used to determine the best model.
- Momentum weighted $\gamma = 0.9$ is used to account for the weight decay.
- Early stopping is considered on the validation set while training the classifier with early epoch as 5, i.e if the validation error goes up for 5 epochs, training stops and the last updated weights are considered.
- Regularization is taken into account to penalize the model for its complexity.
- The model is executed for different activation functions like tanh, sigmmoid and ReLU and their results are mentioned in the following report.
- To further analyze the network topology, the module is enhanced with additional hidden layer, as well as the units of each layer is varied to test the accuracy.
- All the experimental analysis is presented in the following report.

2 Activation Function: tanh

Figure 1 shows the plot of the training vs validation loss with respect to epoch and their respective accuracy. The accuracy obtained with no regularization penality is 89.5 percent.

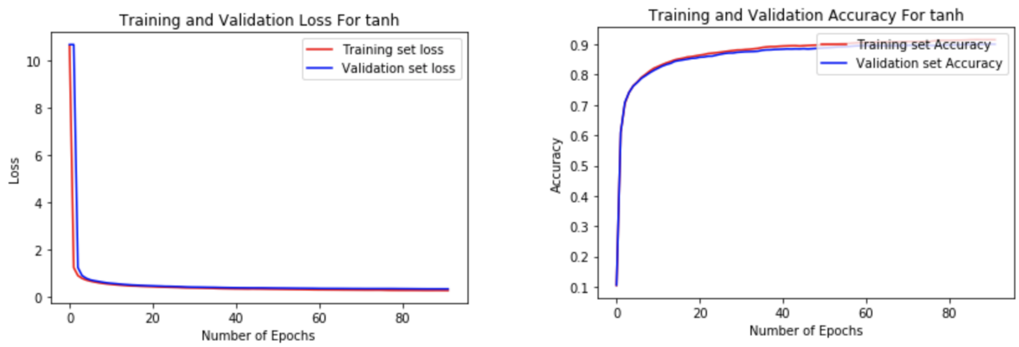

Figure 1: Activation function tanh. Loss and Accuracy plots

2.1 Experiment with regularization

Keeping the activation function as tanh, the l2 penalty was introduced to regularize the complexity. Regularization increases the accuracy. Following are the results obtained.

- Figure 2 shows the activation loss and accuracy on the training and validation set with l2 penalty as 0.001. The best accuracy obtained is 90.78 percent with l2 penalty as 0.001.
- Figure 3 shows the activation loss and accuracy on the training and validation set with l2 penalty as 0.0001. The accuracy obtained is 90.38 percent.

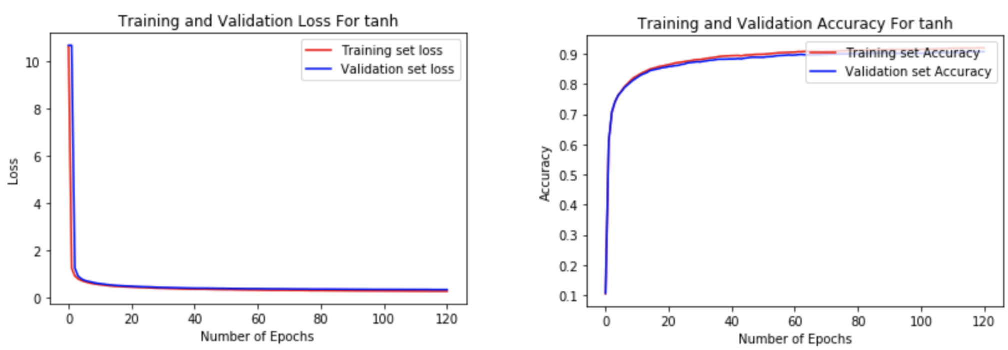

Figure 2: Activation function tanh. Loss and Accuracy plots, l2 penalty 0.001

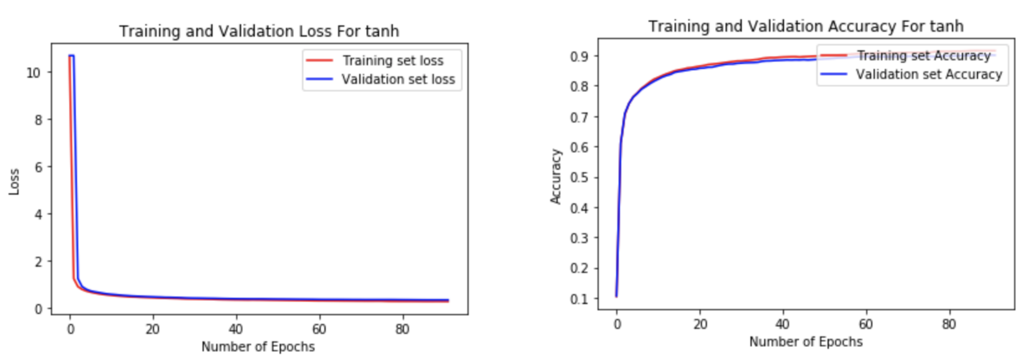

Figure 3: Activation function tanh. Loss and Accuracy plots, l2 penalty 0.001

3 Experiment with Activation

In this part of the report different activation functions namely, sigmoid and ReLU are used to train the model and the results obtained are shown below.

3.1 Activation function: Sigmoid

Figure 4 shows the activation loss and accuracy on the training and validation set with l2 penalty as 0.001. The accuracy obtained is 93.38 percent.

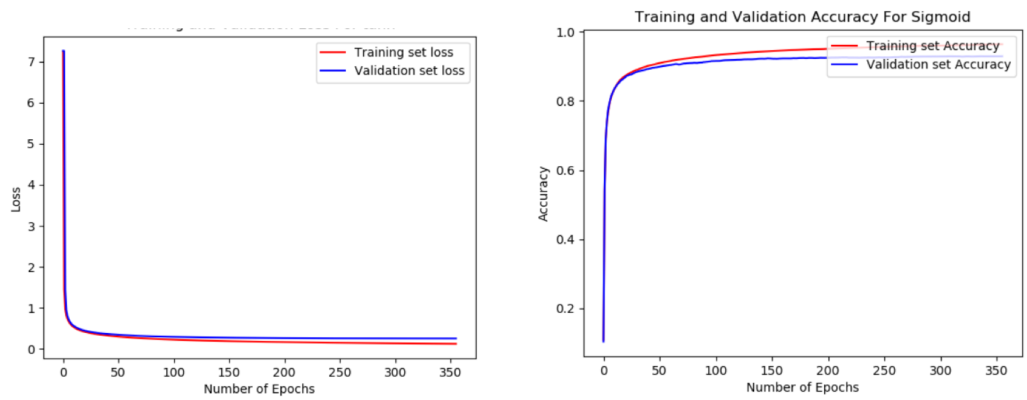

Figure 4: Activation function Sigmoid. Loss and Accuracy plots, l2 penalty 0.001

3.2 Activation function: ReLU

Figure 5 shows the activation loss and accuracy on the training and validation set with l2 penalty as 0.001. The accuracy obtained is 96.24 percent, which is the best amongst all.

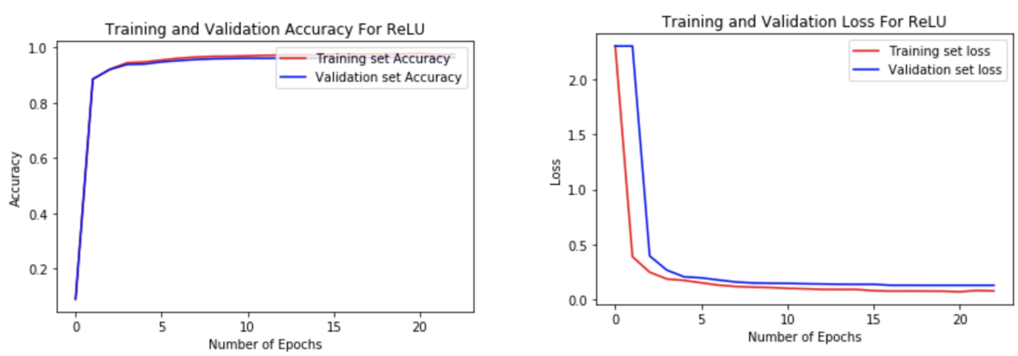

Figure 5: Activation function ReLU. Loss and Accuracy plots, l2 penalty 0.001

4 Experiment with Network Topology

In this part of the problem, the units of the hidden layer are varied and the results obtained are reported. For this problem the l2 penalty is kept 0.001.

- Hidden layer with 25 units. Figure 6 shows the loss and accuracy plots. The accuracy obtained is 88.67 percent.

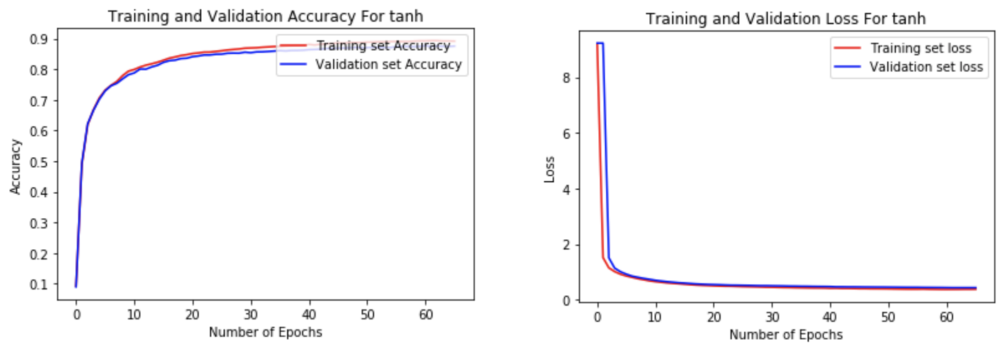

Figure 6: Activation function tanh. Loss and Accuracy plots, l2 penalty 0.001, Hidden layer 25 units

- Hidden layer with 100 units. Figure 7 shows the loss and accuracy plots. The accuracy obtained is 91.07 percent.

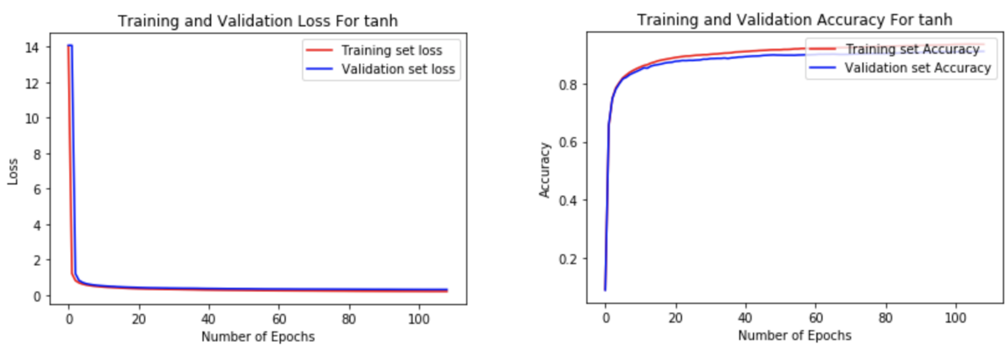

Figure 7: Activation function tanh. Loss and Accuracy plots, l2 penalty 0.001, Hidden layer 100 units

- Two hidden layers with 47 units each. Figure 8 shows the loss and accuracy plots. The accuracy obtained is 90.19 percent.

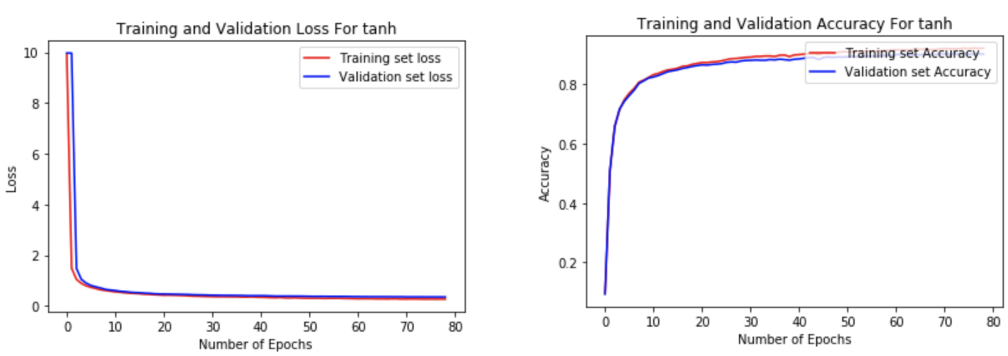

Figure 8: Activation function tanh. Loss and Accuracy plots, l2 penalty 0.001, Two Hidden layer 47 units each

The accuracy increases with the number of hidden units and decreases with the additional hidden layer keeping the other parameters same.

5 Individual Contribution

5.1 Spoorti (A53283648)

- Implementation of forward and Back propagation
- Implementation of Momentum and Regularization
- Implementation of mini batch stochastic gradient descent
- Implementation of Sigmoid and tanh activation functions

5.2 Khushboo (A53271205)

Experimental Analysis with regularization, Network topology and Activation functions
Implementation of ReLU Activation Function
Report formulation