

Deep Learning Lab 2018.

Exercise 1

Sai Sourabh Tiruvaipati

The goal of this exercise is to investigate the neural network architecture, hyperparameters such as epoch numbers, learning rate etc. and their effect on the training of a neural network. In addition to this various activation functions used in neural networks have been worked upon. The neural network is designed to perform a multi-class classification problem. To optimize the weight and the bias, a stochastic gradient descent and gradient descent algorithms have been implemented.

The parts of the notebook marked to do have been followed up. Resources such as The Internet, the Deep Learning book, Lecture slides, discussions with peers have assisted in approaching the coding part.

The layers of the Neural Network have been tested with different activation functions at each layer. Relu has been used as activation function of Layer 1 and Layer 3 respectively and Tanh function has been used as activation function. Initially, the number of units in the hidden layers were set to 100 each. The learning rate and number epochs have been set to 0.06 and 20 respectively with a batch size of 64. Since the validation error was higher than 5% with these hyperparameters, they have been tuned manually. At the time of submission, the number of units in the hidden layers was increased to 450 and 500 respectively.

The final test error is **1.77%**. The tuned hyperparameters to achieve this are 0.09 and 30 for the learning rate and number of epochs, with a batch size of 64. The neural network with these parameters classifies the validation set with a small error of 1.85%. The duration to execute is 740s. The figure depicts the change in error and loss during training and validation with respect to the number of epochs.

