Assignment 5: Problem 1

Page 15 of Lecture 5 Notes introduces the Pre ReLu activation function. The parameter alpha in this function is learnt as part of Training Process. Write down the update rule for alpha by deriving an expression for the gradient $\frac{\partial L}{\partial \alpha}$ in terms of the other gradients that are computed as part of the Backpropagation algorithm.

(2 points)

Assignment 5: Problem 2

With reference to the TensorFlow model “MNIST Fully Connected Model Using Dense Module”, modify the training data set, by reducing its size from 50,000, in order to reproduce the overfitting problem. Plot the Test Data Accuracy and the Validation Data Accuracy together (as a function of the number of epochs) to show that overfitting is occurring.

(2 points)

Assignment 5: Problem 3

- With reference to the TensorFlow program “MNIST Fully Connected Model Using Dense Module”, consider a model with 2 hidden layers consisting of 50 nodes each. Investigate the effects of the following Optimization Algorithms on the rate of convergence, by plotting the Test Loss Function (on a per Epoch basis) vs Number of Epochs (you can also try plotting the Test Accuracy Value if that works better).

1. Regular Gradient Descent
   optimizer = tf.train.GradientDescentOptimizer(learning_rate = 0.001)
2. Momentum
   optimizer = tf.train.MomentumOptimizer(learning_rate = 0.001, momentum = 0.9)
3. Adam
   optimizer = tf.train.AdamOptimizer(learning_rate = 0.001)

(3 Points)

Assignment 5: Problem 4

- The TensorFlow program “CIFAR Linear Model Using Dense Module”, does Linear Classification for the CIFAR-10 image database. Download and run the program and make sure you understand how it works. The MNIST program uses the “mnist.train.next_batch” function to generate the test data for the next batch. The CIFAR model on the other hand, does not come with such a function, and the data feed has to be programmed explicitly.
  Answer the following questions:

1. Write down the equation that the program uses to do image pre-processing.
2. What is the best Test Data Accuracy that you can get with this Linear Model? Hint: Find the best Learning Rate, Batch Size.
3. Replace the Linear Model with a Deep Feedforward Model, with 2 hidden layers with 50 nodes each. Once again find the best Test Data Accuracy, after finding appropriate values for hyper-parameters. Can you get better accuracy by adding more layers and/or adding more nodes per layer?

(3 Points)