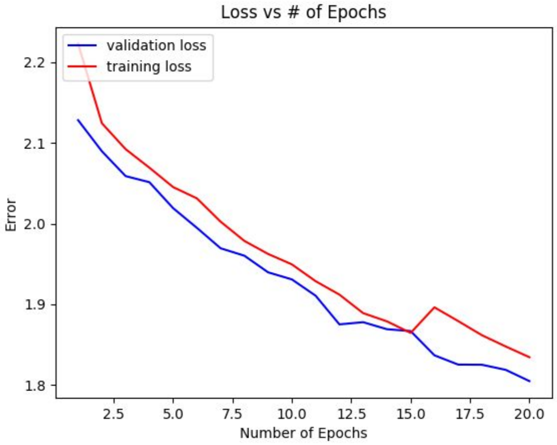
Coulby Nguyen
(1 person group)
5/14/2019

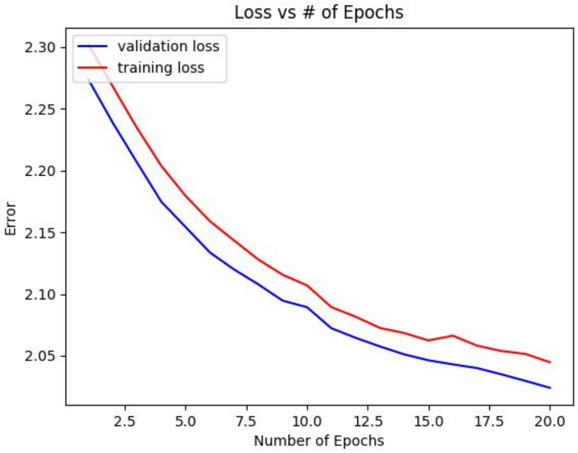
CS 434 Project 3 Write Up

//Note: I am using a set number of epochs = 5 to be able to get data without waiting a lengthy amount of time.

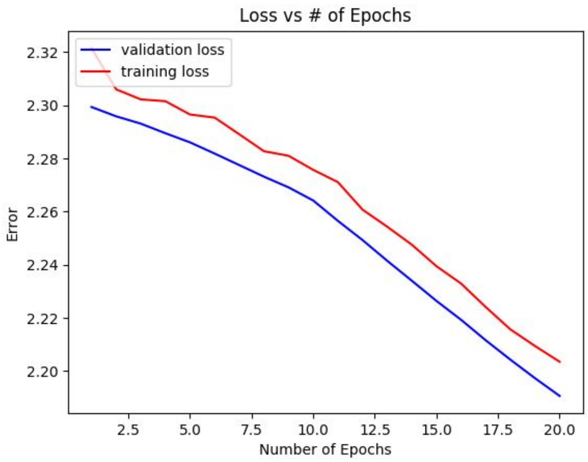
1. Question 1
    a. After testing the learning rates of {0.1, 0.01, 0.001, 0.0001} and plotting the validation loss, and training loss, I have come to the conclusion that the learning rate of 0.001 is the best choice. Though this learning rate yields only 23% from just 5 epochs, the way the data is trending, it seems that it will approach smaller error the fastest out of the learning rates. The signal to stop training is when validation loss converges, or gets incredibly worse.
    b. Learning Rate: 0.1 - 36% test accuracy



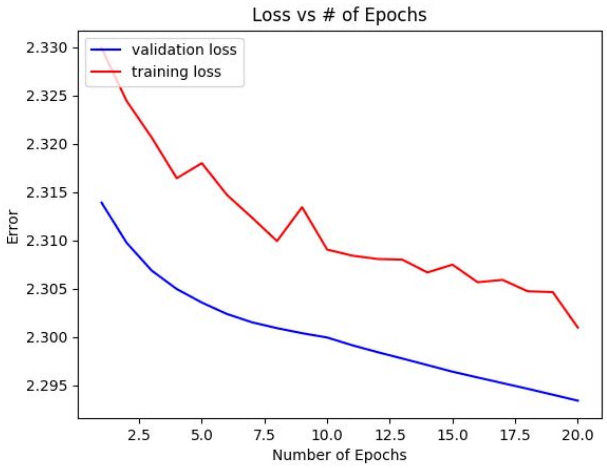
    c. Learning Rate: 0.01 - 29% test accuracy



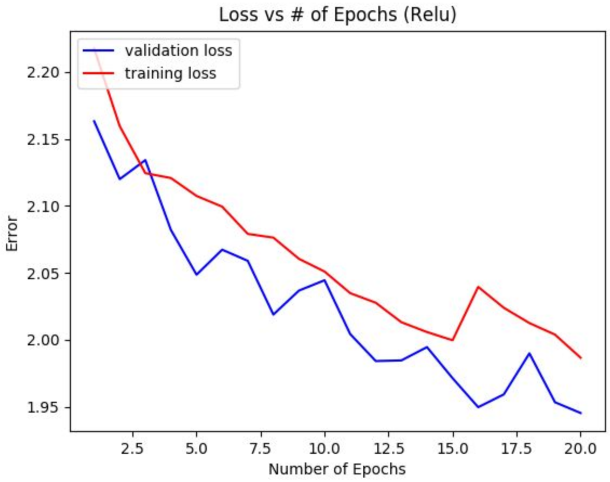
    d. Learning Rate: 0.001 - 23 % test accuracy



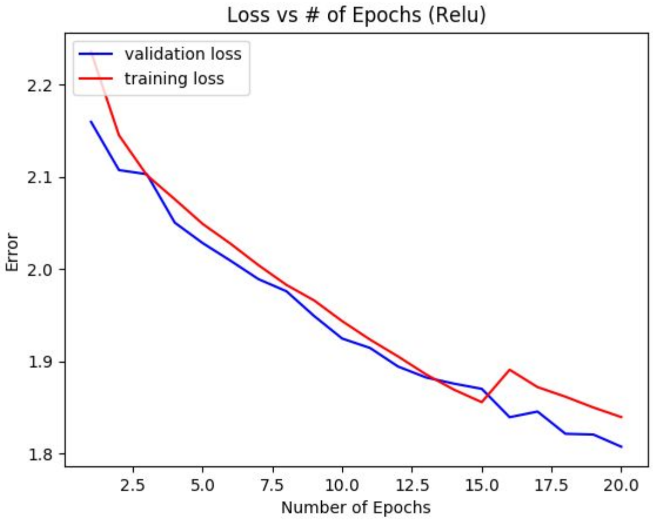
e. Learning Rate: .0001 - 18% test accuracy



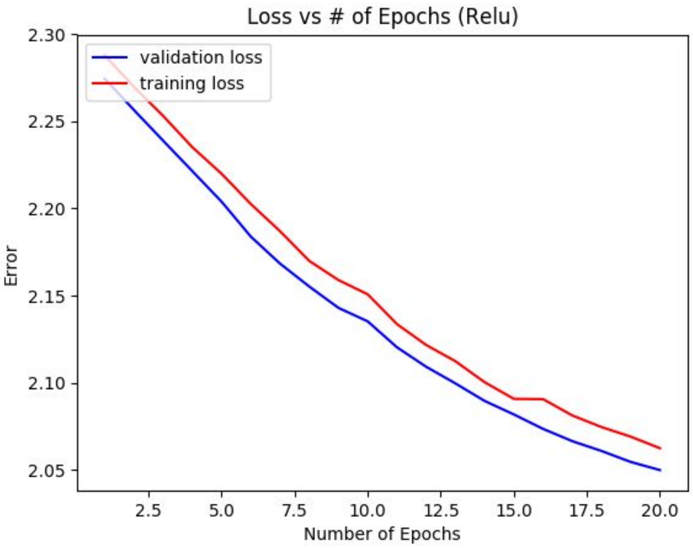
2. Question 2
   a. Using the same loss function as in problem 1 I was able to get better results for the most part, as for each learning rate that were used in both the sigmoid, and relu activation functions, the relu did exceedingly better in the 5 epoch span. However the sole exception was of learning rate size of .1, which yielded a worse test accuracy than the sigmoid. This could be due to overstepping, as seen in the validation loss, the peaks that are formed after every other epoch. For this structure and network I would choose the learning rate of .01 as it yielded the highest test accuracy and the lowest validation/training losses.
   b. Learning Rate: 0.1 - 29% accuracy



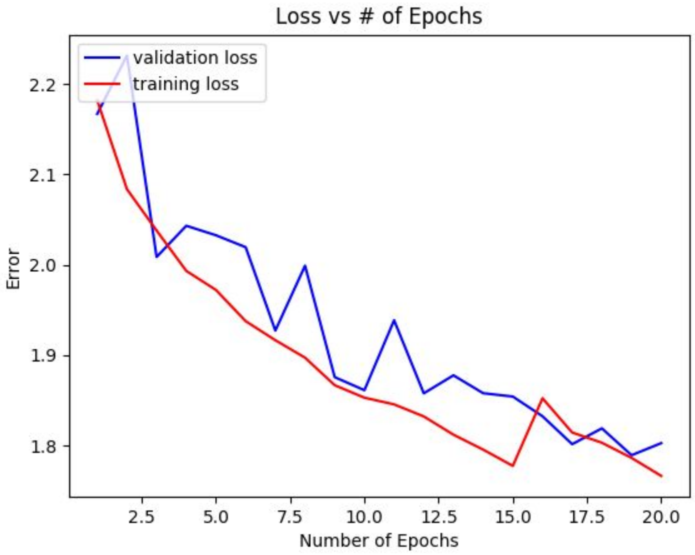
c. Learning Rate: 0.01 - 37% accuracy



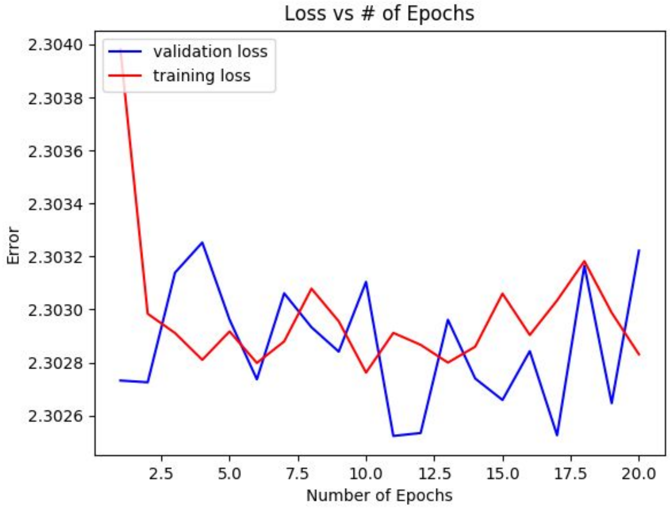
d. Learning Rate: 0.001 - 28% test accuracy



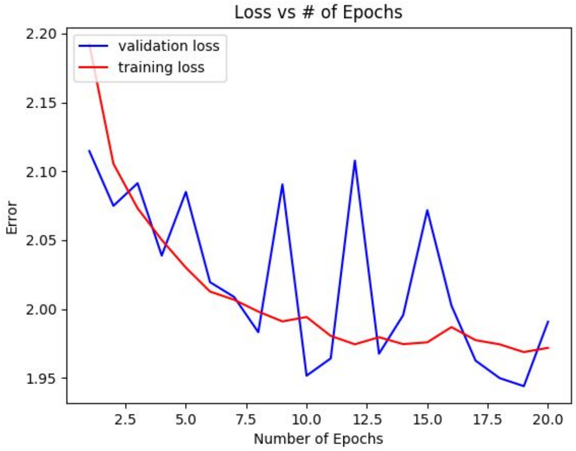
3. Question 3
    a. As seen in the graphs of different values of M and D, as the size of M and D increases the test accuracy plummets. The reason for this is with higher values for these variables, the larger amount of data is misclassified. This could be easily seen with momentum as with higher momentum can affect the classification. However, accuracy also plummets with increase of weight decay. Having set fixed values of m and d, and rising weight decay to over .5, resulted in a steady 10%accuracy. When combined and finely tuned these variable have the capability to converge faster.
    b. Different values of m, d, wd with constant learning rate = .1
        i. m=0, d=0, wd=0



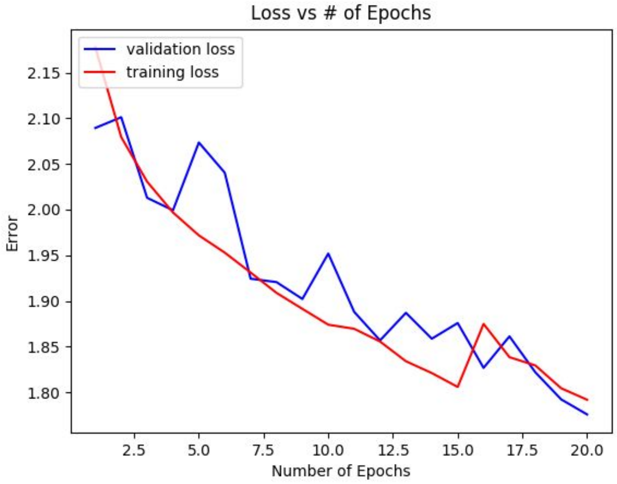
        ii. m=.5, d=.5, wd=.5



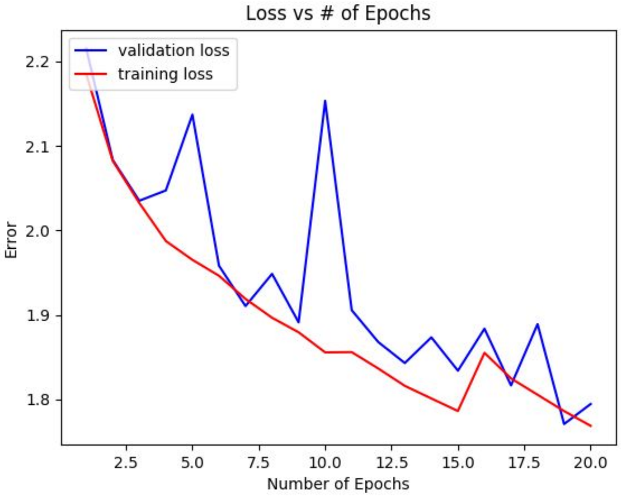
iii. m=.01, d=.01, wd=.01



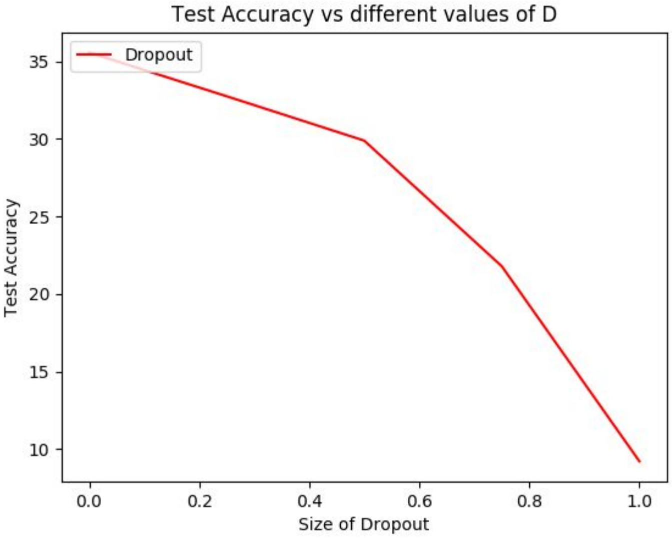
iv. m=.05, d=.001, wd =0



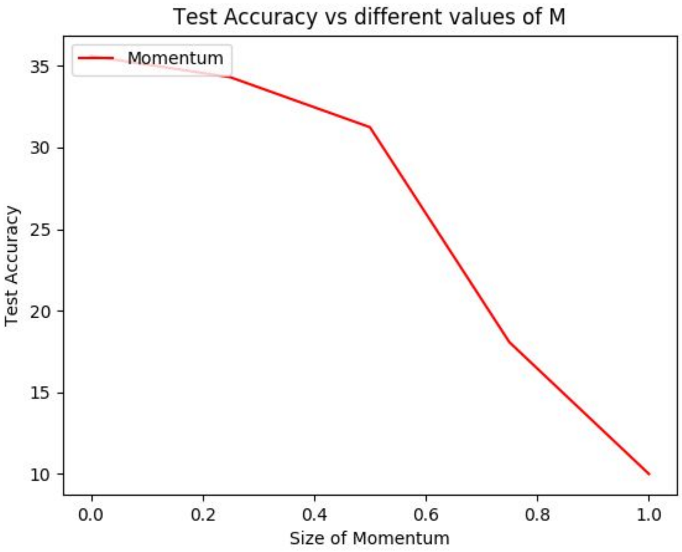
v. m=.01, d=.005, wd =0



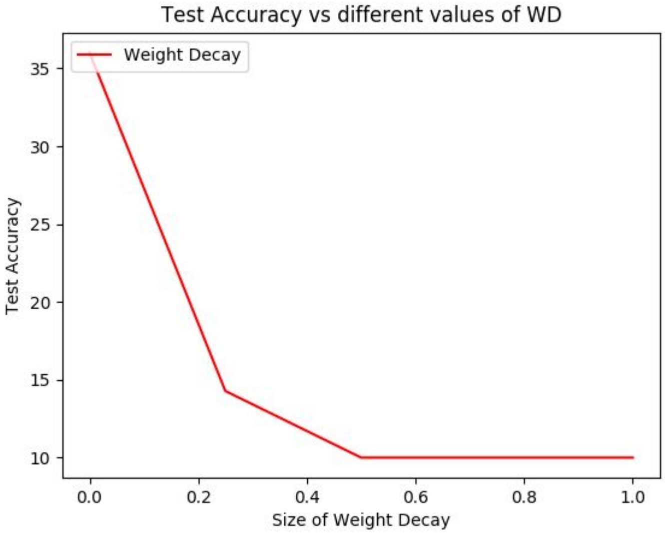
c. Test Accuracy vs different values of Dropout



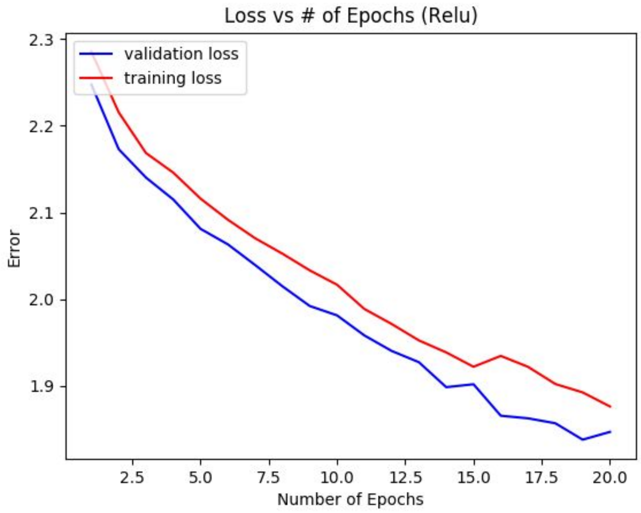
d. Test Accuracy vs different values of Momentum

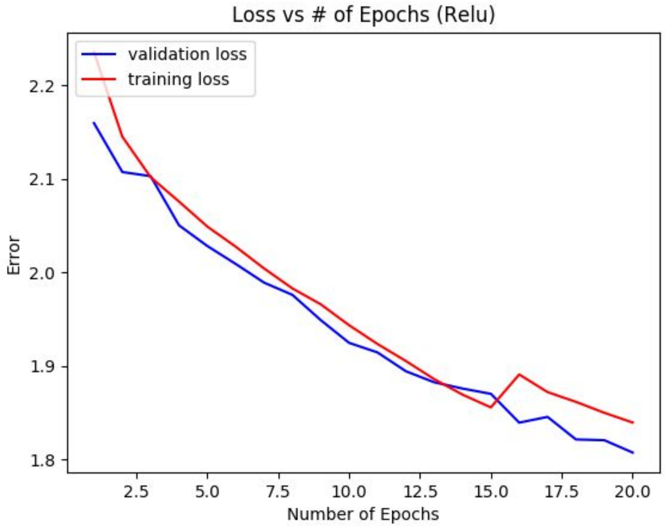


e. Test Accuracy vs different values of Weight Decay

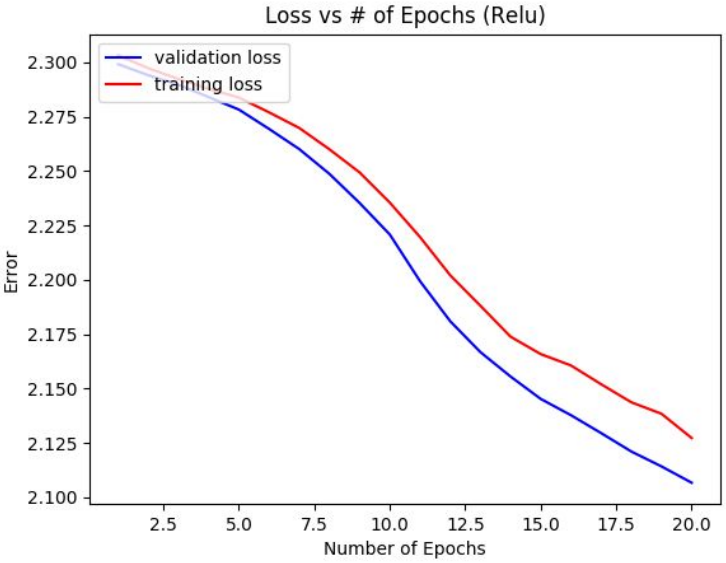


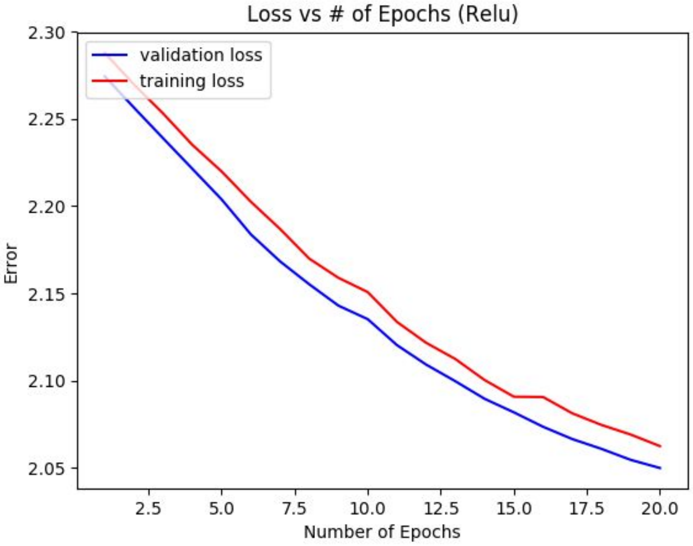
4. Question 4
   a. After creating the 1 layer of 100 hidden nodes and the 2 layers of 50 hidden nodes, and training them, I believe that the 1 layer of 100 hidden nodes is the better structure. The time it took to train was incredibly longer than the time it took to train with just one layer of 100 nodes. Not only was time a factor in deciding the better structure but accuracy. In both training loss, and testing accuracy, the single layer of 100 nodes out performed with the learning rates of .01 and .001. However, in the learning rate of 0.001 there is a point of inflection that changes the concavity of the curve and might lead to better performance with larger amounts of epochs.
   b. Learning Rate 0.01 - 34% test accuracy (2 hidden layers of 50)(top graph) vs Learning Rate 0.01 - 37% test accuracy (1 hidden layer of 100)(bottom graph)





c. Learning Rate 0.001 - 24% test accuracy (2 hidden layers of 50)(top graph) vs Learning Rate 0.001 - 28% test accuracy (1 hidden layer of 100)(bottom graph)


Loss vs # of Epochs (Relu)
validation loss
training loss
Error
2.300
2.275
2.250
2.225
2.200
2.175
2.150
2.125
2.100
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Number of Epochs


Loss vs # of Epochs (Relu)
validation loss
training loss
Error
2.30
2.25
2.20
2.15
2.10
2.05
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Number of Epochs