Jordan Gilman
931942845
CS 434

Implementation Assignment #3 Report

1. Using a batch size of 32, I trained a 2-layer neural net on learning rates of 0.1, 0.01, 0.001, and 0.0001, with 0.1 performing the best with a validation accuracy of 46%. This is not especially impressive, but given that there are 10 classes, it is not entirely useless as a classifier. Figures 1 through 4 show the negative loglikelihoods and validation accuracies as a function of epochs. As a simple stopping measure, I created a stop-function that checks whether the loss has changed within 1% of the average los, given at least 10 epochs have passed. My goal with this stop measure is to stop training once there is no significant change in training performance.

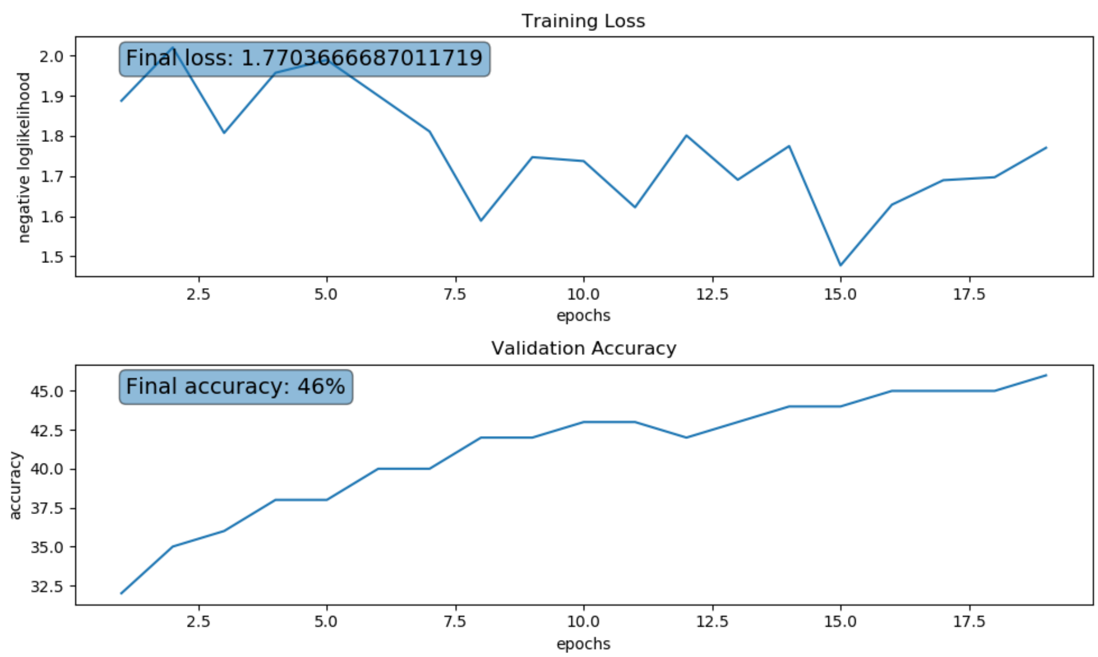

Figure 1. 2-Layer Sigmoid Activated, Learning Rate = 0.1, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.0

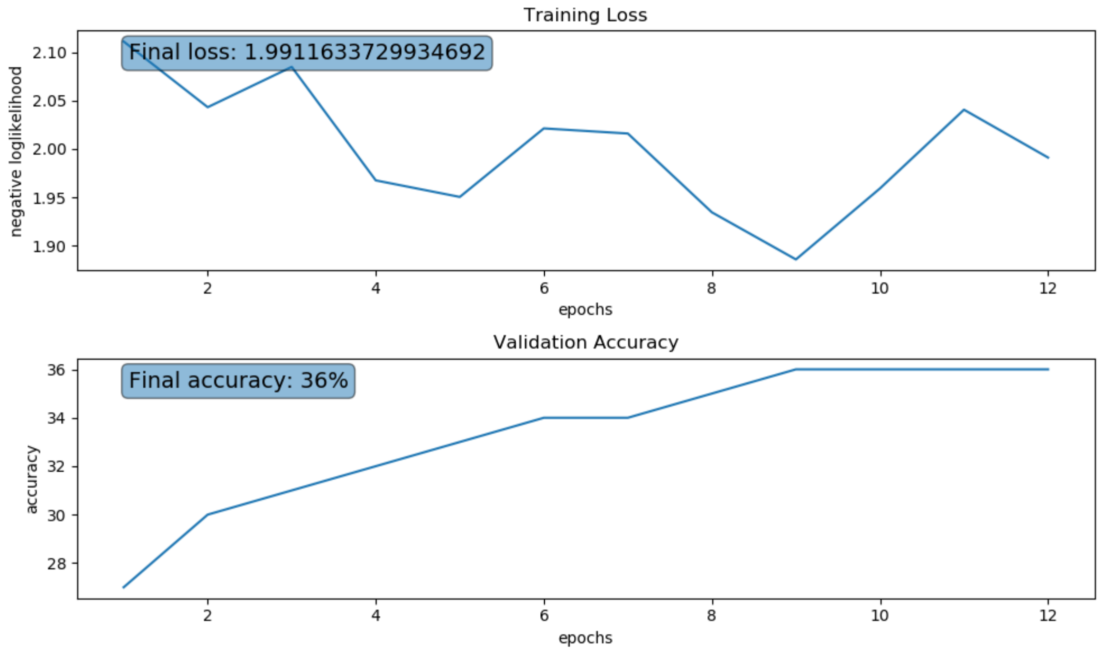

Figure 2. 2-Layer Sigmoid Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.0

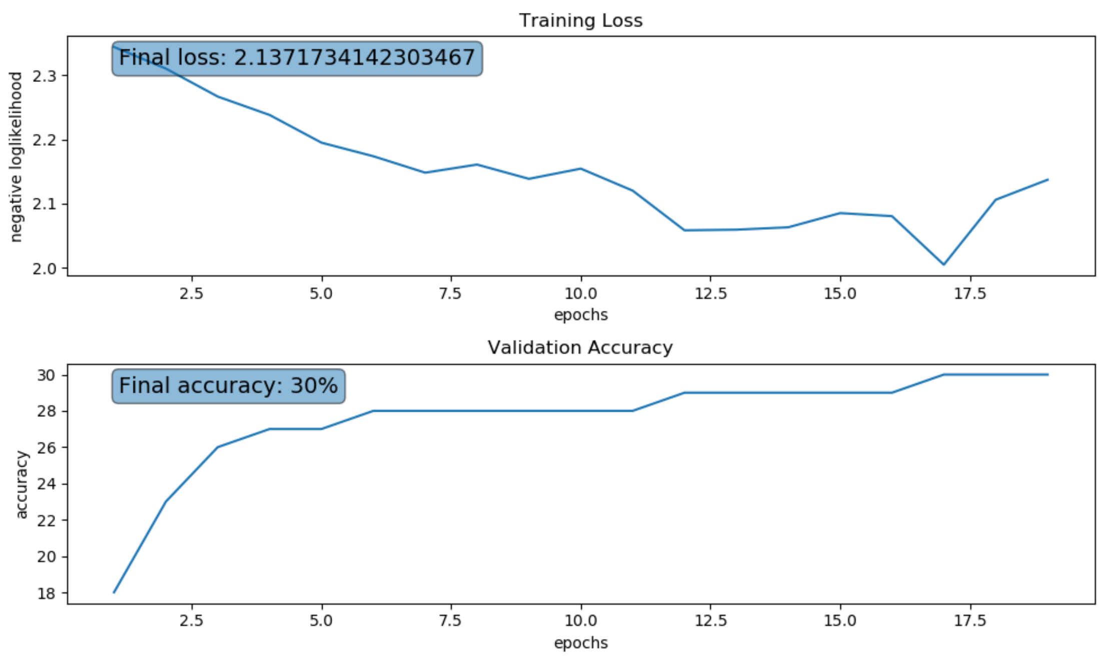

Figure 3. 2-Layer Sigmoid Activated, Learning Rate = 0.001, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.0

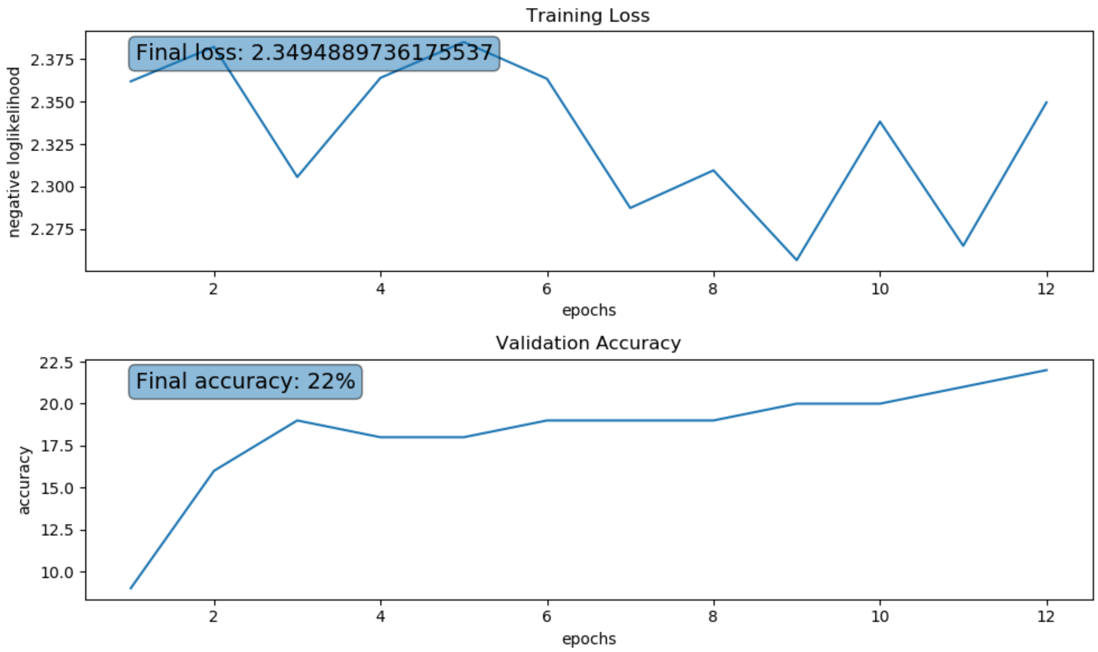

Figure 4. 2-Layer Sigmoid Activated, Learning Rate = 0.0001, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.0

2. Repeating the experiment with ReLU activation resulted in the plots in figures 5 - 8. With ReLU activation, the results were similarly unimpressive, however a learning rate of 0.01 led to the greatest validation accuracy, 43%. I will note that perhaps more epochs being used as a minimum might have led to greater overall accuracy, given the stochastic pattern of the loss function shown in the plots.

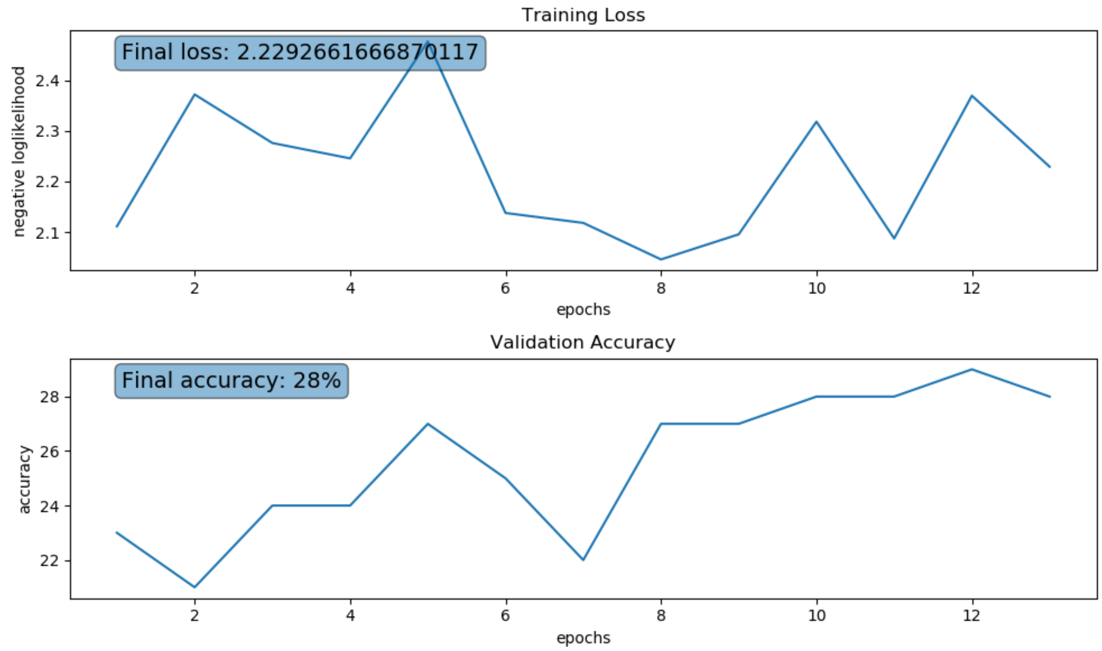

Figure 5. 2-Layer ReLU Activated, Learning Rate = 0.1, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.0

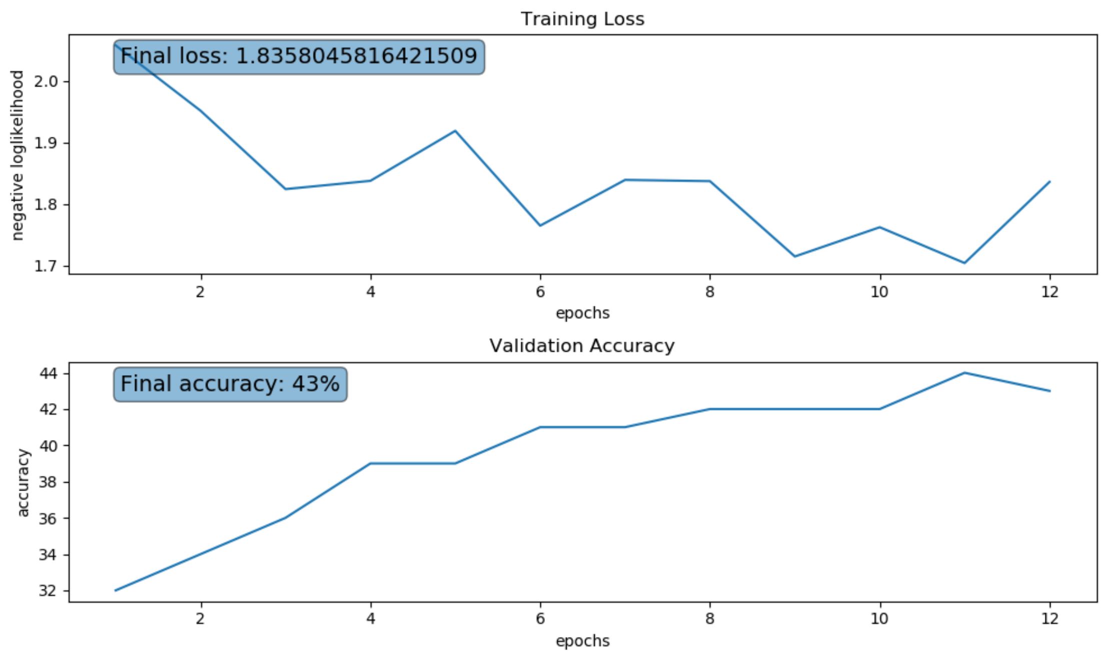

Figure 6. 2-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.0

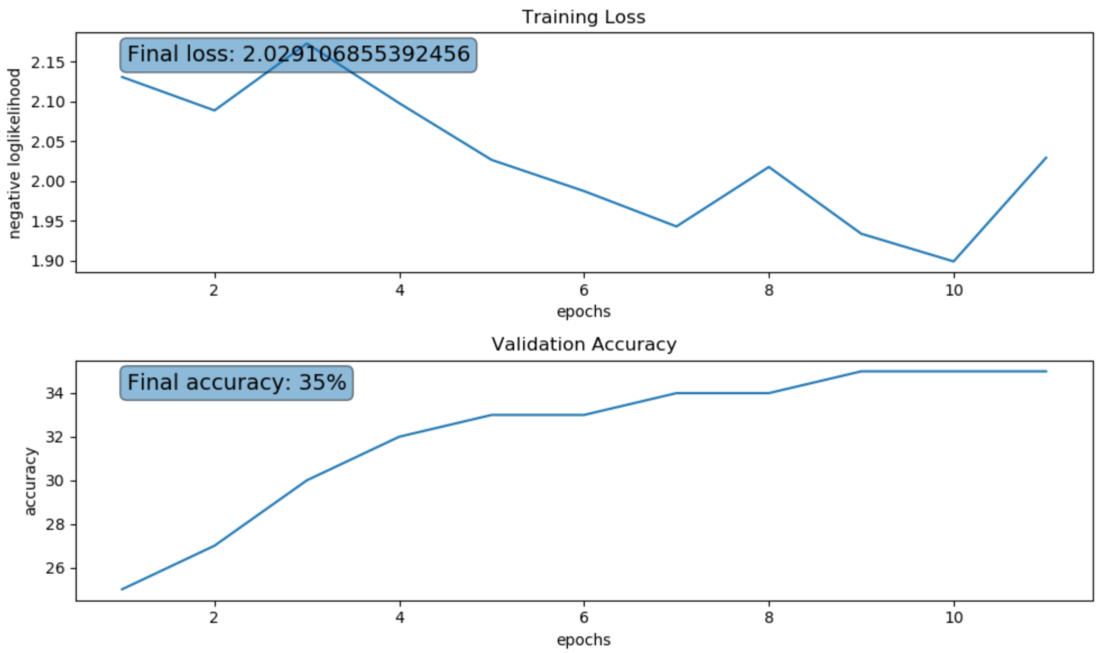

Figure 7. 2-Layer ReLU Activated, Learning Rate = 0.001, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.0

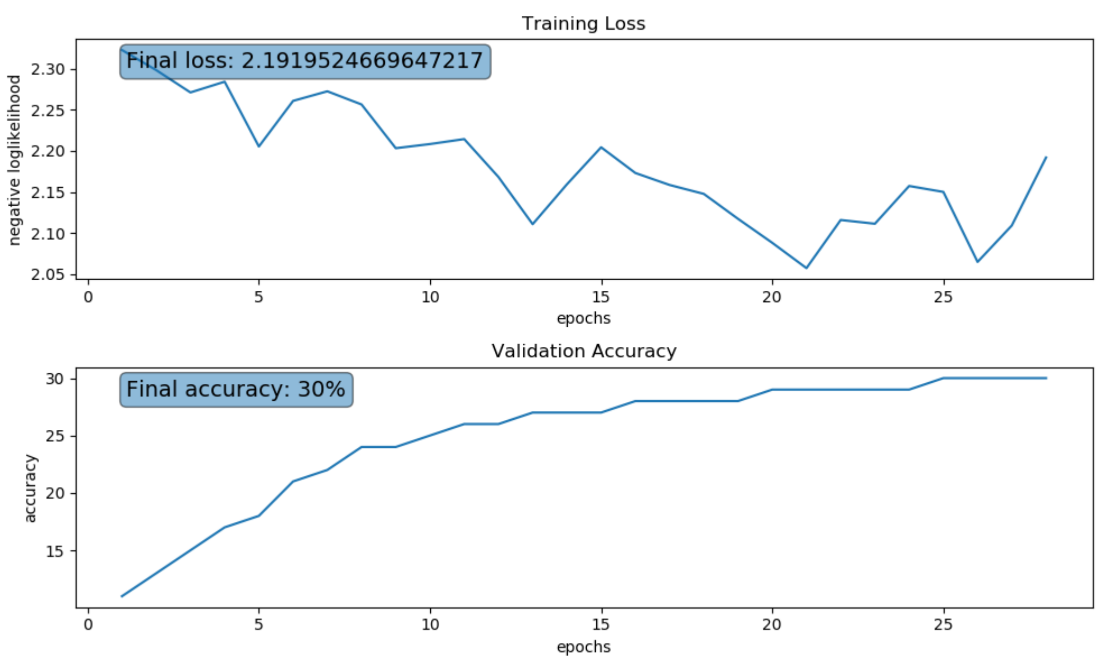

Figure 8. 2-Layer ReLU Activated, Learning Rate = 0.0001, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.0

3. I surveyed different parameters including dropout, momentum, and weight decay, using a learning rate of 0.01 as a constant for comparison's sake.

   The dropout parameter substantially improved performance when lowered to 0.1, with 0.9 performing quite bad as shown in figures 9 and 10, respectively. This tells me that low dropout is valuable to retaining information between epochs, which makes sense for a complex classification problem like CIFAR-10.

   Figures 11 – 13 show the effects of introducing momentum into our training. A high momentum resulted in very poor performance, however momentum of 0.1 converged on 49% validation accuracy. As previously stated, this classification problem is complex so it seems high momentum in the overshoots minima whereas there doesn't appear to be any benefit of avoiding local optima.

   Weight decay, like the other two parameters, performed best at a low value, but had the worst results of the three given its introduction as shown in figures 14-16. Notably, none of the nets trained for a significant number of epochs, so perhaps a longer training time would have allowed the loss the converge.

Figure 9. 2-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.1, Momentum = 0.0, Weight Decay = 0.0

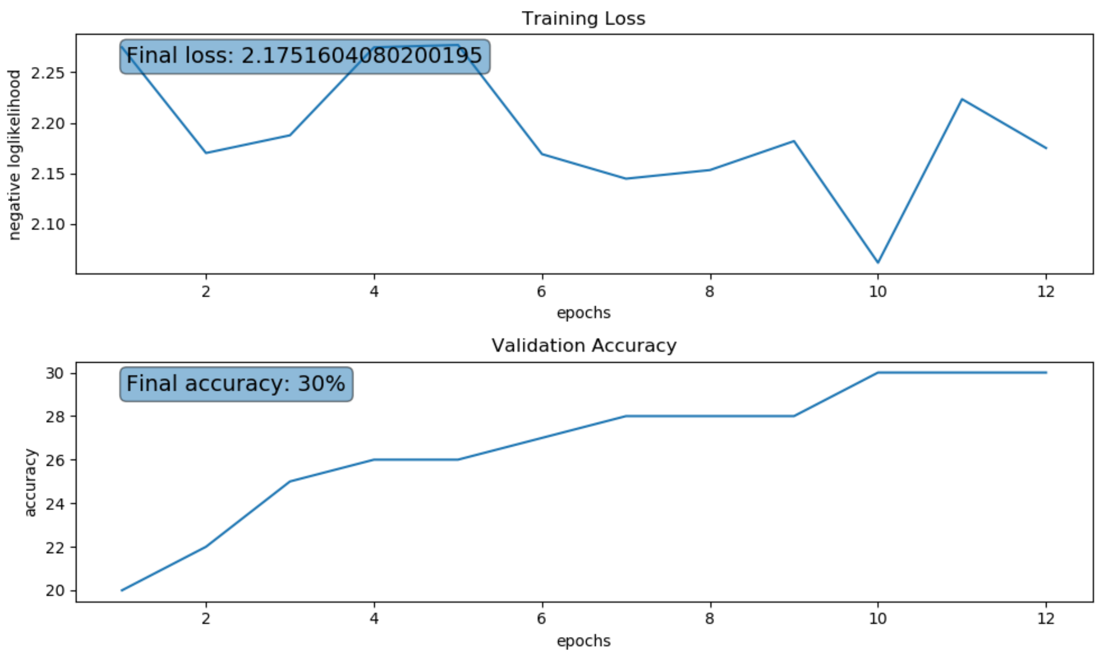

Figure 10. 2-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.9, Momentum = 0.0, Weight Decay = 0.0

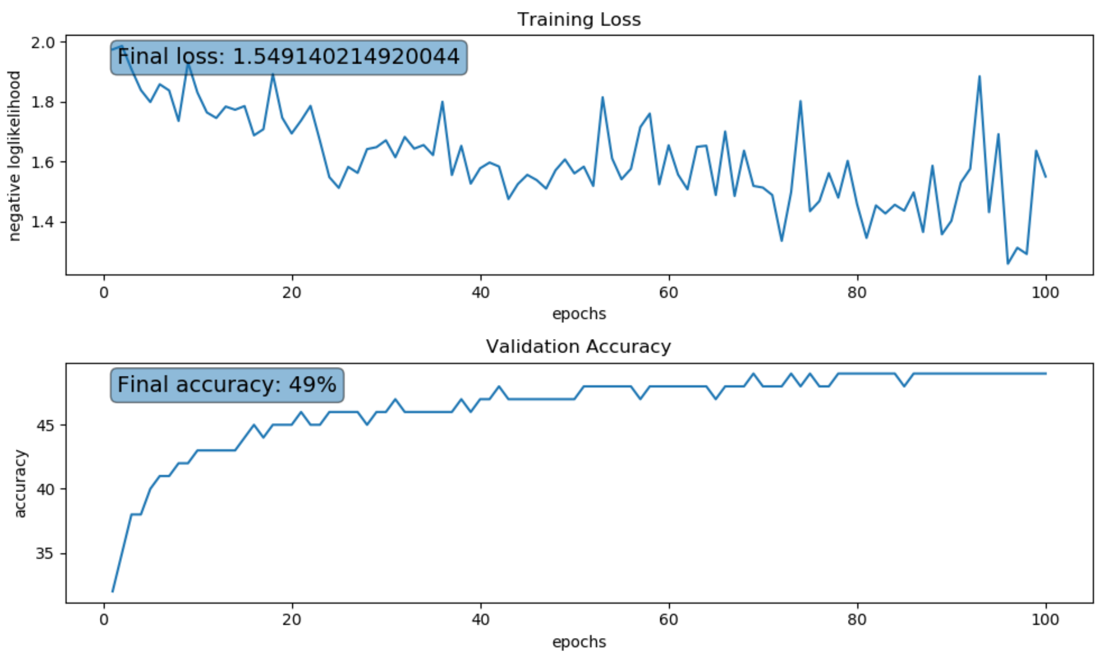

Figure 11. 2-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.1, Weight Decay = 0.0

Figure 12. 2-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.5, Weight Decay = 0.0

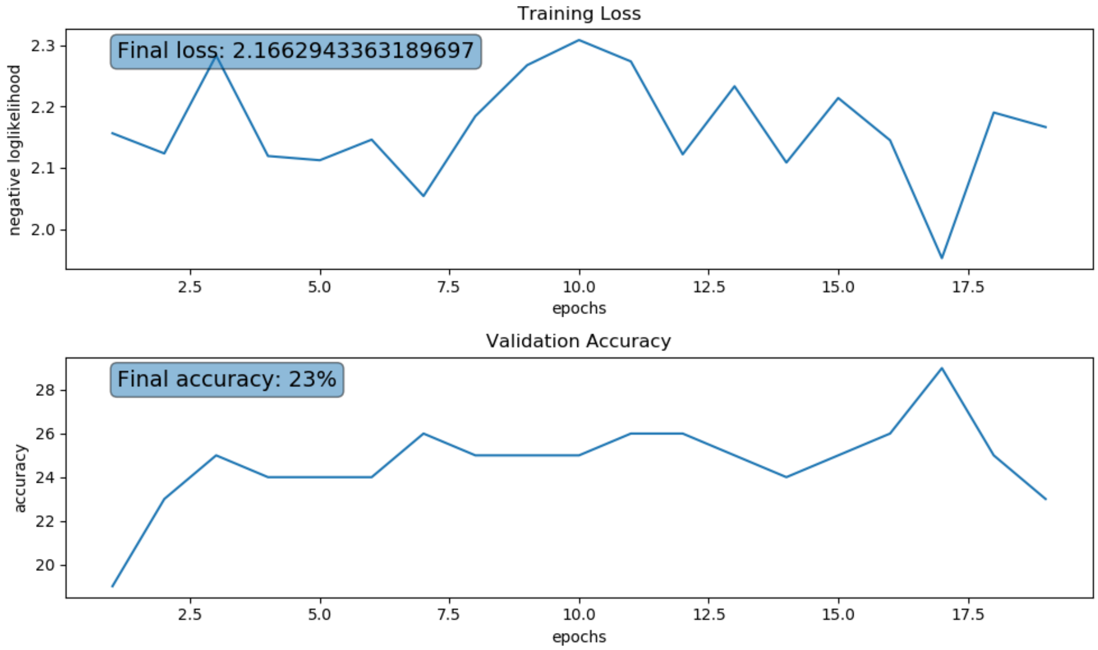

Figure 13. 2-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.9, Weight Decay = 0.0

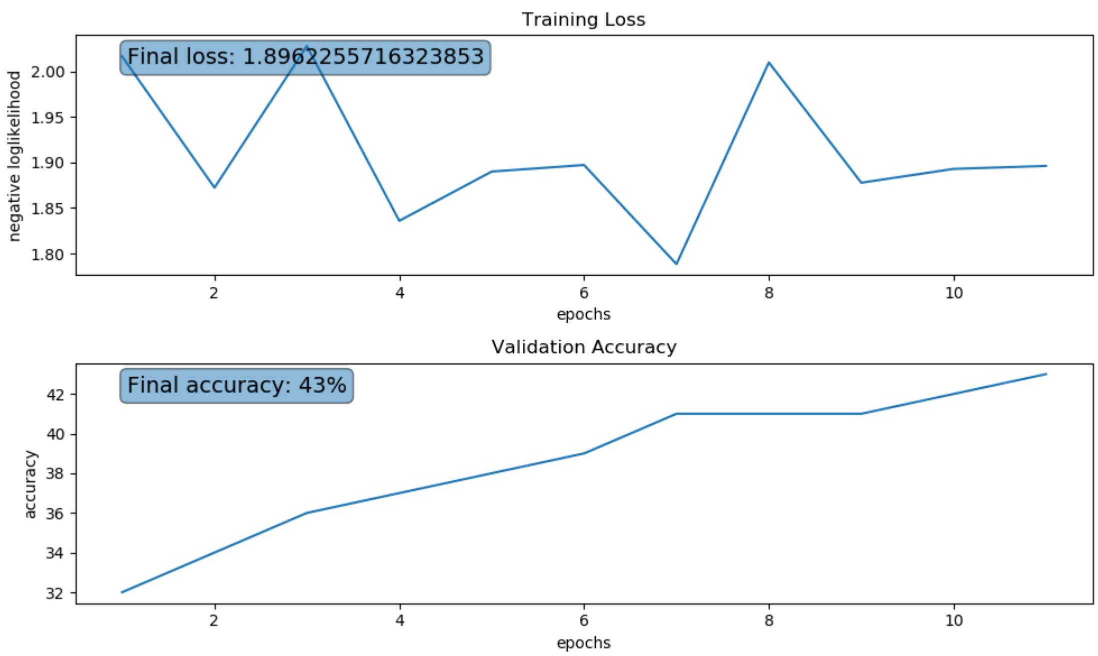

Figure 14. 2-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.001

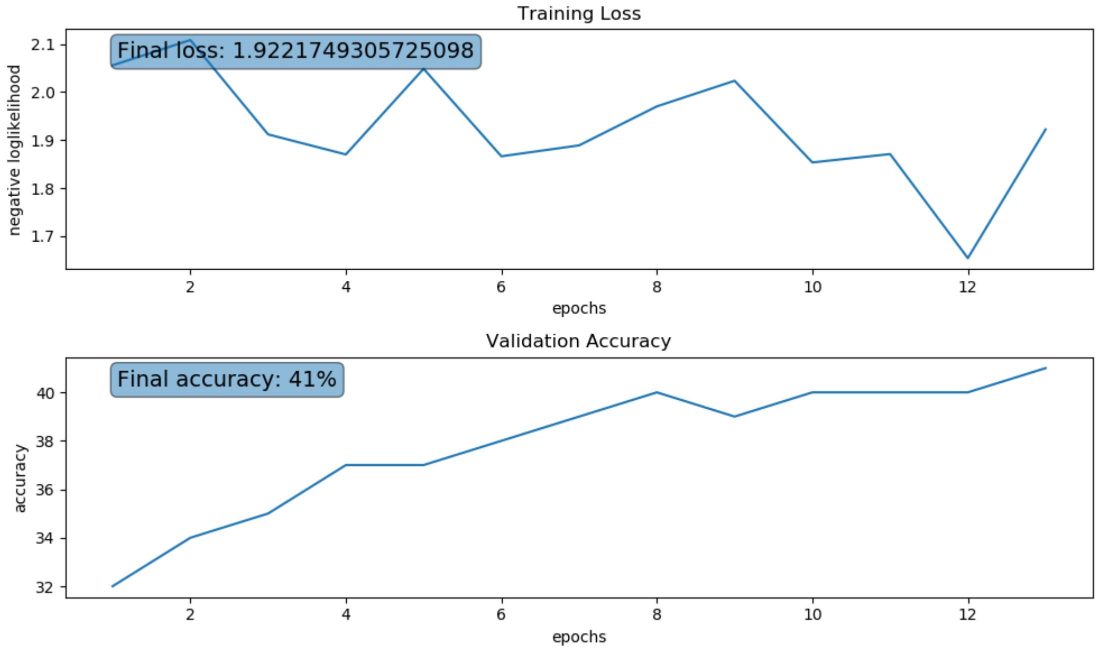

Figure 15. 2-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.01

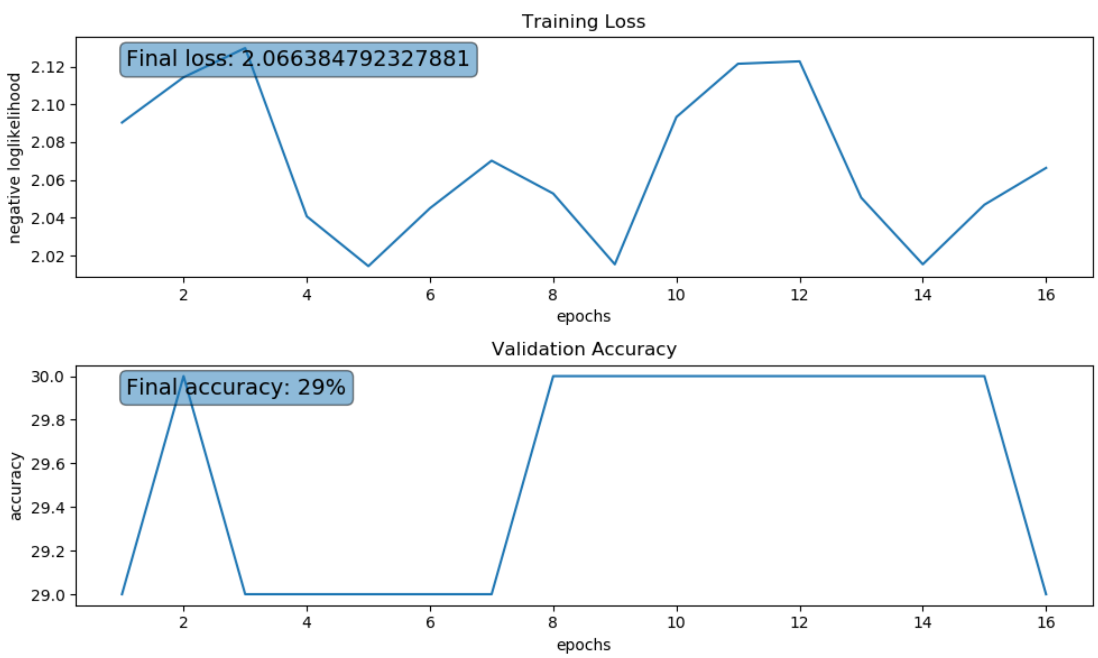

Figure 16. 2-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.1

4. Now we perform the same survey as above, but with a 3-layer ReLU activated network. Instead of a single 100-node hidden layer, we use two 50-node hidden layers with the hopes of increasing our network performance. Figures 17-24 show the training loss and validation accuracy results.

The patterns noticed previously between different values of parameters mirrors the results obtained in the 2-layer model, however there are two noticeable differences. First, my termination function tended to terminate much earlier, almost always under 20 epochs. Although in the graphs, the validation accuracy seemed to flatten out, it's difficult to determine whether the training loss could have continued to make progress. Secondly, the overall validation accuracy was lower, across the board in the 3-layer model. Although performance was worse, training convergence seemed overall better, which perhaps led to the earlier termination of training. Although the earlier termination meant training went much faster, it seems the trade-off is lower overall performance.

I attempted to obtain the best results with the 3-layer model by combining the best values of each of the tuning parameters and ended up with the results in figure 25, validation accuracy of 48% which mirrors the 2-layer model. This seems likely to be the limit of our rudimentary models. State-of-the-art methodologies are in the range of 75-96% accurate with much more sophisticated models.

Figure 17. 3-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.1, Momentum = 0.0, Weight Decay = 0.0

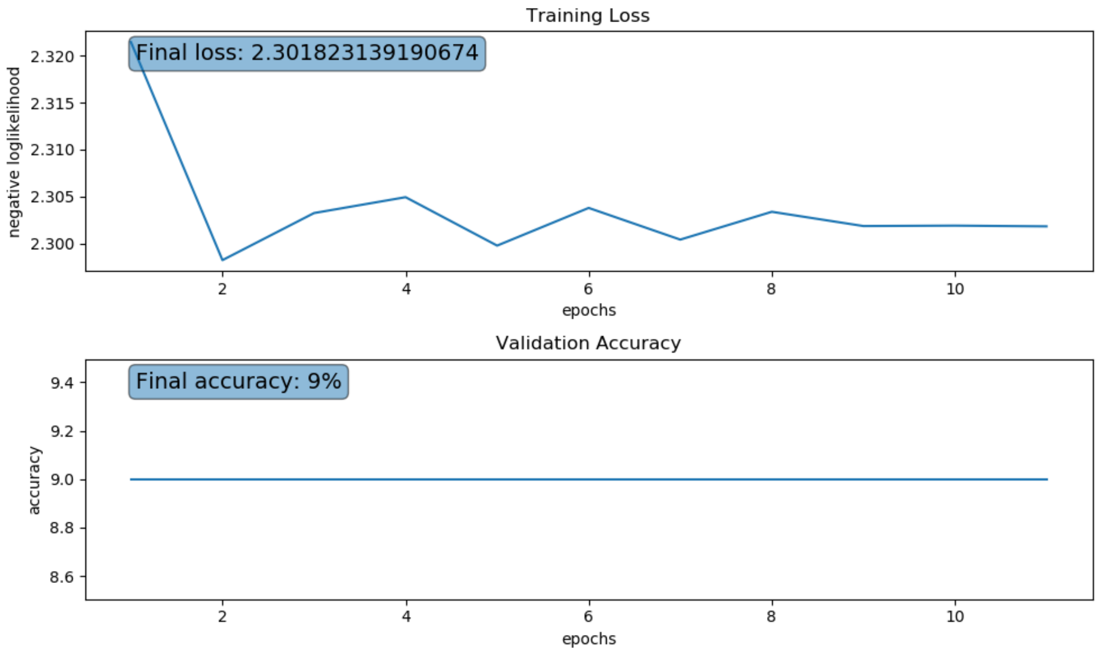

Figure 18. 3-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.9, Momentum = 0.0, Weight Decay = 0.0

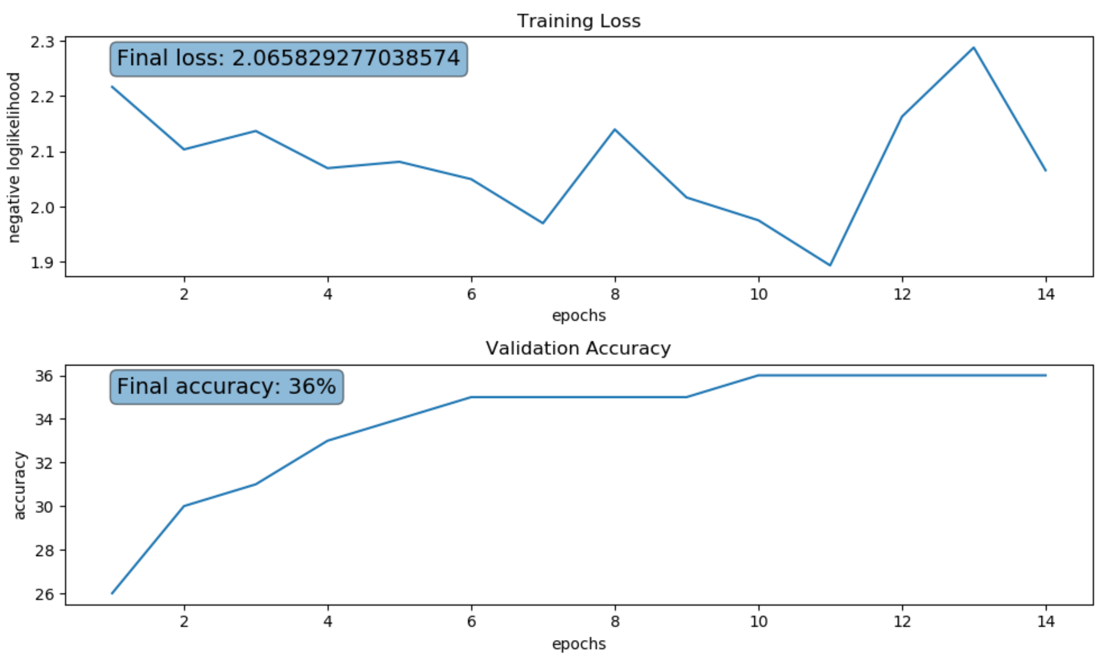

Figure 19. 3-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.1, Weight Decay = 0.0

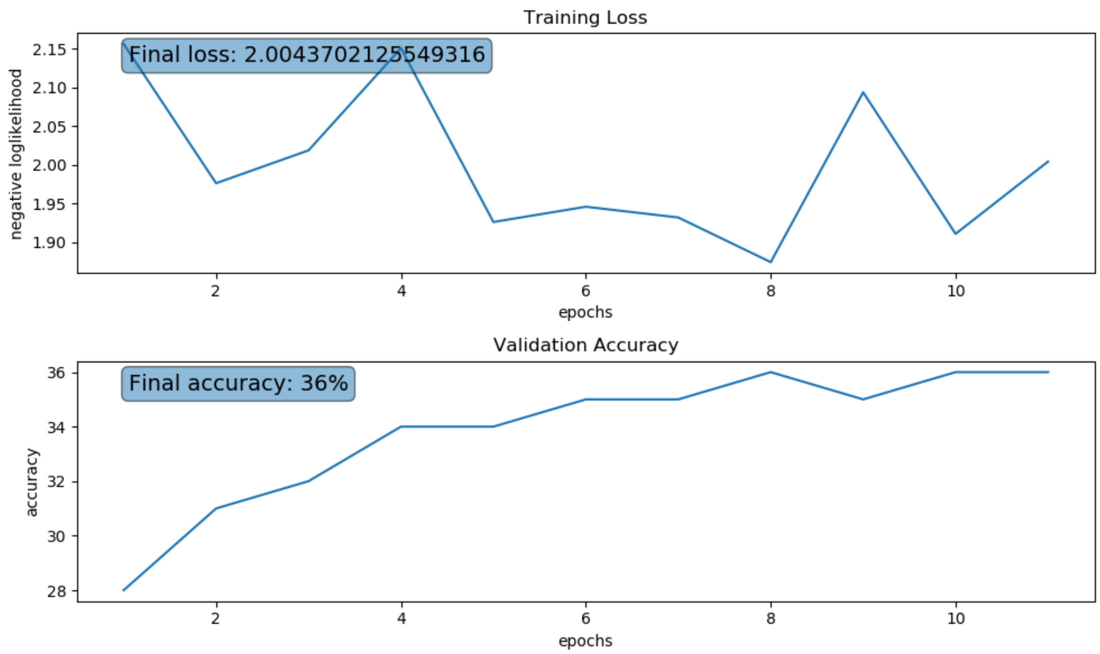

Figure 20. 3-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.5, Weight Decay = 0.0

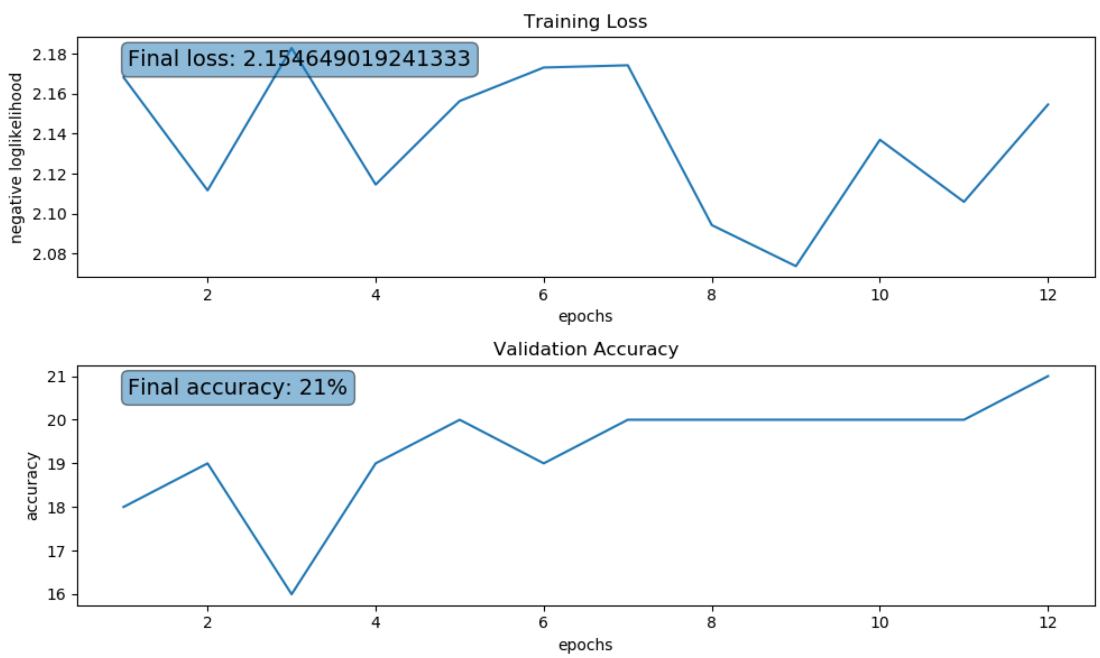

Figure 21. 3-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.9, Weight Decay = 0.0

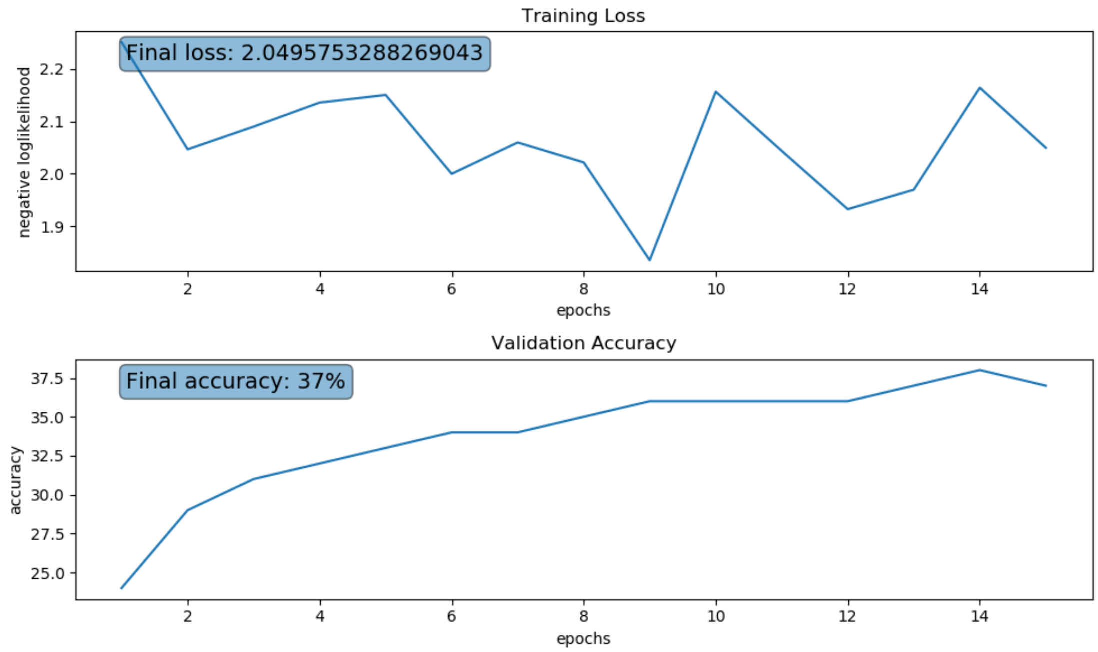

Figure 22. 3-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.001

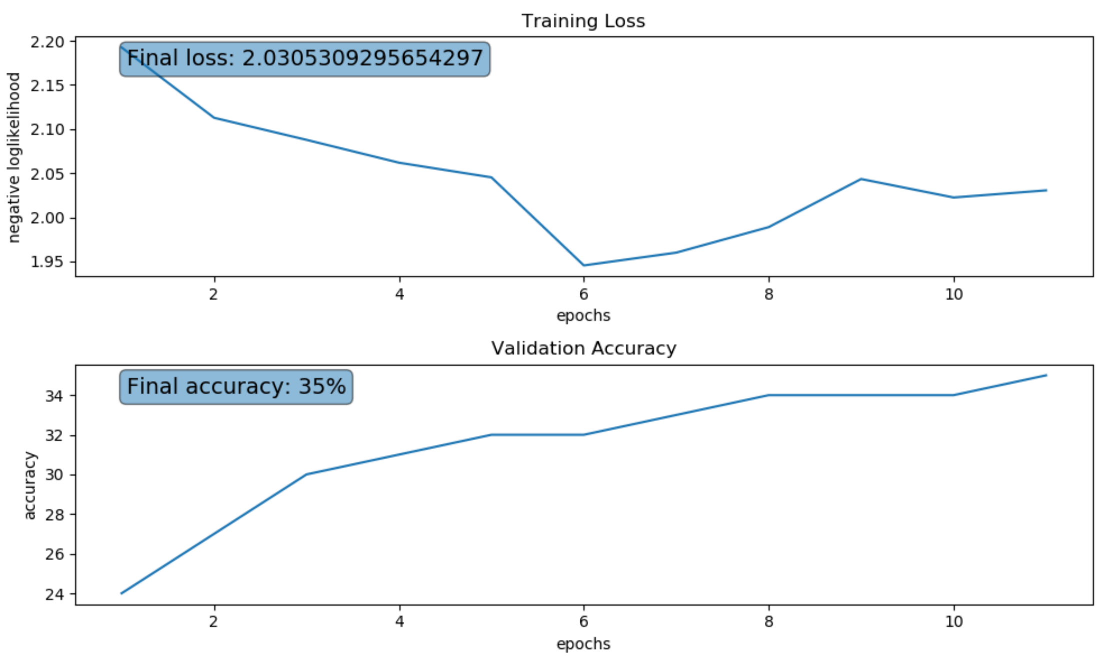

Figure 23. 3-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.01

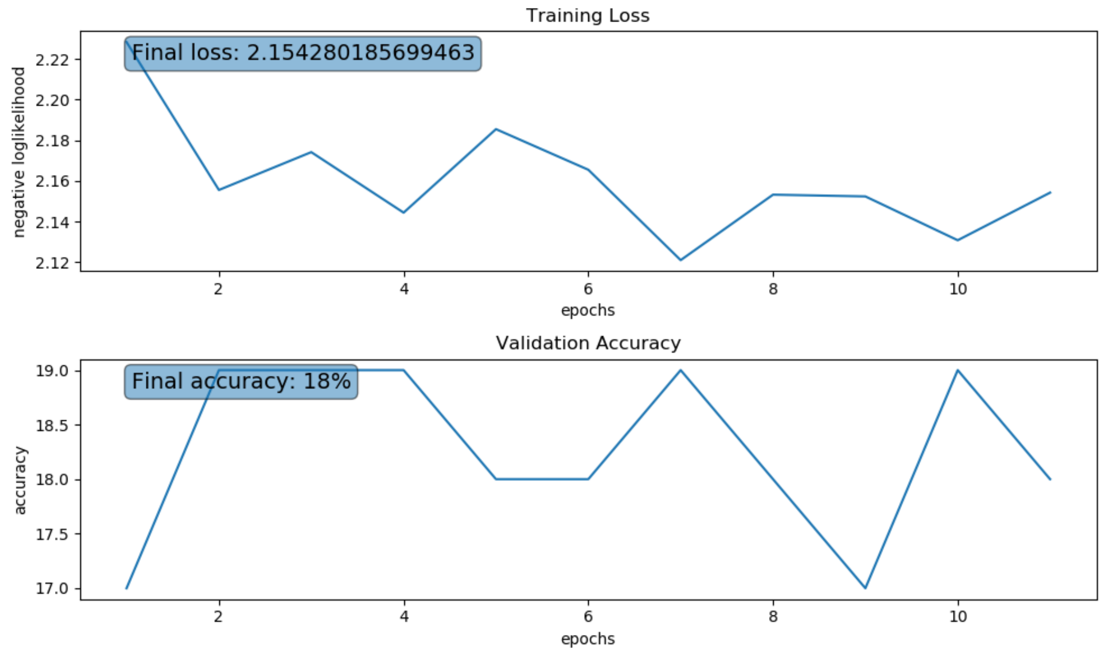

Figure 24. 3-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.5, Momentum = 0.0, Weight Decay = 0.1

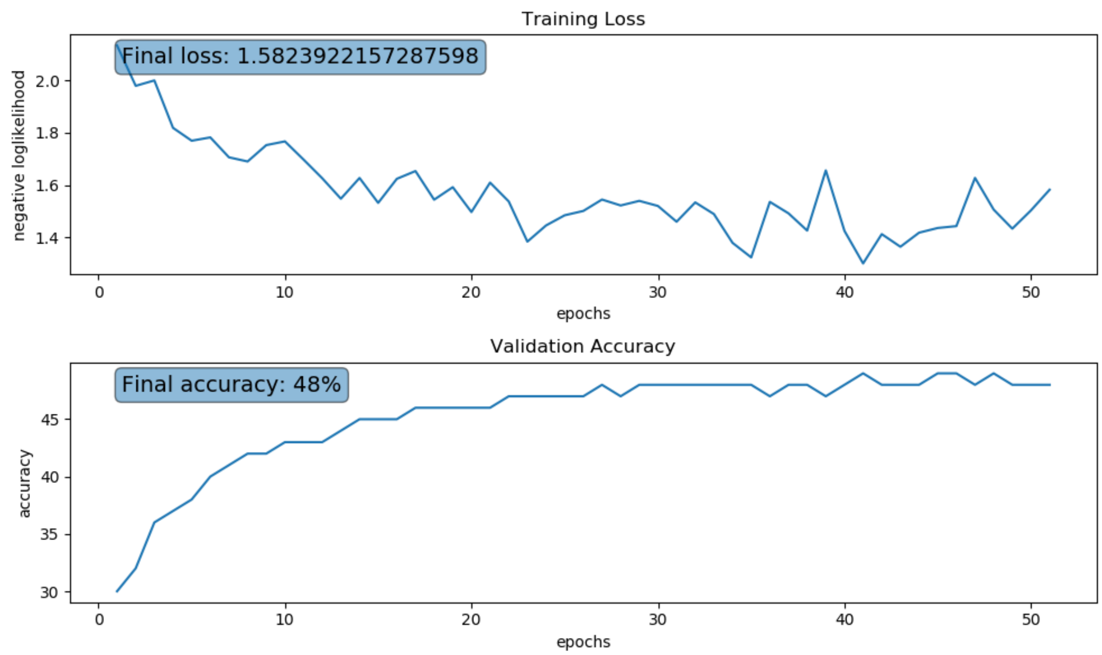

Figure 25. 3-Layer ReLU Activated, Learning Rate = 0.01, Dropout = 0.1, Momentum = 0.1, Weight Decay = 0.001