Xavier Robbins

Lab 9 : Conclusion

CS3450-031

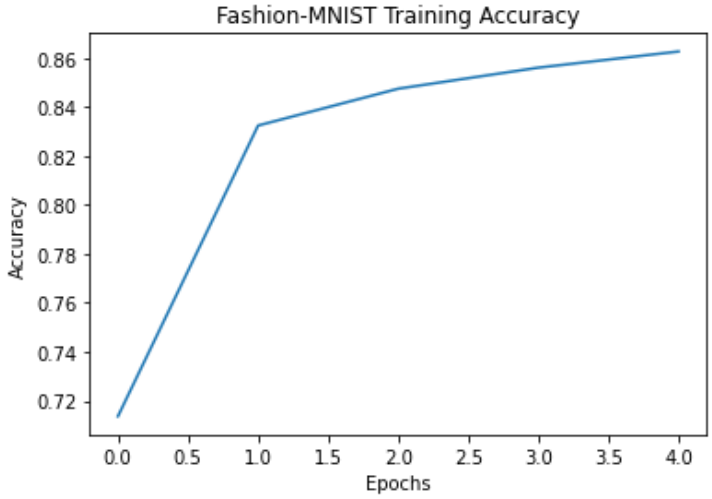
5.15.23

Fashion MNIST Training Attempts

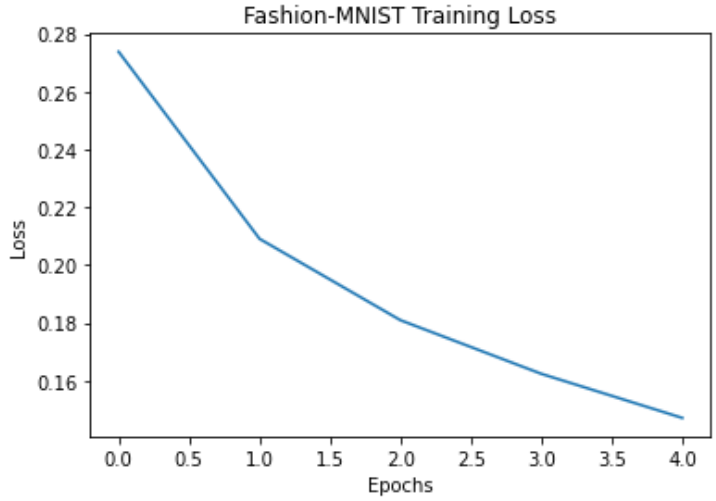
Run	# Epochs	Batch Size	Learning Rate	Regularization	Epsilon	Test Loss	Test Acc.
1	5	8	0.01	0.0001	1e-7	0.1744	81.0%
2	5	4	0.001	0.0001	1e-7	0.0291	85.3%
3	10	16	0.1	0.0001	1e-7	5.38	10%
4	10	8	0.01	0.0001	1e-7	0.0167	81.54%

All network architectures for the above attempts had one hidden layer with 256 neurons.

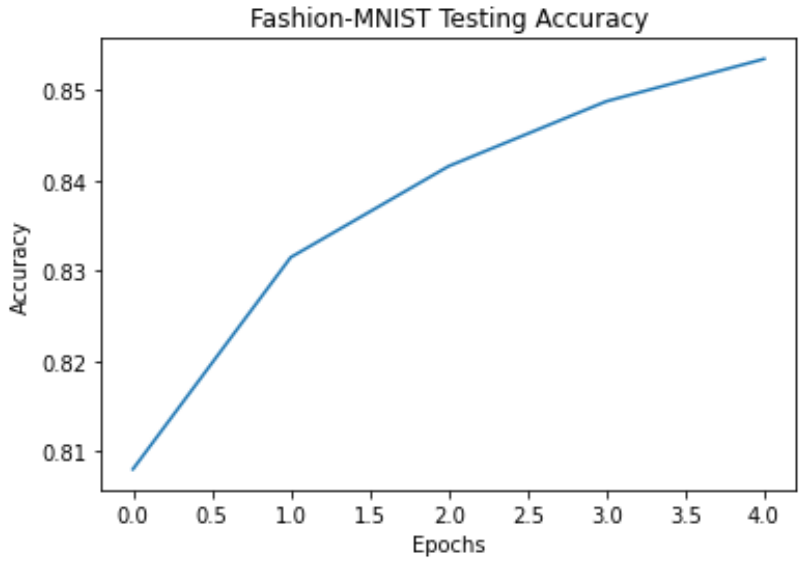
Fashion MNIST Training Accuracy (Run 2):



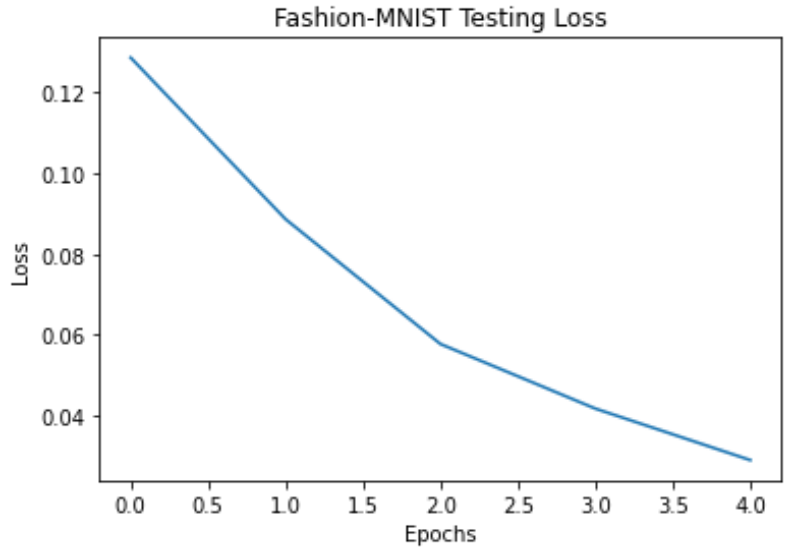
Fashion MNIST Training Loss (Run 2):



Fashion MNIST Test Accuracy (Run 2):



Fashion MNIST Test Loss (Run 2):

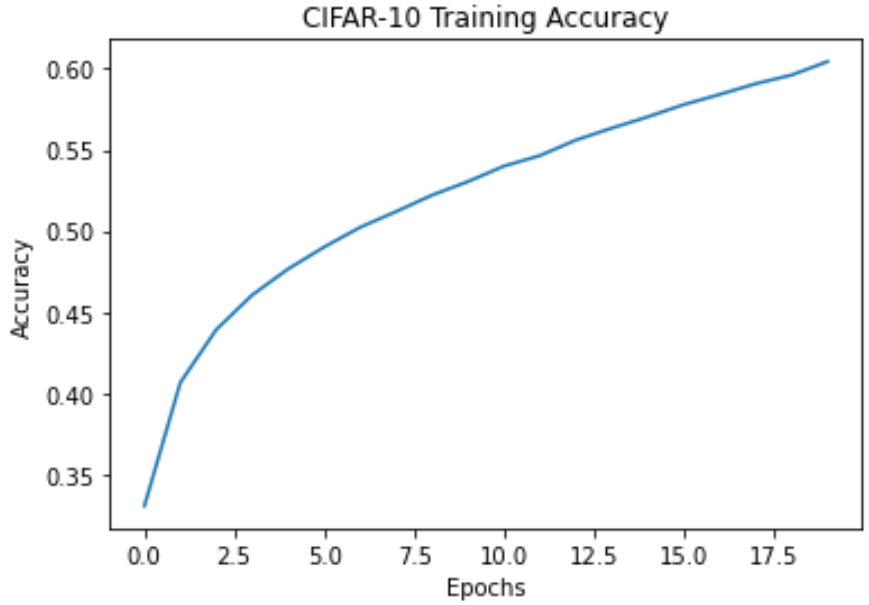


CIFAR-10 Training Attempts

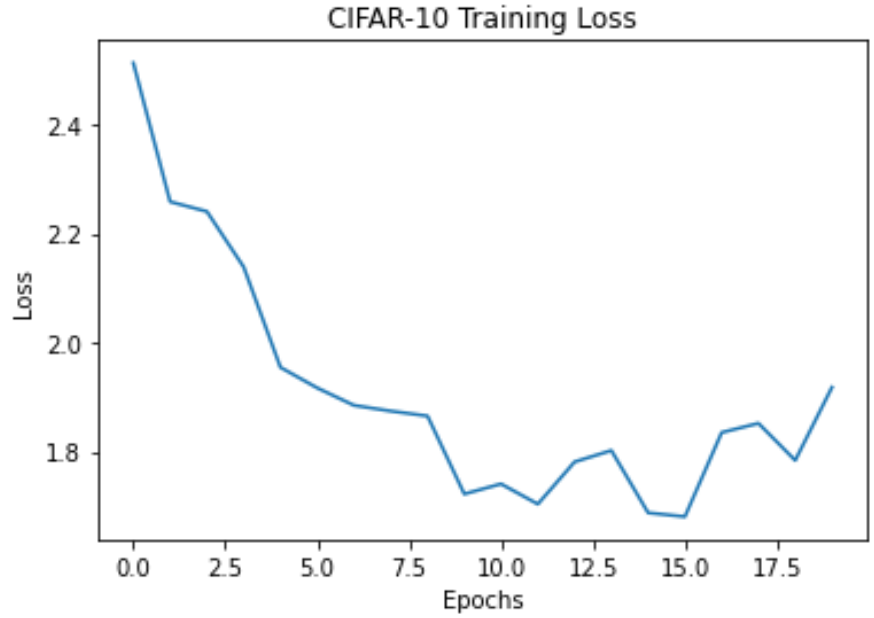
Run	# Epochs	Batch Size	Learning Rate	Regularization	Epsilon	Test Loss	Test Acc.
1	5	4	0.01	0.0001	1e-7	0.9800	45.6%
2	10	8	0.001	0.0001	1e-7	1.6284	47.7%
3	10	16	0.0001	0.0001	1e-7	4.1492	40.5%
4	20	8	0.001	0.0001	1e-7	2.15	49.2%

All network architectures for the above attempts contain one hidden layer with 512 neurons

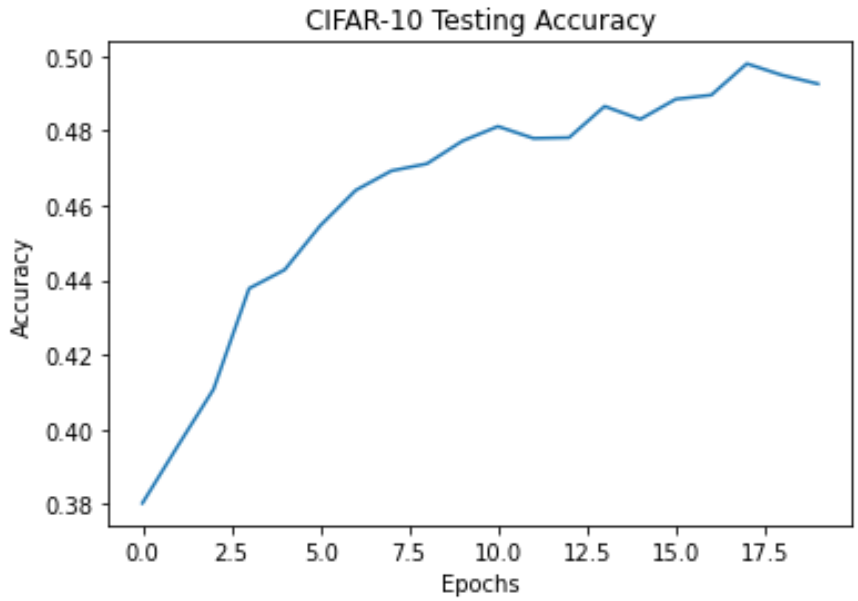
CIFAR-10 Training Accuracy (Run 4):



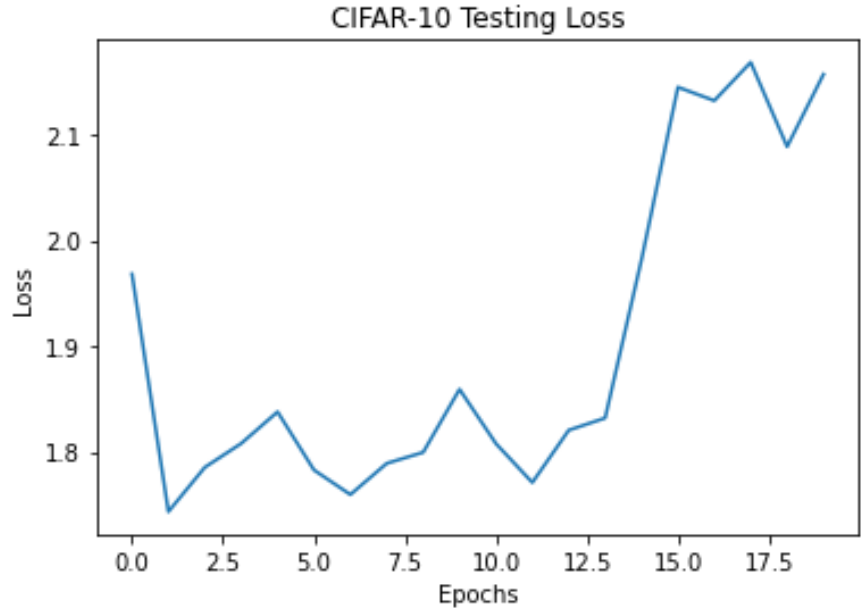
CIFAR-10 Training Loss (Run 4):



CIFAR-10 Test Accuracy (Run 4):



CIFAR-10 Test Loss (Run 4):



Conclusion:

Through training on these datasets I found that typically, more epochs and a larger batch size allows for better testing results after training. On the Fashion-MNIST dataset I was able to achieve an 85.3% accuracy and cross entropy loss of 0.0291. This was done using a batch size of 4 with a learning rate of 0.001 over 5 training epochs. With CIFAR-10 an accuracy of 49.2% with a cross entropy loss of 2.15. This was done using a batch size of 8 with a learning rate of 0.001 over 20 training epochs. The accuracy and loss curves for training and testing with the Fashion-MNIST seem to indicate a good fit over the training data as the loss and accuracy curves match closely for training and testing. The accuracy and loss curves for the CIFAR-10 dataset are not as nice, particularly the loss does not seem to be decreasing consistently through training and testing. Particularly in testing, in the last 10 epochs, the loss increases sharply. This behavior is apparent in the training loss but not as severely. This could indicate overfitting of the network in training, and a more optimal model could be achieved with fewer epochs.

This lab sequence was challenging, but ultimately rewarding. I particularly enjoyed this set of labs because I have tried to implement a neural network by hand myself before but have never successfully done it. I also feel that I gained a relatively solid understanding of the math that goes into forward and backward propagation in a network. Deriving the backpropagation equations was quite easy once I got the hang of it, and without taking into account batching I was easily able to understand the equations. Implementing individual layers and their corresponding tests was quite easy as well and I did not have many issues with this portion of the sequence. Debugging issues came into account when putting the network all together and training on the linear dataset. By not considering batching while deriving my backpropagation equations I had not used the proper operations in some of the layer's backpropagation equations. This resulted in shape errors when using batched learning in the network, and using matrix multiplication in place of an outer product multiplication in the linear layer backpropagation fixed my issues. Moving onto the Fashion-MNIST and CIFAR-10 I did not have any issues with debugging and was able to experiment with different hyper parameters to try and get the best accuracy possible.