Q2 – Learning Rate

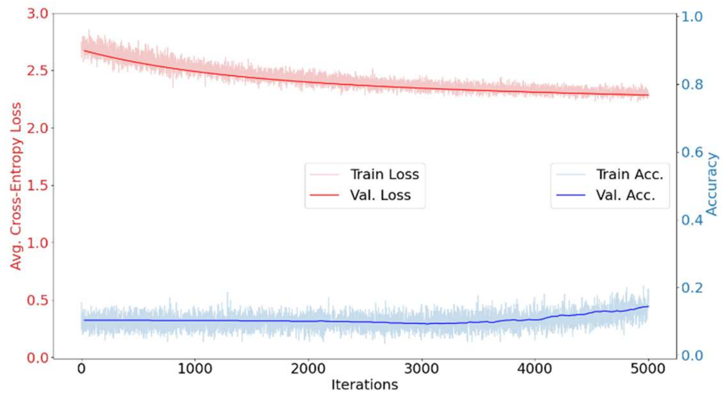

Figure 1: Step Size = 0.0001

Figure one has very consistently low accuracy and high loss. This is expected because the step size is very low, which means that the algorithm can't travel a meaningful distance down the gradient. I expect that if given more time (i.e. more epochs), the accuracy and loss would rise and fall respectively, just at very slow rates.

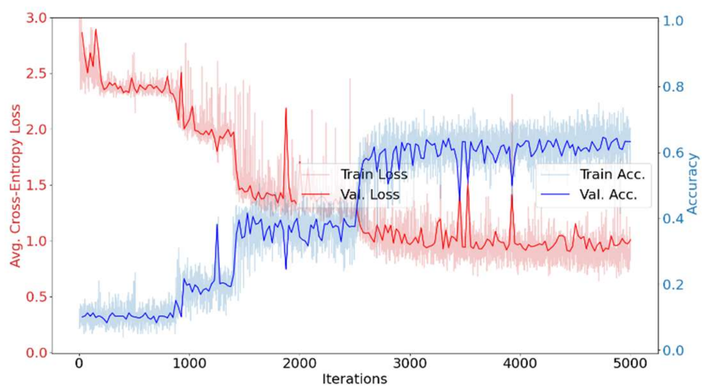

Figure 2: Step Size = 5

Figure two seems to have much more variance than the curve with default parameters. This is because the step size is slightly too high, causing the model to over- and under-guess.

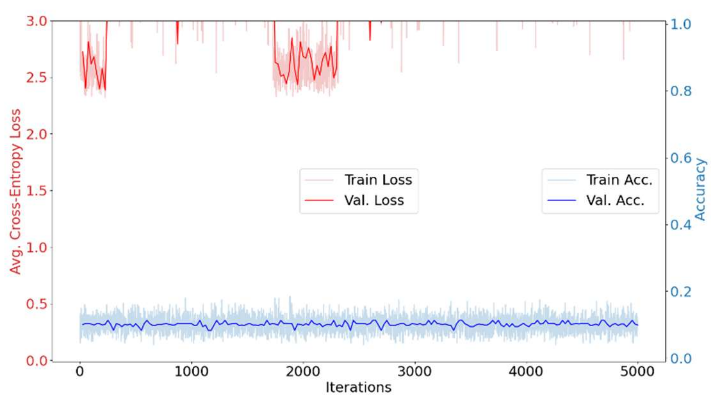

Figure 3: Step Size = 10

Figure three has extremely low accuracy and high loss. This is because the step size is much too high, making the model wildly inaccurate.

Q3 – ReLU’s and Vanishing Gradients

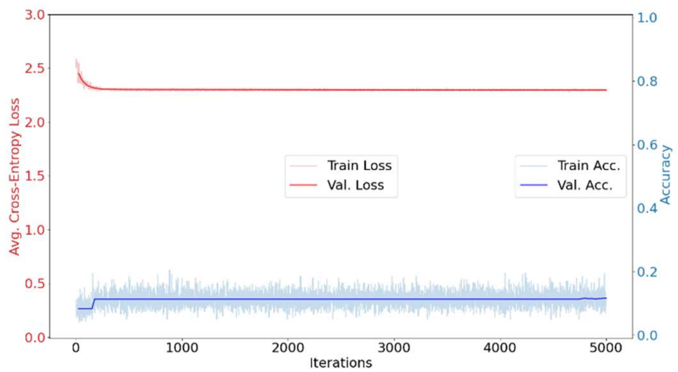

Figure 2: 5 Layer w/ Sigmoid

Figure four shows a 5-layer model with Sigmoid Activation. It seems to have similar issues as figure one where it isn’t traveling down the gradient quickly enough. This is because we increased the number of layers, making the gradient smaller in magnitude and learning slower.

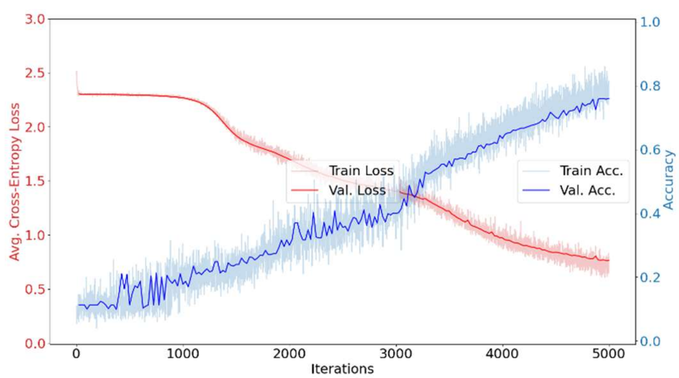

Figure 5: 5 Layer w/ Sigmoid Step Size = 0.1

Figure five has much quicker learning than the previous figure but still doesn’t match the quality of the curve with default parameters. Increasing the step size allowed the model to learn more quickly, making it more accurate.

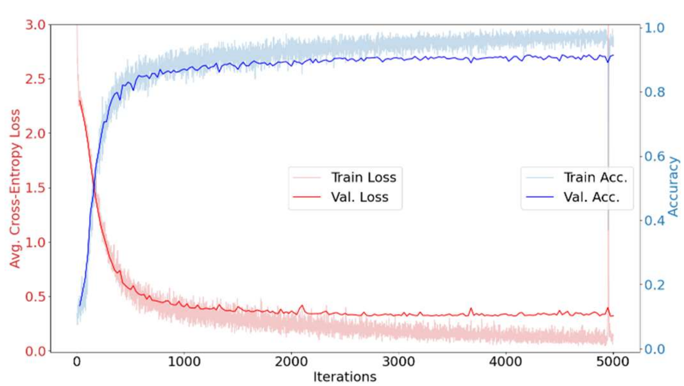

Figure 6: 5 Layer w/ ReLU

Figure six shows the best accuracy and loss yet. I believe this is because the derivative of the ReLU activation function is 1 if x > 0, so as we compute more layers, the gradient doesn’t get smaller and smaller until it vanishes.

Q4 – Measuring Randomness

Seed	Training Accuracy	Validation Accuracy
15024	87.4%	90.0%
0	85.14%	86.1%
99943503	87.66%	90.0%
4545061	86.43%	87.4%
495938403	86.86%	87.3%

This shows that there's some degree of error that will be related to random seeding. However, the difference seems to be marginal enough that my conclusions should still hold.