1.

$$\left(\frac{\theta 1}{\theta 0}\right)\frac{\prod_{i=1}^{d}\theta_{i1}^{xi}(1-\theta_{i1})^{1-xi}}{\prod_{i=1}^{d}\theta_{i0}^{xi}(1-\theta_{i0})^{1-xi}} > 1 \Rightarrow$$

$$log\left(\frac{\theta 1}{\theta 0}\right) + \sum_{i=1}^{d} log\left(\left(\frac{\theta i1}{1-\theta i1}\right)^{xi}(1 - \theta i1)\right) - \sum_{i=1}^{d} log\left(\left(\frac{\theta i0}{1-\theta i0}\right)^{xi}(1 - \theta i0)\right) \Rightarrow$$

$$log\left(\frac{\theta 1}{\theta 0}\right) + \sum_{i=1}^{d}\left(log\left(\frac{\theta i1(1-\theta i0)}{\theta i0(1-\theta i1)}\right)^{xi}\left(\frac{1-\theta i1}{1-\theta i0}\right)\right) \Rightarrow log\left(\frac{\theta 1}{\theta 0}\right) + \sum_{i=1}^{d} xi\left(log\left(\frac{\theta i1}{\theta i0}\right) > 0\right.$$

2.

$P(y = 1|X_1 = x_1, X_2 = x_2) = \frac{P(X_1=x_1|y=1)^2}{P(X_1=x_1|y=1)^2+P(X_1=x_1|y=0)^2}$ since P(y=1)=P(y=0) =>

$$\frac{1}{1+\left(\frac{P(X_1=x_1|y=0)}{P(X_1=x_1|y=1)}\right)^2}$$

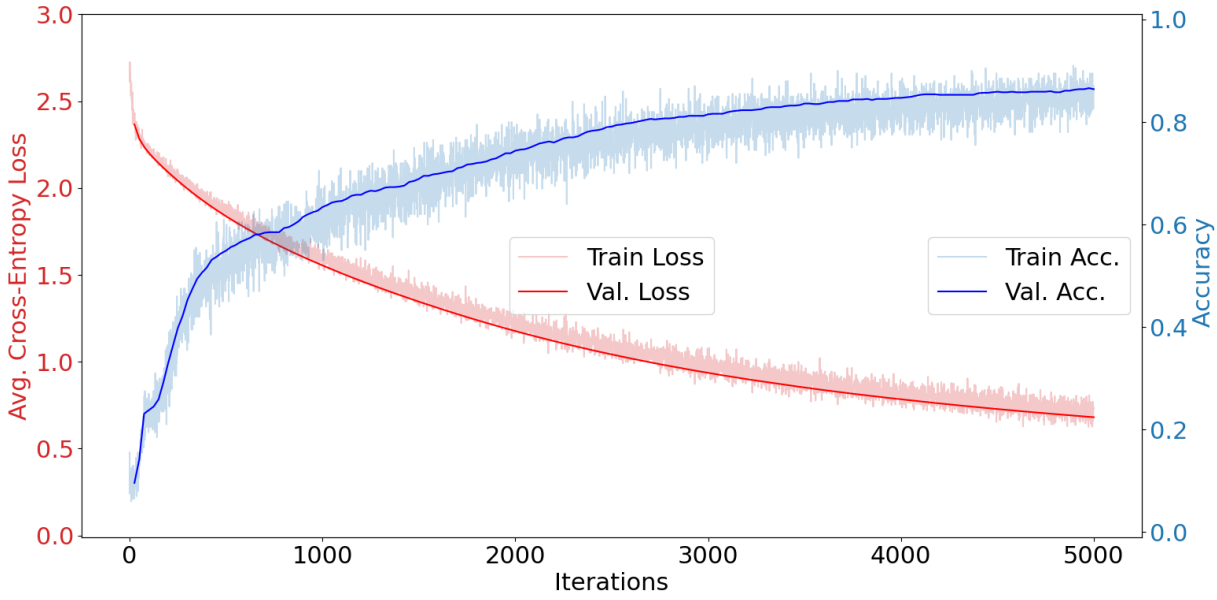
$P(y = 1|X_1 = x_1)$ = same stuff but without the added X_2 = $\frac{1}{1+\frac{P(X_1=x_1|y=0)}{P(X_1=x_1|y=1)}}$

3.

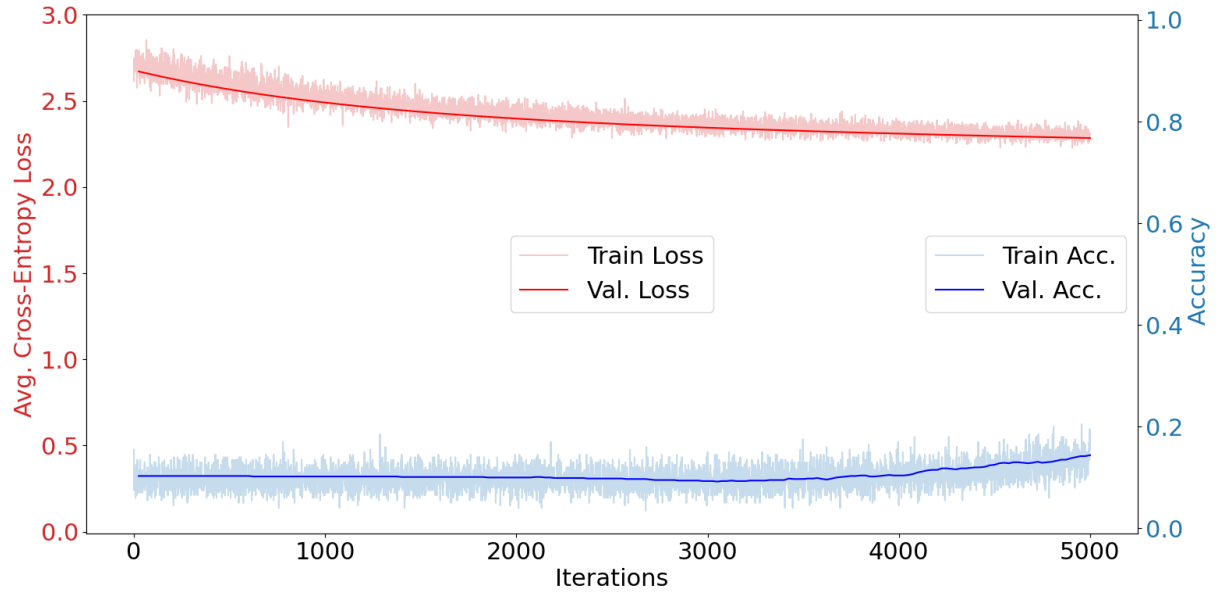
```
def backward(self, grad):
    self.grad_weights = (np.transpose(self.input) @ grad) # = dL/dW
    self.grad_bias = np.sum(grad, axis=0)                 # = dL/db
    self.grad_input = (grad @ np.transpose(self.weights)) # = dL/dX

    return self.grad_input
```

4.

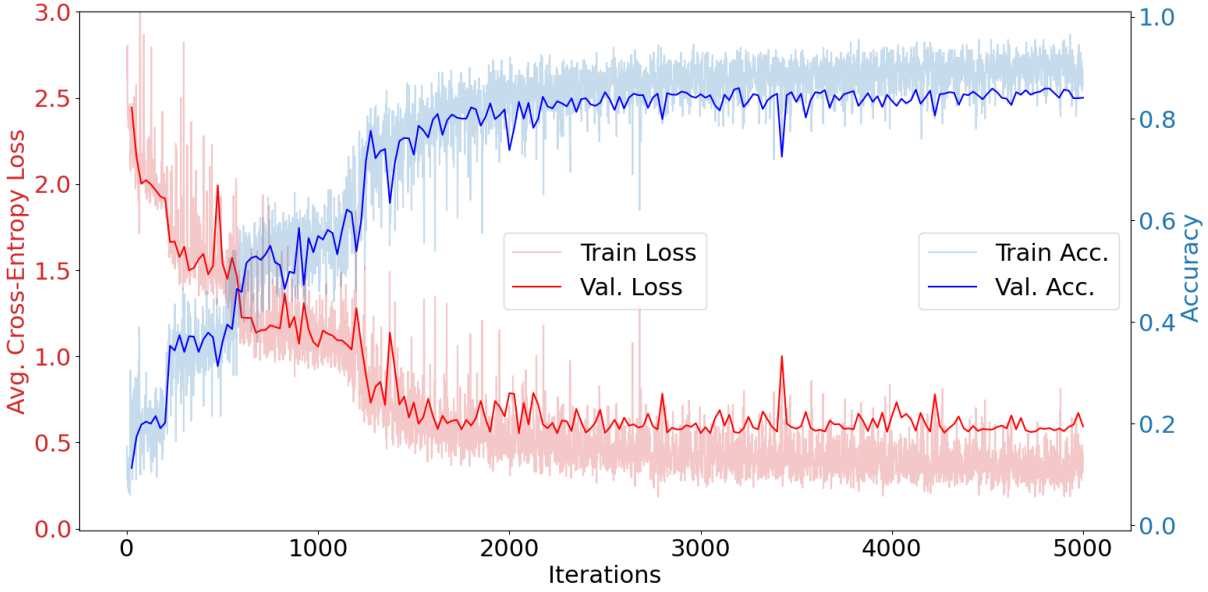
a. 0.01 Default



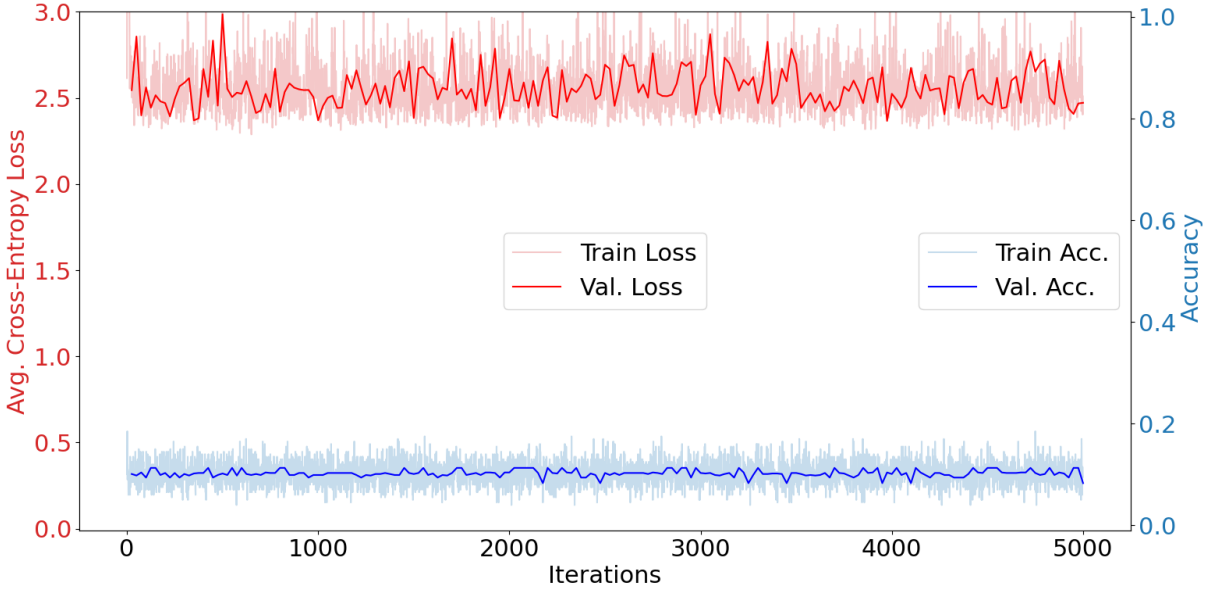
b. 0.0001



c. 5.0



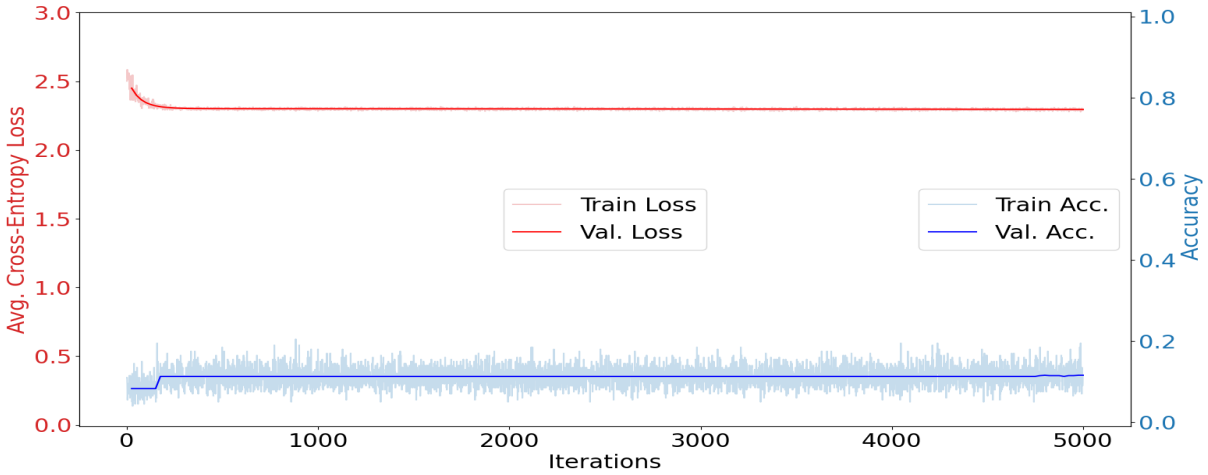
d. 10.0



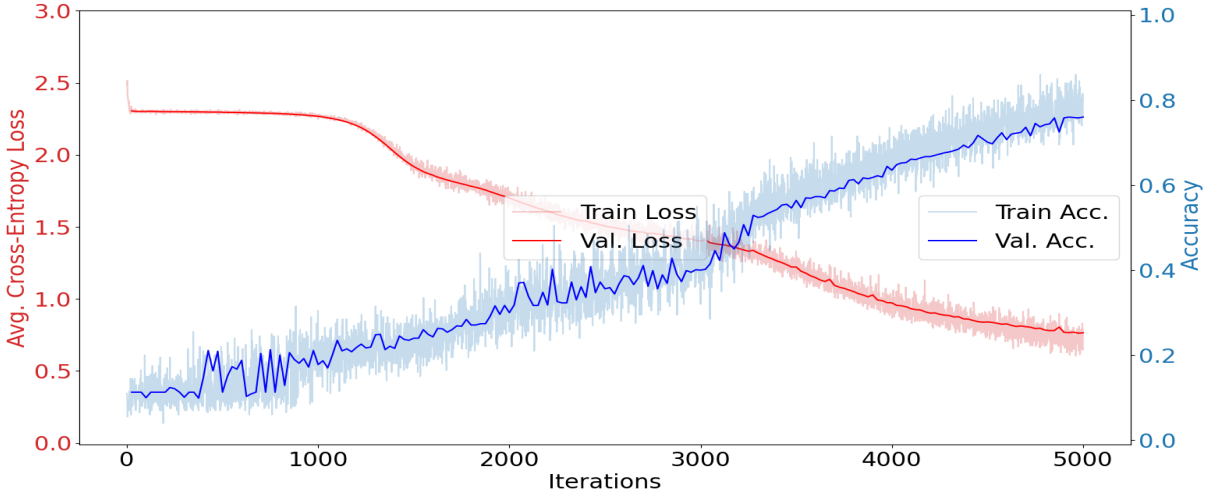
a. The training lines converge at a certain stepr range. Being separate and smooth with too small of steps while not crossing and being erratic which too high of steps.
b. There would be more iterations but the values would follow the same trends although the average should even out a little more.

5.

a. Sigmoid with default



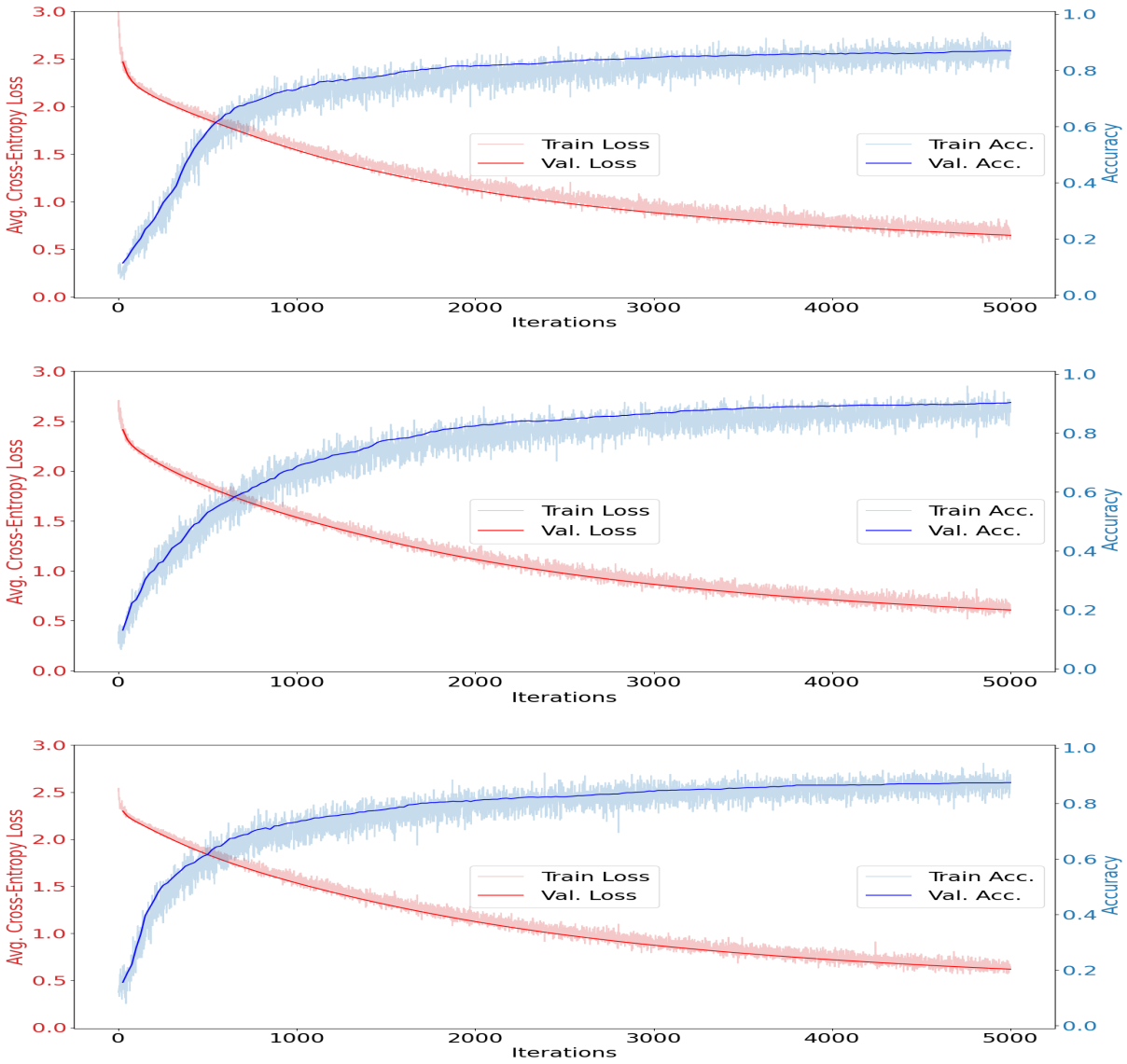
b. Sigmoid with 0.1 steps

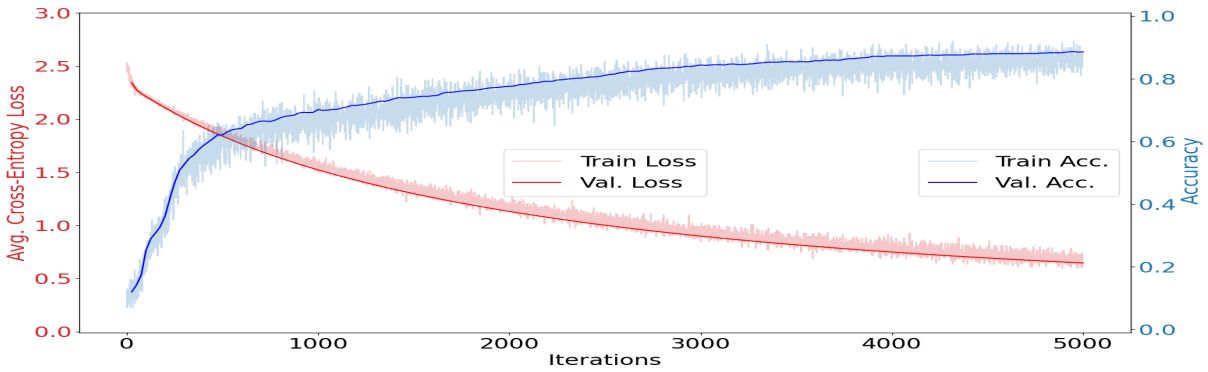


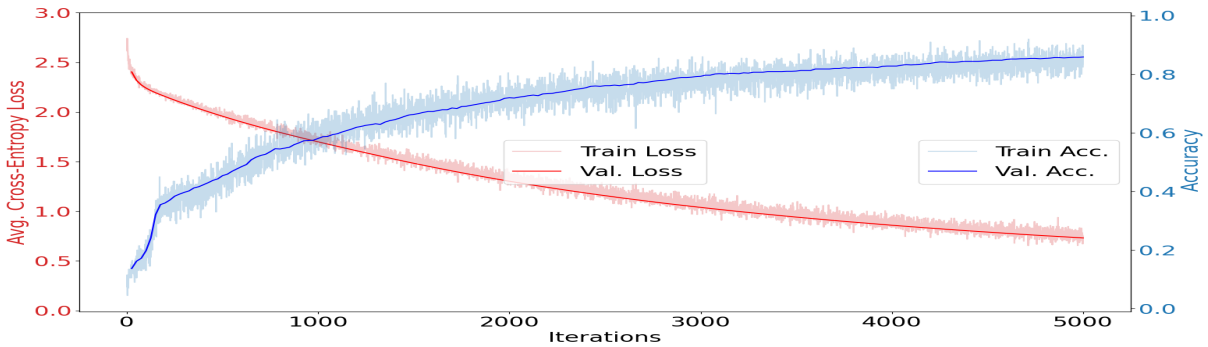
c. Relu with default

a. ReLU is almost like default although the curves cross earlier on. The Sigmoid has less loss variance but it has more accuracy variance. Similar to the previous comparison, it crosses at a certain range, steps too large and too small will not cross.
b. The steps were large enough to make variance in its tests to then increase its accuracy.
c. ReLU has more uniform weights which makes it faster and less prone to variance so it will describe the accuracy quickly instead of over multiple iterations like Sigmoid.

6.







a. It varies the first few iterations for accuracy until it starts to level off.

1. Approximately how many hours did you spend on this assignment?

11/15: 12:30pm - 2:00pm
5:20pm - 7:20pm
11/16: 7:30pm - 8:00pm
11/17: 4:20pm - 6:45pm
11/18: 4:00pm - 6:33pm

2. Would you rate it as easy, moderate, or difficult?

Easy, once I wrote Q3 and Q7, it was just testing.

3. Did you work on it mostly alone or did you discuss the problems with others?

Alone with a few questions to the professor over discord.

4. How deeply do you feel you understand the material it covers (0%–100%)?

I am not going to remember this for more than a few days, honestly.

5. Any other comments?

All the credit riding on these assignments is kinda stressful. I don't feel I can do the extra credits so my grade is difficult for me to maintain. Of course, if i was just a better student and doing more for this class then I wouldn't be struggling.