CS:434 → Homework 3

#1

$$P(y=1 \mid x_1 \dots x_d) = \frac{P(y=1) \prod^{d} P(x_i / y=1)}{P(x_1 x_2, \dots x_d)} = \theta_1 \prod^{d} \theta\, x_i\, (1-\theta_1)\; 1-x_1 \quad -①$$

$$P(y=0 \mid x_1 \dots x_d) = \frac{P(y=0) \prod_{i=1}^{d} P(x_i \mid y=0)}{P(x_1 x_2 \dots x_d)} = \theta_0 \prod_{i=1}^{d} \theta_i\, 0^{x_i} (1-\theta_i 0)\, 1 - x_i \quad -②$$

$$\frac{P(y=1 \mid x_1 \dots x_d)}{P(y=0 \mid x_1 \dots x_d)} > 1 \Rightarrow b + \sum_{i=1}^{d} w_i x_i > 0 \rightarrow Ⓐ$$

Consider...

$$\frac{P(y=1 \mid x_1 \dots x_d)}{P(y=0 \mid x_1 \dots x_d)} > 1 \Rightarrow \frac{\theta_1 \prod^{d} \theta\, x_i\, (1-\theta_1)\; 1-x_1 \quad -①}{\theta_0 \prod_{i=1}^{d} \theta_i\, 0^{x_i} (1-\theta_i 0)\, 1 - x_i \quad -②} > 1 \rightarrow \frac{\theta_1}{\theta_0} * \frac{\prod_{i=1}^{d} \theta_{i1}^{x_i} (1-\theta_{i1})^{1-x_i}}{\prod_{i=1}^{d} \theta_{i0}^{x_i} (1-\theta_{i0})^{1-x_i}} > 1 \Rightarrow \frac{\theta_1}{\theta_0} * \frac{\prod_{i=1}^{d} \left(\frac{\theta_{i1}}{1-\theta_{i1}}\right)^{x_i} (1-\theta_{i1})}{\prod_{i=1}^{d} \left(\frac{\theta_{i0}}{1-\theta_{i0}}\right)^{x_i} (1-\theta_{i0})} > 1$$

Take the log →

$$\log \left[\frac{\theta_1}{\theta_0} * \frac{\prod_{i=1}^{d} \left(\frac{\theta_{i1}}{1-\theta_{i1}}\right)^{x_i} (1-\theta_{i1})}{\prod_{i=1}^{d} \left(\frac{\theta_{i0}}{1-\theta_{i0}}\right)^{x_i} (1-\theta_{i0})} \right] > 0 \rightarrow \log\left(\frac{\theta_1}{\theta_0}\right) + \sum_{i=1}^{d} \log\left(\left(\frac{\theta_{i1}}{1-\theta_{i1}}\right)^{x_i} (1-\theta_{i1})\right) - \sum_{i=1}^{d} \log\left(\left(\frac{\theta_{i0}}{1-\theta_{i0}}\right)^{x_i} (1-\theta_{i0})\right) > 0$$

$$\rightarrow \log\left(\frac{\theta_1}{\theta_0}\right) + \sum_{i=1}^{d} \left[\frac{\frac{\theta_{i1}}{1-\theta_{i1}}}{\frac{\theta_{i0}}{1-\theta_{i0}}}\right]^{x_i} \left[\frac{1-\theta_{i1}}{1-\theta_{i0}}\right] > 0$$

$$\rightarrow \log\left(\frac{\theta_1}{\theta_0}\right) + \sum_{i=1}^{d} \left[\frac{\theta_{i1}(1-\theta_{i0})}{\theta_{i0}(1-\theta_{i1})}\right]^{x_i} \left[\frac{1-\theta_{i1}}{1-\theta_{i0}}\right] > 0$$

$$\rightarrow \log\left(\frac{\theta_1}{\theta_0}\right) + \sum_{i=1}^{d} x_i \left[\log \left[\frac{\theta_{i1}(1-\theta_{i0})}{\theta_{i0}(1-\theta_{i1})}\right] + \log\left[\frac{1-\theta_{i1}}{1-\theta_{i0}}\right]\right] > 0$$

$$\left.\begin{aligned} b &= \log\left(\frac{\theta_1}{\theta_0}\right) \\ w_i &= \log\left(\frac{\theta_{i1}(1-\theta_{i0})}{\theta_{i0}(1-\theta_{i1})}\right) + \log\left(\frac{1-\theta_{i1}}{1-\theta_{i0}}\right) \\ &= \log\left(\frac{\theta_{i1}(1-\theta_{i0})}{\theta_{i0}(1-\theta_{i1})}\right) \times \left(\frac{1-\theta_{i1}}{1-\theta_{i0}}\right) \end{aligned}\right\} \quad \begin{aligned} &\text{bias} \rightarrow \log\left(\frac{\theta_1}{\theta_0}\right) \\ &\text{weight} \rightarrow \log\left(\frac{\theta_{i1}}{\theta_{i0}}\right) \end{aligned}$$

#3

Implemented code ✓

#4

here are all the graphs →

figure 1: Step size → 0.0001

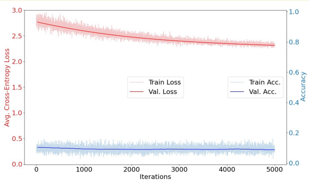


figure 2: Step size → 0.01

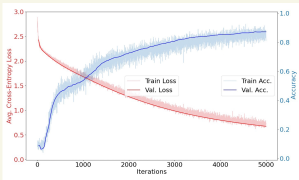


figure 3: Step size → 5

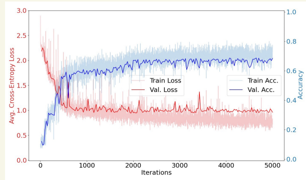


figure 4: Step size → 10

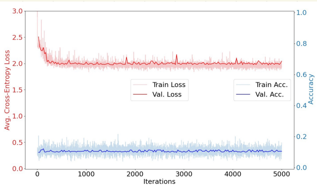
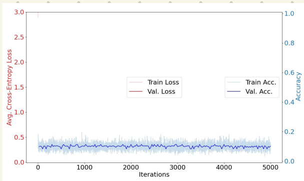


figure 5: Step size → 200



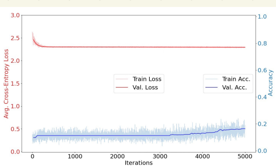
A) Figure 1 and Figure 4 look the most similar (0.0001 and 10).
Figure 2 and Figure 3 also look alike (0.01 and 5)
Figure 5 was a mistake I included.
I find it interesting that the step size from 0.01 to 5 intersects. On those two graphs the blue line seems to be increasing before retuning to a more constant state. Additionally, the red line seems to have a more drastic decrease on these two figures.
For figure 1, 4 and Five the blue line remainds reality the same.
For size 0.0001 and 10 the red lines are pretty similar, however for 10 the line is more jagged.
The jagged lines are pretty consistent in every line that has a step size of 5 and up.
The smaller the step size the smoother the line appears.

B) If the max epochs was to increase eventually this could cause over fitting and the model will not learn from the data. It is very important to find the right amount of epochs to get an optimal model. Too few causes under fitting.

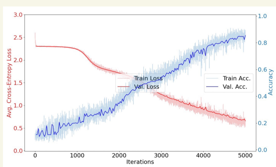
#5

here are all the graphs:

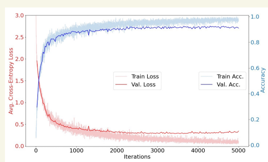
test 1:



test 2:



test 3:



A) the first test had the smoothest shape overall. Having 5-layer with sigmoid Activation represented the graph with 2 layers but a step size of ten. The only difference is the red line is smoother for this graph
The second test intersects to make an x. The training acc. (blue line) is the most bumpy and is very visably increasing in comparison to the other graphs. It is very intresting to see the train loss (red line) starts of pretty steady and as the line decreases it starts to osolate more.
Test 3 is like a sidways probla and is similar to the step size of 5 graph. For both lines of this graph they consistantly ocosiate a little and are not super smooth.
All of the graphs have slight negative trends for the train loss. test 2 is a more obious negative trend.

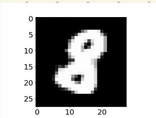
⟶

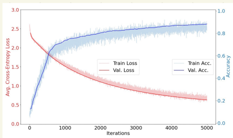
B) For test 2 train Acc. does experience a gradual increase, in comparison with test 1 the over all increase that is seen in test 2 is quiet drastic. The only differences between step 1 and 2 is that in step 2 the step size is now 0.1. From this we can conclude that the higher a step size is the more drautic of an incline will be seen on the graph.

C) Test 3 and Test 1 look wildly different. This is due to the fact that test 1 uses sigmoid and test 3 uses ReLU. ReLU allows training for deep neural nets, which is not possible with sigmoid-like activation functions.

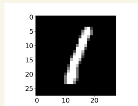
#6

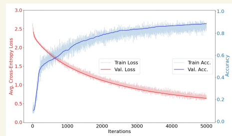
Seed 10



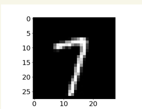


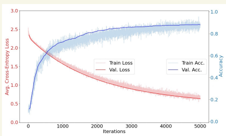
Seed 55



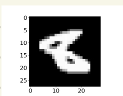


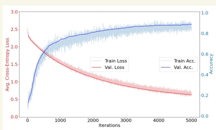
Seed 222



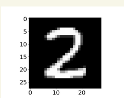


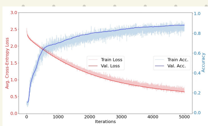
Seed 579





Seed 4141





When looking at all five random seeds I found the only things changing was the lines for the value acc. and value loss. The training looks to be consistant through out the five test.

Additionally all the numbers for the digitized data set changed with each test.

knowing that the seed has an impact on the line this much

#7

Username → Definitely not Alexis Marie Doyle

debrief :

1. 2 days
2. diff
3. alone
4. 63%
5. N/A