

4SL4 Assignment 4
Parthav Patel
400251342

Train/Test/Valid = 60/20/20%

Sigmoid was used as the activation function at all layers.

The model was trained with Stochastic Gradient Descent and early stopping.

- Every 10 epochs, the validation error is computed, and checked against the current lowest validation error. If there is no improvement, the training stops and the weights are returned. Lower intervals were tested, but this would cause the algorithm to stop too early.
- The number of epochs was capped at 500.
- The weights are initialized using `np.random.rand`. This provides random numbers in the range (0,1), so for a more accurate initial set of weights, 0.5 was subtracted to get a range of (-0.5,0.5)

N1/N2 hyperparameters:

- Each combination of $N1+N2 = 8$ was tested (1 and 7, 2 and 6, etc...)
- For each test, three different seeds/initial weights were chosen, from which the lowest validation error was selected.
- The combination of N1/N2 was chosen based on what provides the lowest validation error.

Table of training/validation errors:

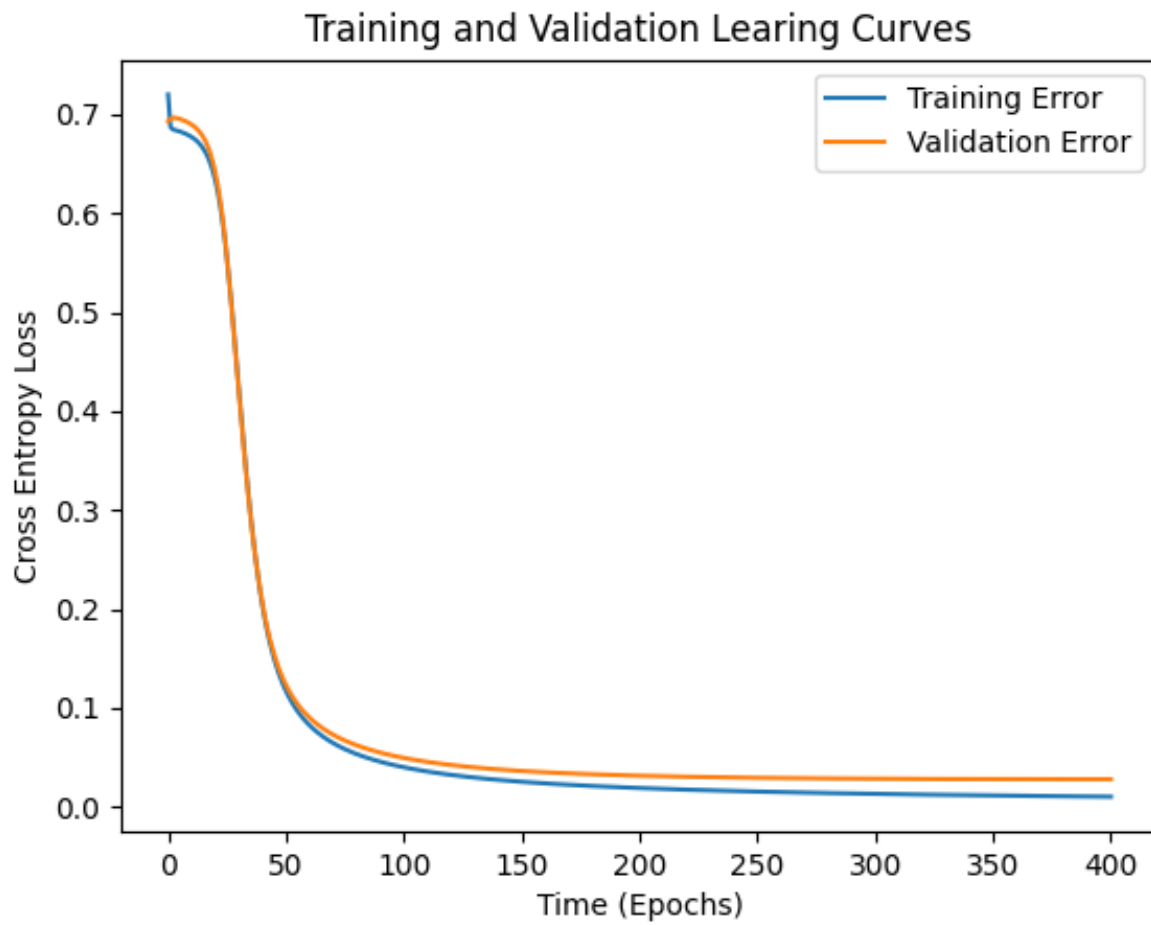
Seeds used: [1342, 4132, 2134]

	Training Error	Validation Error
N1=1, N2=7	0.009364584634729023	0.027127350840339348
N1=2, N2=6	0.005546714693098666	0.023575452043773076
N1=3, N2=5	0.009352468012167823	0.02648015264463669
N1=4, N2=4	0.003948665731527548	0.02723538901833994
N1=5, N2=3	0.009312024931630464	0.027097771119632394
N1=6, N2=2	0.006004882485575528	0.017207018723934795
N1=7, N2=1	0.009036371390962257	0.028654745141805524

Best Pair of N1/N2 for this model = 6/2, with a validation error of 0.017207018723934795

Final Model:

Learning Curves:



Seed = 1342, N1 = 6, N2 = 2

Misclassification Rates:

Training: 0.0

Validation: 0.00546448087431694

Test: 0.01090909090909091

Cross-Entropy Loss:

Training: 0.006004882485575528

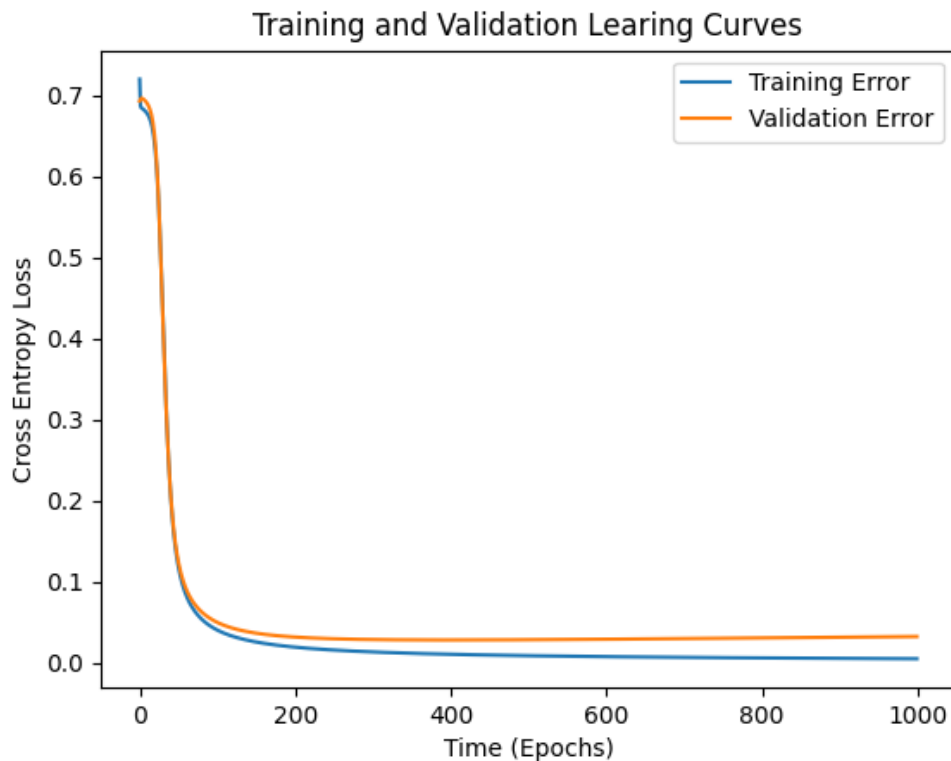
Validation: 0.017207018723934795

Test: 0.02930780228180852

Discussion:

The performance of this model is quite effective. This can be seen in the relatively low misclassification rates and cross-entropy losses on the test and validation sets. According to the learning curves above, due to early stopping, the algorithm was stopped just after 400 epochs, as we can start to see a rise in validation error. To further illustrate this, in the plot below, I've disabled the early stopping and increased the max number of epochs to 1000. We can then see a steady increase in validation error. Therefore, we can conclude that our early stopping mechanism successfully worked in mitigating overfitting. To further improve the model, I think it would be better to finetune the hyperparameters in our model. Cross-validation could be used to determine the optimal number of hidden units in each layer, as well as the learning rate, alpha. Additionally, different methods to prevent overfitting could be tested, such as dropout, dataset augmentation, model averaging, etc...

Learning Curve with 1000 max epochs and no early stopping:



Initial Weights used with seed = 1342:

W1:

```
[[ 1.      0.07159295 -0.21903136  0.47890573 -0.37647803]
 [ 1.     -0.36711676 -0.20076167 -0.45967896 -0.48585324]
 [ 1.      0.16734263  0.20915111 -0.37471133  0.19670391]
 [ 1.     -0.20091489 -0.01368432  0.43012584 -0.02198196]
 [ 1.      0.17665059 -0.06362748 -0.34446041  0.07997095]
 [ 1.     -0.05165596  0.14439173 -0.28934794  0.2586795 ]]
```

W2:

```
[[ 1.     -0.35730455  0.46474876 -0.1843708  0.45654161 -0.46875258 0.24779729]
 [ 1.     -0.10267337 -0.05012168 -0.38262639 -0.4115214  -0.19361772 -0.21882045]]
```

W1:

```
[[ 1.      0.06219012 -0.20013284]]
```

Final Weights after training:

W1:

```
[[ 7.25683788e-02 -1.16812048e+00 -1.16738695e+00 -8.67090687e-01 -9.60402607e-02]
 [-1.06521591e+00 -2.48843806e+00 -2.28866336e+00 -1.97527850e+00  9.63782304e-02]
 [ 1.51197969e+00  8.27613050e-01  6.32683405e-01 -5.23363699e-01 -8.04739742e-03]
 [-1.82623070e+00 -2.50915162e+00 -2.18724811e+00 -3.79906088e+00  6.12178801e-01]
 [ 1.06616099e+00  1.54402063e-01  1.42187061e-02 -5.05385624e-01  4.82814506e-02]
 [ 1.56805649e+00  1.07273223e+00  8.38010040e-01 -3.13893649e-01  6.70011855e-04]]
```

W2:

```
[[ 0.27926645  0.25808388  1.89828809 -1.02946475  2.16590809 -0.81977665 -0.8110795 ]
 [ 2.58445519 -1.71821855 -3.87167065  1.39436062 -5.43538065  0.15680835 1.88202485]]
```

W1:

```
[[ 2.70676981  4.21055117 -9.93196981]]
```