AssignmentReport-Group1

February 18, 2022

1 Assignment 1 Report

2 Task 1
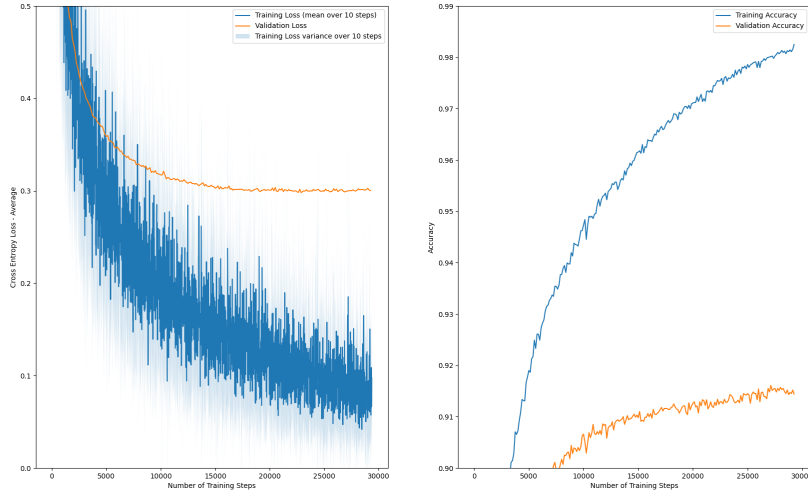
2.1 task 1a)

Task 1a:

$$\frac{\partial C}{\partial w_{ji}} \overset{\text{chain rule}}{=} \left(\frac{\partial C}{\partial z_j}\right) \cdot \frac{\partial z_j}{\partial w_{ji}} = \delta_j \cdot \frac{\partial}{\partial w_{ji}}\left(\sum_{i=0}^{I} w_{ji} x_i\right) = \delta_j x_i$$

$$\Rightarrow w_{ji} := w_{ji} - \alpha \frac{\partial C}{\partial w_{ji}} = \underline{w_{ji} - \alpha \delta_j x_i}$$
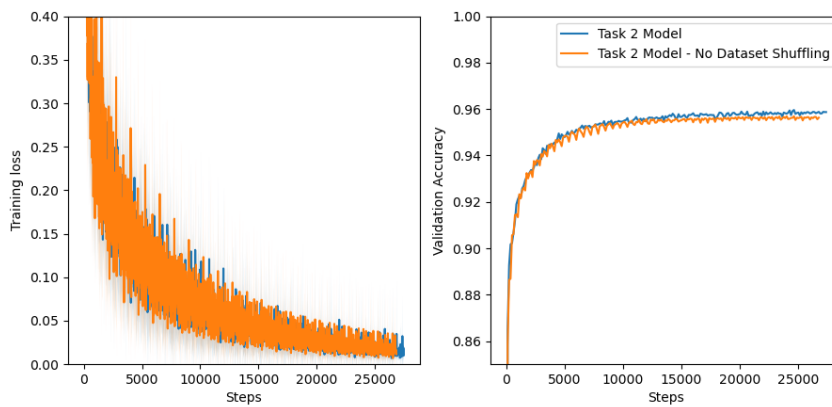
3 Task 2

3.1 Task 2c)

3.2 Task 2d)

There is a total of 50880 parameters in the network of which 50240 are weights between the input layer and the hidden layer and 640 are weights between the hidden layer and the output layer.
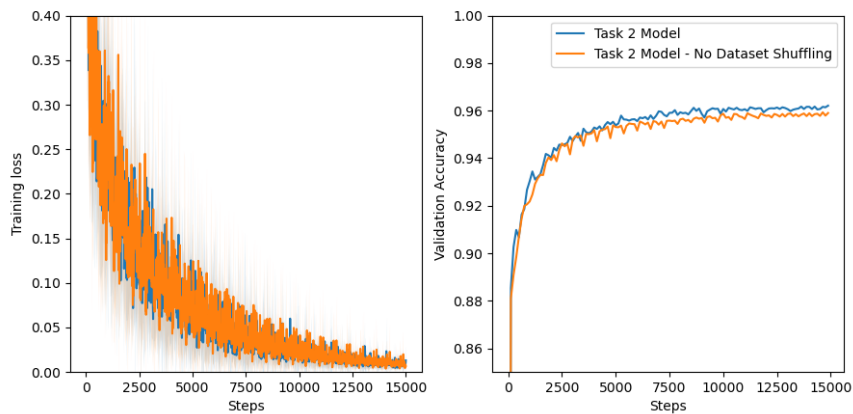
4 Task 3

4.1 Task 3a - Improved weight initialization

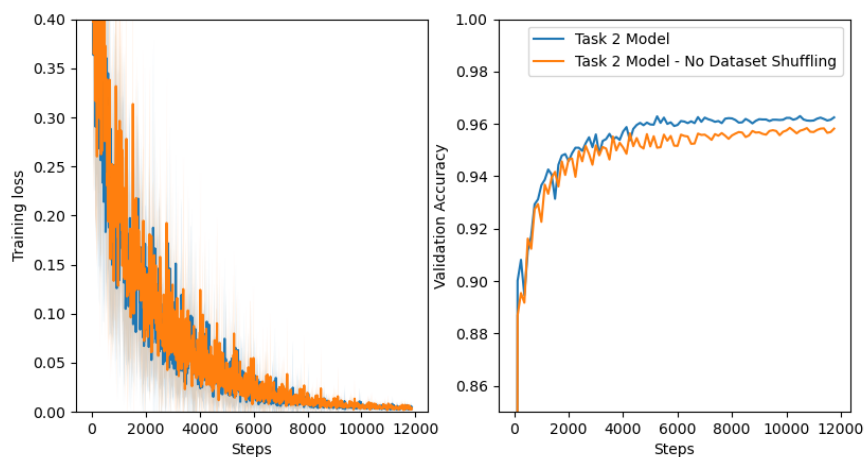Early stop after 43/44 epochs compared to 47.

4.2 Task 3b - Improved sigmoid function

Early stop after 24 epochs.

## 4.3 Task 3c - Momentum

Early stop after 19 epochs.

All three improvements significantly reduces convergance speed. With all three improvements early stop kicks in after 19 epochs compared to 47.

The improved weight initialization improves validation accuracy by a great deal from about 0.915 to almost 0.96. The improved simoid function seems to improve the validation accuracy with shuffeling just a tiny bit, and the momentum improves it some more. With the last improvement we also see a clearer difference between the acuracy with and without shuffeling, which makes sense because the momentum makes a connection between successive batches, thus reducing generality if not shuffeled between each epoch.
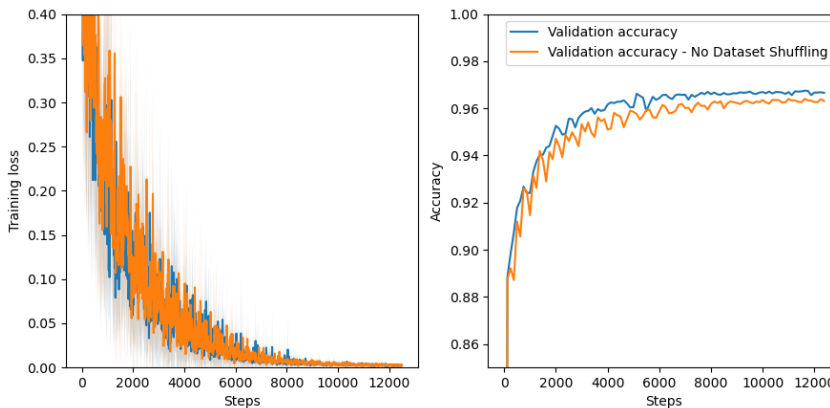
5 Task 4

5.1 Task 4a)

To few hidden units (32) reduces the learning speed (faster iterations and convergence after 16/17 epochs), but also the accuracy down to 0.94-0.95
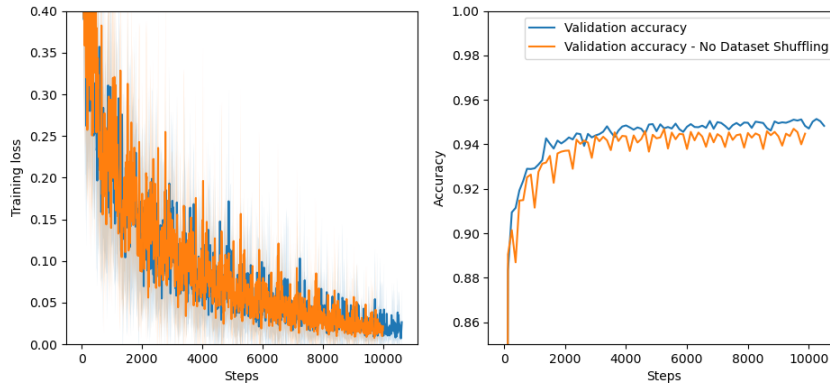

5.2 Task 4b)

To many hidden unites (128) increases the learning speed (slower iterations and convergance after 20 epochs), but the accuracy increases slightly. In general however a too complex network will tend to overfit and thus reduce the accuracy.

5.3 Task 4d)

In the network from task 3 there is 50880 parameters.
In the new network there is 51300 parameters. The network has two hidden layers of 60 nodes each.
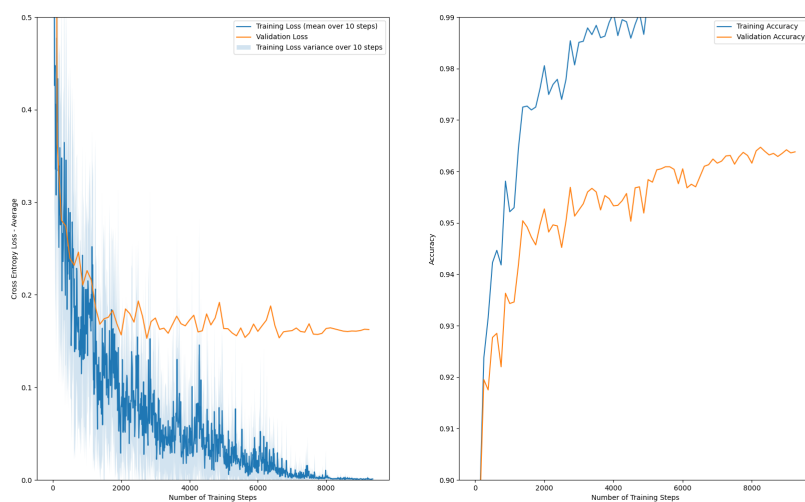
The new model gives about the same accuracy, but it converges alot faster. We can also see that it's overfitting, giving 100% accuracy on the training data.

5.4 Task 4e)