

# Assignment2\_Group15

February 23, 2024

## 1 Assignment 2 report, group 15

### 2 Task 1

#### 2.1 task 1a)

#### 2.2 task 1b)

### 3 Task 2

#### 3.1 Task 2a)

- The mean value for X\_train was: 33.55274553571429
- The Standard deviation for X\_train was: 78.87550070784701

#### 3.2 Task 2c)

#### 3.3 Task 2d)

- The number of parameters were calculated in the get\_number\_of\_params() function in task2.py
- The number of params with 784 input layer, 64 hidden layer and 10 output layer was 50890 (Have added for bias term here)

## 4 Task 3

Here is our plot with the different combinations of “technique” to improve the model. The tricks of the trade!

- Max accuracy for scenario Original 0.8945
  - Original base line model
- Max accuracy for scenario Improved Sigmoid 0.877
  - Improved sigmoid activation function, which shows in increase in learning speed, but falls off in accuracy over time
- Max accuracy for scenario Improved Weight Init 0.9523
  - We see that this are a large improvement both in learning speed and accuracy. There also seem to take a long time before the model overfits and the early stopping kicks in.
- Max accuracy for scenario With Momentum 0.9196
  - Same as for the improved weight init, but seems to be a bit slower in learning speed and overfits faster with less accuracy

- Max accuracy for scenario All Improvements 0.9607
  - This is the best model, with all the improvements combined. It has the best learning speed and accuracy, and converges fast to a good solution.

## 5 Task 4

### 5.1 Task 4a)

Here is the similar plot from task 3, but now the number of hidden units changed from 64 to 32

- Max accuracy for scenario Original 0.8895
- Max accuracy for scenario Improved Sigmoid 0.8811
- Max accuracy for scenario Improved Weight Init 0.9396
- Max accuracy for scenario With Momentum 0.9118
- Max accuracy for scenario All Improvements 0.9432

We observe that the accuracy is lower for all the scenarios. The performance is worse.

### 5.2 Task 4b)

Here is the similar plot from task 3, but now the number of hidden units changed from 64 to 128

- Max accuracy for scenario Original 0.901
- Max accuracy for scenario Improved Sigmoid 0.8748
- Max accuracy for scenario Improved Weight Init 0.9572
- Max accuracy for scenario With Momentum 0.9213
- Max accuracy for scenario All Improvements 0.9679

We observe that the accuracy is a bit higher for all the scenarios compared to the 64 hidden units. Compared to the 32 hidden units, the accuracy is a lot higher for all the scenarios.

This is not what we expected based on the task, but the difference is not too much.

### 5.3 Task 4d)

For this task, we observed by trial and fail that two hidden layers of 59 units is about the same number of parameters as a single hidden layer of 64 units.

- The number of params, as stated earlier, with a single 64 units hidden layer is 50890.
- The new number of params with two hidden layers of 59 units is 50455.

We also observed that the two hidden layers of 59 units had a higher accuracy than the single hidden layer of 64 units.

- Max accuracy for scenario with all tricks of the trade is 0.9626
  - This is slightly higher again, which means the two hidden layers are slightly better on the classification task.

### 5.4 Task 4e)

This task we take a look at the training loss of a 10 hidden layer network with 64 units in each layer. We are comparing this to the baseline model with all “tricks of the trade” applied.

- Min Loss for scenario all Improvements 2.4165162806958054
- Min Loss for scenario all Improvements 2.4034787796197197

We see that the loss of the deep model increases significantly compared to the baseline model. This is due to the vanishing gradient problem. The deep model is not able to learn as well as the baseline model.