Task 2:

2a)

mean = 33.55274553571429
standard deviation = 78.87550070784701

2b)

See code for implementation.

2c)

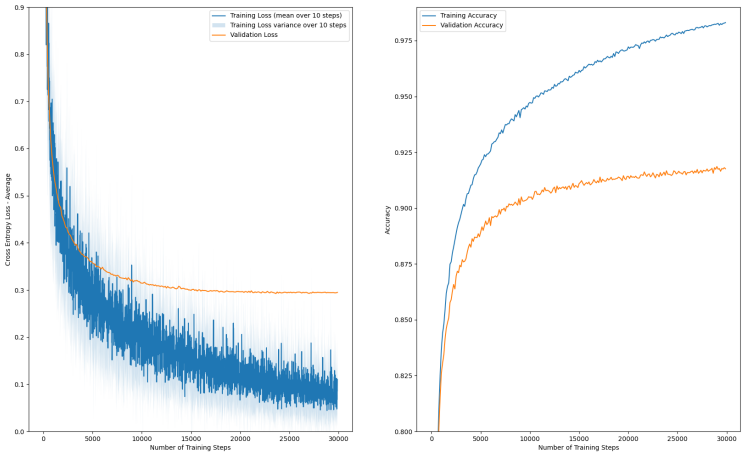

Figure 1: Plot of the training and validation loss and accuracy over training.

2d)

We have an input layer with the size of 784. In addition we have 1 bias unit. The hidden layer will get input from 785 nodes. Further, it has 64 nodes, which means there are 64 weights. We will get 785 * 64 parameters in this layer. In the next layer we have 10 nodes and it gets input from 64 nodes. This 64 *10 parameters. In total we will have :

$$Numberofparameters = 785 \cdot 64 + 64 \cdot 10 = 50880 \quad (1)$$

Task 3 :

3a)

See code for implementation.

3b)

See code for implementation.

3c)

See code for implementation.

Task 4 :

4a)

4b)

4c)

4d)

4e)

4f)