

Q1.1

Q2.1.1 Network Initialization

A network initialized to all zeros will result the nerurons to ouptput zero, with forward propagation. In addition, the back propagation will result to the same gradients for each test train. With specific layer weights and biases, the parameters will be the same throughout the training process and the optimum will not be obtained.

Q 2.1.3 Network Initialization

Random numbers initialization is desired, because it breaks the network symmetry. There is a faster chance for the network to reach global optimum without the symmetry.

Scaling the initialization depending layer size is done to make the output value (forward propagation) and gradients (back propagation) not too big or too small by making the variance of the parameter's gradients constant for every layer. This will prevent the vanishing problem or gradient explosion.

Q 3.1

Best accuracy of 75.6 % is achieved with batch size of 48 and learning size of 0.01, along with the given 40 epochs & hidden_size of 64.

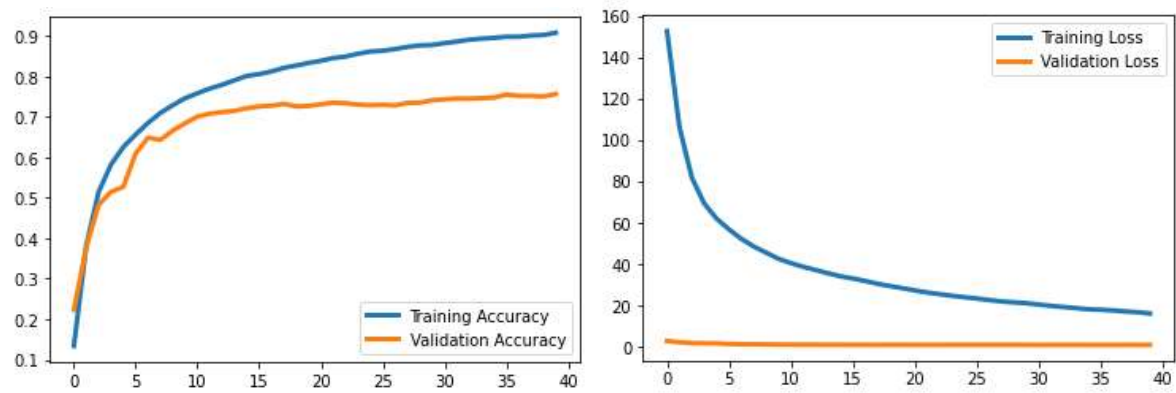


Figure 1. Best learning rate of 75.6% using 0.01 learning size.

Q 3.2

Best accuracy of 75.6 % is achieved with batch size of 48 and learning size of 0.01, along with the given 40 epochs & hidden_size of 64. The plots are shown below. Further figures show the loss and accuracy using learning size of 0.1 and 0.001 respectively. A lower learning rate of 0.001, the time needed to reach the desired accuracy becomes longer, making the final accuracy at the same epoch value to be slightly lower. With a larger learning rate, the gradient updates at a much larger rate to obtain the local optimum. In other words, the gradient difference is too big that it skips notable learning points.

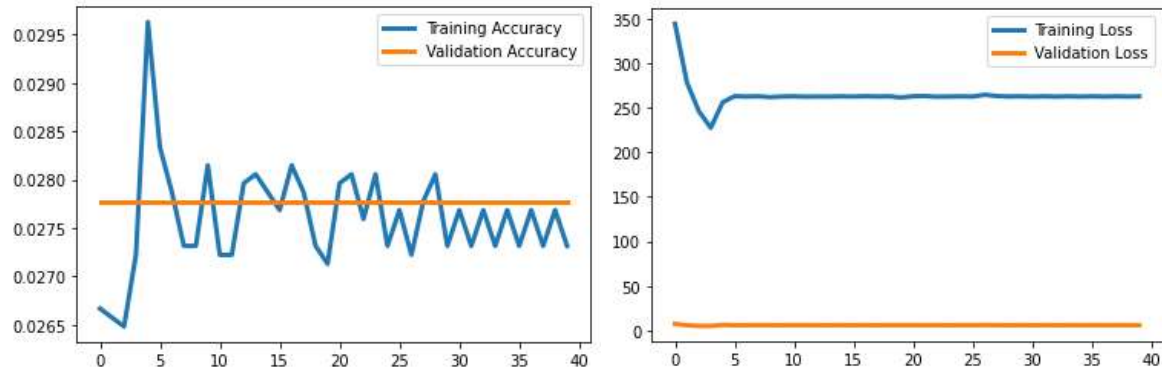


Figure 2. Learning rate of 2.78% using 0.1 learning size.

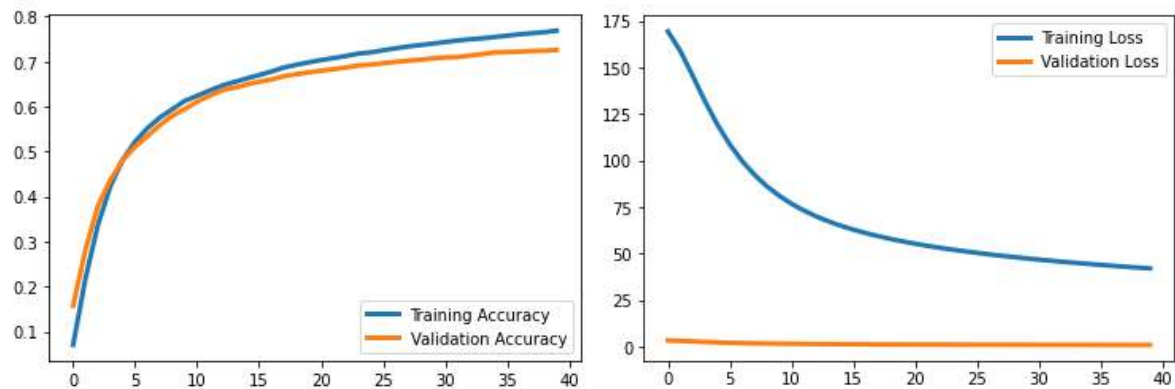


Figure 3. Learning rate of 72.58% using 0.001 learning size.

Q 3.3

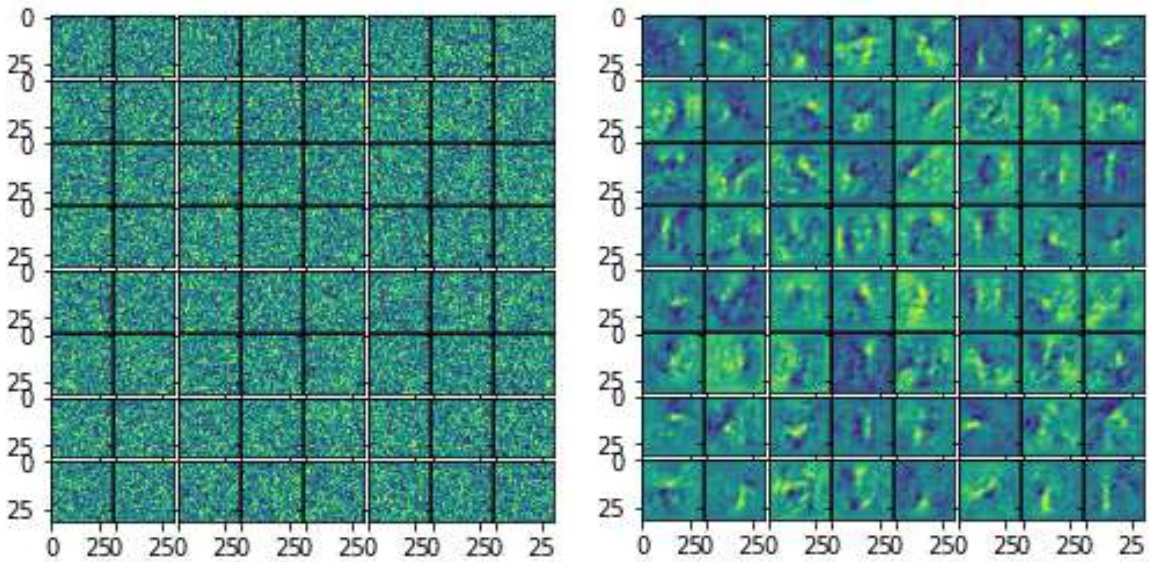


Figure 4. Learned weights for first layer and learned weights (left and rights respectively).

The left side represents the first layer immediately after initialization and shows random noisy data. However, in the right image of learned weights, there is obvious black and light spots, which describes pattern in each grid. Thus, this confirms that the network is learning through training loops.

Q 3.4

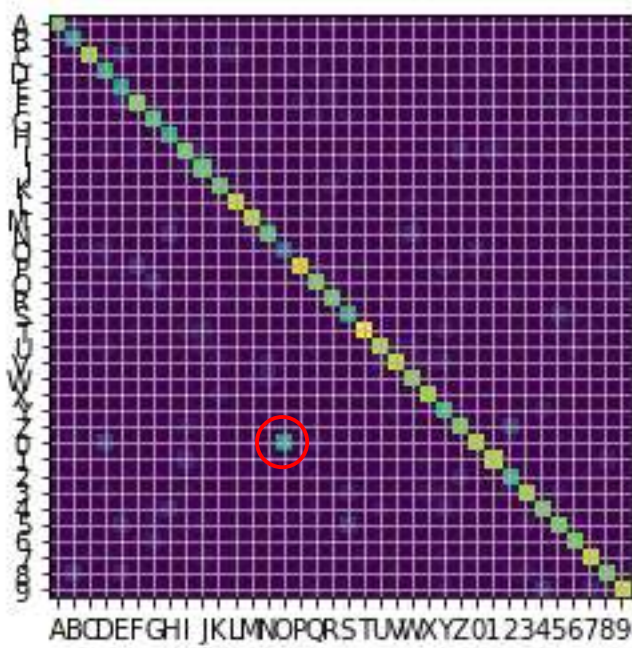
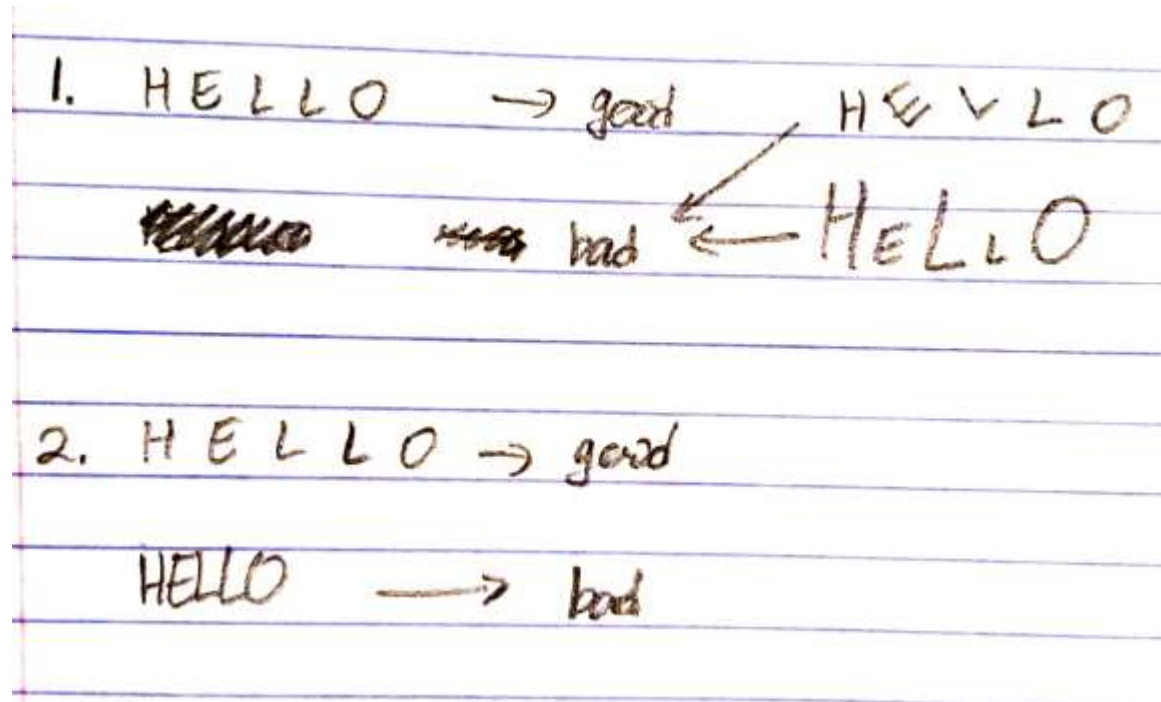


Figure 5. Confusion matrix of best model.

There is one major pair of class in the confusion matrix which is (O, 0). This is understandable as the two characters look the same with circular shape and little difference in thickness. Truthfully, some human handwriting has these two characteristics frequently mixed up.

Q 4.1

Two big assumptions needed for the sample method to work are consistent correct oriented characters and separated characters within each other.



Q 4.3 Find Letters Results

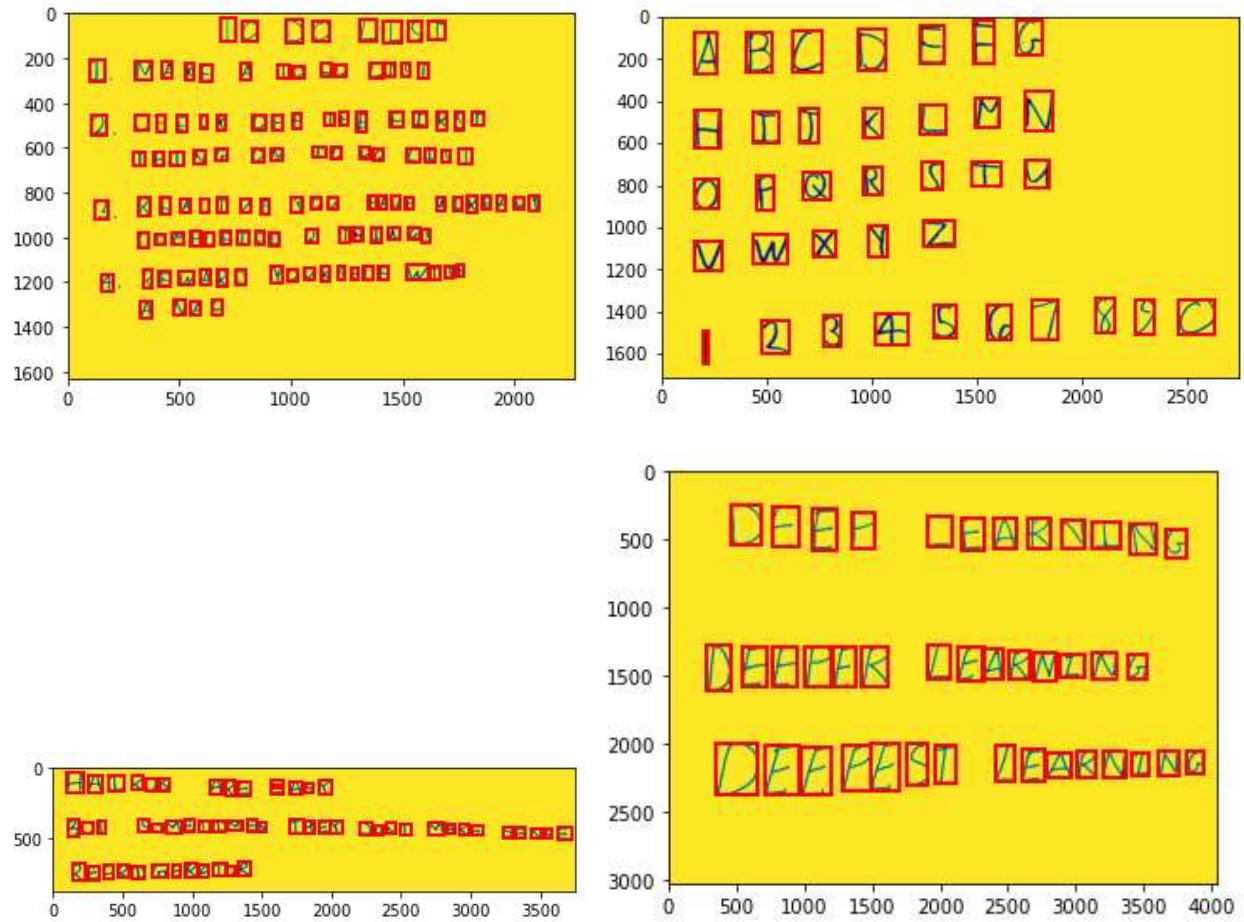


Figure 6. Result images from Find Letters function.

Q4.4

01 list

```
T7J7LC8T
IN0KLATDLDLI8T
JCHFCKBFF7H5FIR8T
THINGQNTQDQLI8T
YRIACILEYQUHAVEALRLADT
CQMPLEFTCDITHINGI
9RWARDYQWKSELFWITH
ANAP
```

02 letters

```
CYCJYFG
HITKLMN
QPQK8TU
VWXYZ
1Z3QSG7Y7J
```

03 haiku

```
HAIWUFARFEMASY
5CTSZMETIMBSTHEYDQNTMAKGSGMGE
RBFRIGERQTQR
```

04 deep

```
JYTYJY2KJIXG
YCPYF7LF0KYING
JTFPFYTCCFRNING
```