**REPORT**

a. Network Architecture

input layer – 784 nodes data image,

one hidden layer – 512 nodes

output layer – 10 nodes for each digit from 0-9

B/w input and hidden layer, **relu** function was used as the activation function

B/w hidden and output layer, **softmax** function was used as the activation function

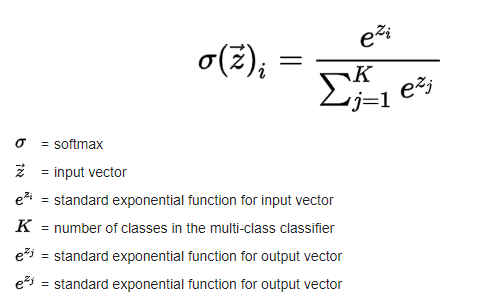
Error function used for softmax function – **cross entropy** function

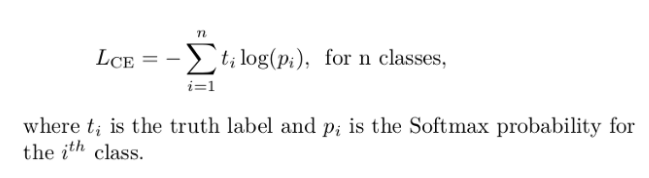
b. Activation function

**relu function**:

y(x) = max (0, x)

**softmax function**:



c. Loss function 

d. Learning Algorithm

**Mini Batch gradient descent**

4000 epochs

1 epoch → 40 batches of size 20 each.

e. Learning Rate

constant learning rate of 0.01

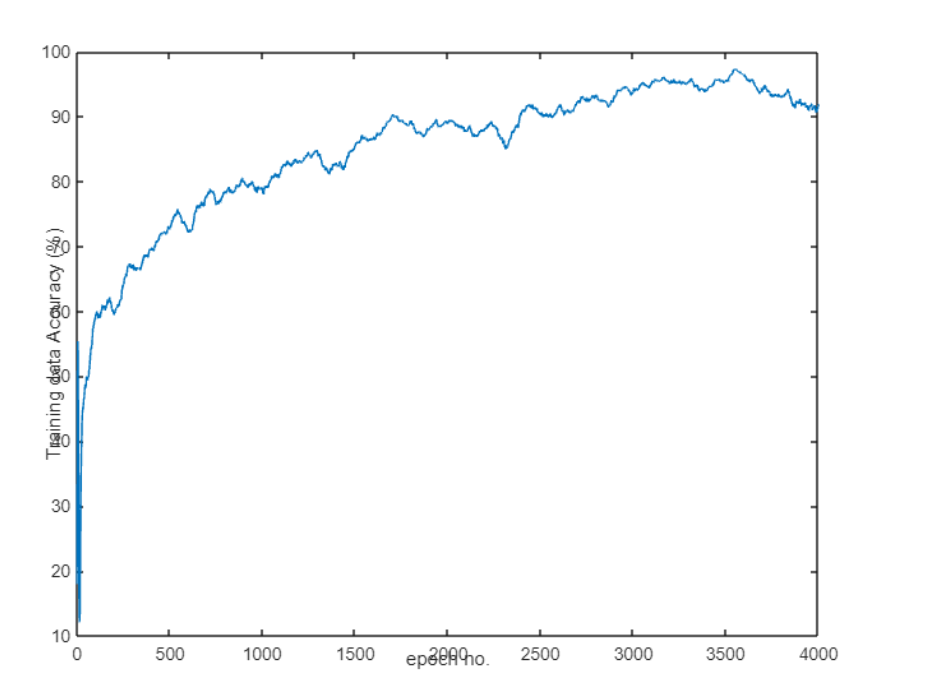
f. Loss and Accuracy of testing data set

Final Accuracy – 71.5%

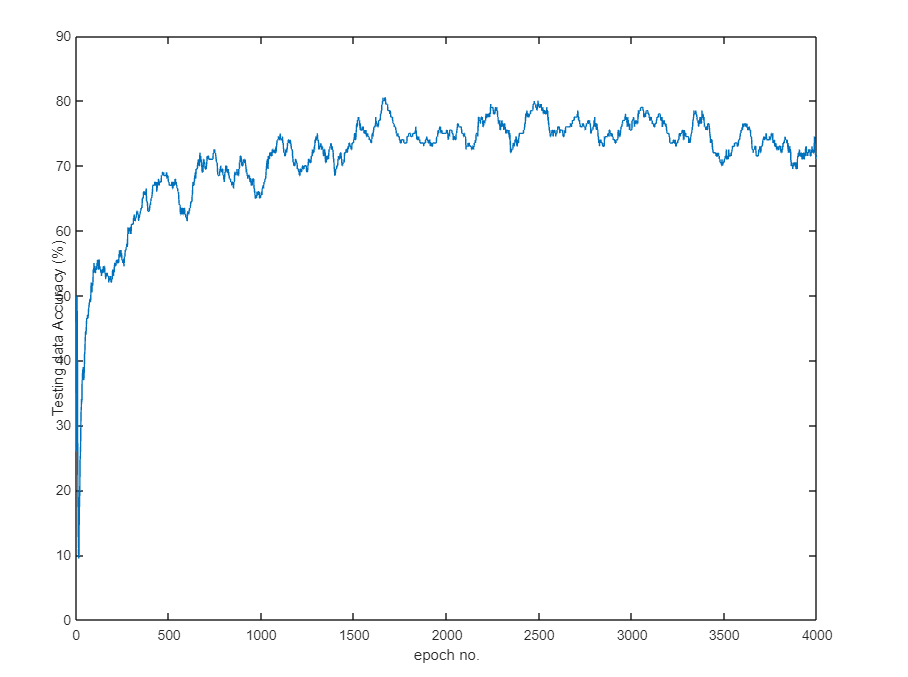
Final average Loss – 0.0018

**PLOTS**

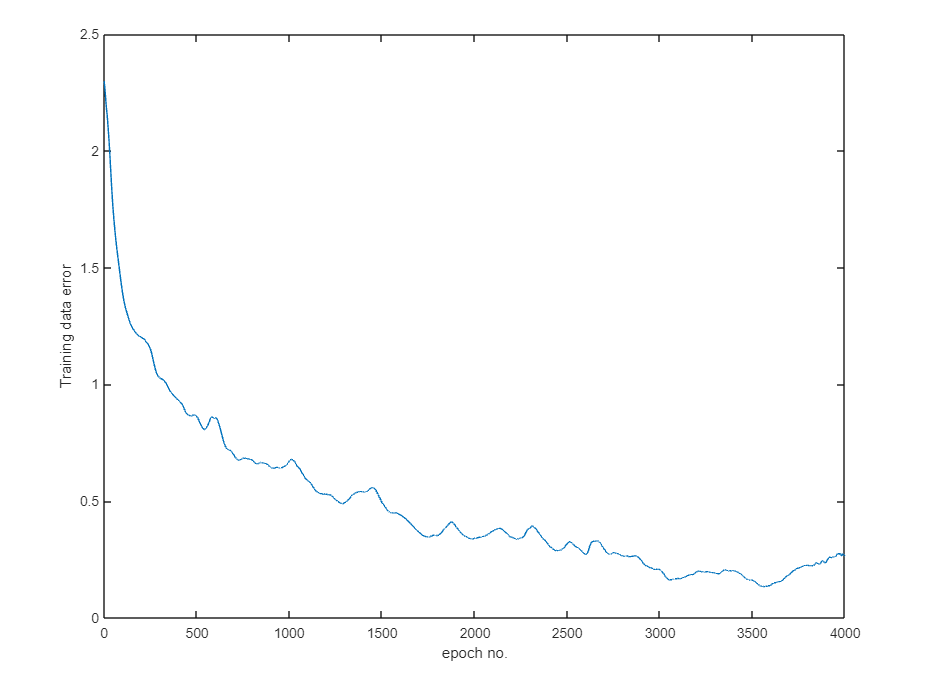
Training data accuracy

****

Testing data accuracy



Training data Loss



Testing data Loss

