Vansh Gupta  
2019EE10143  
Assignment-3

**Part 1A.**

In my implementation for NN, I have introduced a lot of flexibility, as follows:

* Number of hidden layers
* Number of units in each layer
* Activation function for each layer: ReLU, Tanh, Sigmoid
  + Last layer has softmax activation
* Learning rate
  + Initial learning rate can be specified
  + User can enable dynamic lr, which changes with epochs
* Number of epochs
* Batch size
* Mode of regularization: Early stopping
  + I skipped L2 as it had been covered in one of the assignments before. I instead focused on a new method of regularization
  + Patience and threshold can also be specified
* Cost function: MSE, SSE, Cross-Entropy
* Number of folds for using a k-folds CV
* Train-Dev-Test split
* Data normalization

This gave my implementation enough flexibility to work in any way that I want it to.  
I did not add any data normalization in current set of experiments as the data we were given was already in the range of 0 to 1.

For data splitting, I adopted the following technique:

I first took my training set and split it into 2 parts in ratio 4:1. This gave me 2400 sentences in training data and 600 in testing. Further, I adopted a 5-folds CV in my implementation to get 1920 instances in my final training set and 480 in the development/validation set. I train on my training set, select the best hyperparameters based on the accuracy on my validation set and finally, report the accuracy on my test set

I ran a search for the following hyperparameters:

Epochs = [1, 10, 20, 50]  
Learning rate = [0.001, 0.01, 0.1]  
Cost\_fn = ['mse', 'sse']  
Batch\_sizes = [1, 10, 100]

I adopted the early stopping for regularization which was true for the entire experiment.  
In the models where early stopping did take place, I rechecked by disabling this parameter, and validated that early stopping does help with the problem of overfitting

Further, I used a dynamic learning rate for the entirety of my executions as I had previously confirmed that these help in faster and better convergence as they make sure model is learnt more initially and as the epochs pass, it gets closer to the minima and should learn slowly so as not to miss the minima.

Note: I could have added more parameters, but on the CPU that I was using (Google Colab, 7~8 times faster than mine), even these were taking more than 10 hours and notebooks on Colab crash after 12 hours or a few hours of inactivity. This restricted my search. Even with these, for epochs=50 and 100, the model was taking more than 40 minutes for 1 iteration and it became impossible to work with these. So many iterations were then done with just 1 validation set

One thing I could have done (And tried as well) was to reduce the number of training instances and reported results on that. However, I noticed that there wasn’t much difference among these two and I would still not have been able to incorporate the values that I wanted to. At least, currently the values reflect answers on the entire set.

These are the results that I received (Note: The time includes both training and testing as asked in the assignment):

The first model has following layers:

Layer type: Fully Connected   
Input Dimension: 784 Output Dimension: 128  
Layer type: Activation Layer  
Activation Function: ReLU  
Layer type: Fully Connected  
Input Dimension: 128 Output Dimension: 10  
Layer type: Softmax   
Input Dimension: 10

--------------------------------------------------

Iteration Number: 1

Number of Epochs: 1

Learning Rate: 0.001

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.089184 | Final loss:0.089184

Fold Number: 2/5

Initial loss:0.092600 | Final loss:0.092600

Fold Number: 3/5

Initial loss:0.092868 | Final loss:0.092868

Fold Number: 4/5

Initial loss:0.092216 | Final loss:0.092216

Fold Number: 5/5

Initial loss:0.092947 | Final loss:0.092947

Mean train accuracy: 11.489583

Mean dev accuracy: 10.625000

Total time taken in 5-folds CV on given set of hyperparameters: 19.214358 seconds

--------------------------------------------------

Iteration Number: 2

Number of Epochs: 1

Learning Rate: 0.01

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.089157 | Final loss:0.089157

Fold Number: 2/5

Initial loss:0.092313 | Final loss:0.092313

Fold Number: 3/5

Initial loss:0.091435 | Final loss:0.091435

Fold Number: 4/5

Initial loss:0.092364 | Final loss:0.092364

Fold Number: 5/5

Initial loss:0.092222 | Final loss:0.092222

Mean train accuracy: 11.322917

Mean dev accuracy: 11.541667

Total time taken in 5-folds CV on given set of hyperparameters: 18.889737 seconds

--------------------------------------------------

Iteration Number: 3

Number of Epochs: 1

Learning Rate: 0.1

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.089630 | Final loss:0.089630

Fold Number: 2/5

Initial loss:0.091634 | Final loss:0.091634

Fold Number: 3/5

Initial loss:0.091180 | Final loss:0.091180

Fold Number: 4/5

Initial loss:0.091191 | Final loss:0.091191

Fold Number: 5/5

Initial loss:0.090950 | Final loss:0.090950

Mean train accuracy: 14.343750

Mean dev accuracy: 14.000000

Total time taken in 5-folds CV on given set of hyperparameters: 18.920217 seconds

--------------------------------------------------

Iteration Number: 4

Number of Epochs: 1

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.443631 | Final loss:0.443631

Fold Number: 2/5

Initial loss:0.460160 | Final loss:0.460160

Fold Number: 3/5

Initial loss:0.442526 | Final loss:0.442526

Fold Number: 4/5

Initial loss:0.438810 | Final loss:0.438810

Fold Number: 5/5

Initial loss:0.450914 | Final loss:0.450914

Mean train accuracy: 24.864583

Mean dev accuracy: 24.625000

Total time taken in 5-folds CV on given set of hyperparameters: 18.785089 seconds

--------------------------------------------------

Iteration Number: 5

Number of Epochs: 1

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.448467 | Final loss:0.448467

Fold Number: 2/5

Initial loss:0.443328 | Final loss:0.443328

Fold Number: 3/5

Initial loss:0.451519 | Final loss:0.451519

Fold Number: 4/5

Initial loss:0.448825 | Final loss:0.448825

Fold Number: 5/5

Initial loss:0.449321 | Final loss:0.449321

Mean train accuracy: 23.177083

Mean dev accuracy: 23.500000

Total time taken in 5-folds CV on given set of hyperparameters: 18.534141 seconds

--------------------------------------------------

Iteration Number: 6

Number of Epochs: 1

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.454173 | Final loss:0.454173

Fold Number: 2/5

Initial loss:0.459959 | Final loss:0.459959

Fold Number: 3/5

Initial loss:0.444918 | Final loss:0.444918

Fold Number: 4/5

Initial loss:0.441046 | Final loss:0.441046

Fold Number: 5/5

Initial loss:0.453390 | Final loss:0.453390

Mean train accuracy: 23.000000

Mean dev accuracy: 20.791667

Total time taken in 5-folds CV on given set of hyperparameters: 18.593548 seconds

--------------------------------------------------

Iteration Number: 7

Number of Epochs: 1

Learning Rate: 0.001

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.086340 | Final loss:0.086340

Fold Number: 2/5

Initial loss:0.088661 | Final loss:0.088661

Fold Number: 3/5

Initial loss:0.087401 | Final loss:0.087401

Fold Number: 4/5

Initial loss:0.086940 | Final loss:0.086940

Fold Number: 5/5

Initial loss:0.085239 | Final loss:0.085239

Mean train accuracy: 41.927083

Mean dev accuracy: 40.833333

Total time taken in 5-folds CV on given set of hyperparameters: 19.192737 seconds

--------------------------------------------------

Iteration Number: 8

Number of Epochs: 1

Learning Rate: 0.01

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.086870 | Final loss:0.086870

Fold Number: 2/5

Initial loss:0.086973 | Final loss:0.086973

Fold Number: 3/5

Initial loss:0.087076 | Final loss:0.087076

Fold Number: 4/5

Initial loss:0.088169 | Final loss:0.088169

Fold Number: 5/5

Initial loss:0.087536 | Final loss:0.087536

Mean train accuracy: 39.760417

Mean dev accuracy: 37.583333

Total time taken in 5-folds CV on given set of hyperparameters: 18.877996 seconds

--------------------------------------------------

Iteration Number: 9

Number of Epochs: 1

Learning Rate: 0.1

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.087619 | Final loss:0.087619

Fold Number: 2/5

Initial loss:0.087288 | Final loss:0.087288

Fold Number: 3/5

Initial loss:0.085798 | Final loss:0.085798

Fold Number: 4/5

Initial loss:0.088445 | Final loss:0.088445

Fold Number: 5/5

Initial loss:0.085861 | Final loss:0.085861

Mean train accuracy: 39.697917

Mean dev accuracy: 39.541667

Total time taken in 5-folds CV on given set of hyperparameters: 18.821635 seconds

--------------------------------------------------

Iteration Number: 10

Number of Epochs: 1

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.293416 | Final loss:0.293416

Fold Number: 2/5

Initial loss:0.297315 | Final loss:0.297315

Fold Number: 3/5

Initial loss:0.311606 | Final loss:0.311606

Fold Number: 4/5

Initial loss:0.320506 | Final loss:0.320506

Fold Number: 5/5

Initial loss:0.313933 | Final loss:0.313933

Mean train accuracy: 80.395833

Mean dev accuracy: 78.416667

Total time taken in 5-folds CV on given set of hyperparameters: 18.715522 seconds

--------------------------------------------------

Iteration Number: 11

Number of Epochs: 1

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.309683 | Final loss:0.309683

Fold Number: 2/5

Initial loss:0.301022 | Final loss:0.301022

Fold Number: 3/5

Initial loss:0.318283 | Final loss:0.318283

Fold Number: 4/5

Initial loss:0.328871 | Final loss:0.328871

Fold Number: 5/5

Initial loss:0.297979 | Final loss:0.297979

Mean train accuracy: 78.895833

Mean dev accuracy: 76.625000

Total time taken in 5-folds CV on given set of hyperparameters: 18.420845 seconds

--------------------------------------------------

Iteration Number: 12

Number of Epochs: 1

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.309502 | Final loss:0.309502

Fold Number: 2/5

Initial loss:0.317175 | Final loss:0.317175

Fold Number: 3/5

Initial loss:0.305725 | Final loss:0.305725

Fold Number: 4/5

Initial loss:0.318600 | Final loss:0.318600

Fold Number: 5/5

Initial loss:0.320567 | Final loss:0.320567

Mean train accuracy: 78.927083

Mean dev accuracy: 75.833333

Total time taken in 5-folds CV on given set of hyperparameters: 18.630308 seconds

--------------------------------------------------

Iteration Number: 13

Number of Epochs: 1

Learning Rate: 0.001

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.048697 | Final loss:0.048697

Fold Number: 2/5

Initial loss:0.048408 | Final loss:0.048408

Fold Number: 3/5

Initial loss:0.048678 | Final loss:0.048678

Fold Number: 4/5

Initial loss:0.047769 | Final loss:0.047769

Fold Number: 5/5

Initial loss:0.050086 | Final loss:0.050086

Mean train accuracy: 84.479167

Mean dev accuracy: 81.708333

Total time taken in 5-folds CV on given set of hyperparameters: 19.131150 seconds

--------------------------------------------------

Iteration Number: 14

Number of Epochs: 1

Learning Rate: 0.01

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.049151 | Final loss:0.049151

Fold Number: 2/5

Initial loss:0.049256 | Final loss:0.049256

Fold Number: 3/5

Initial loss:0.049132 | Final loss:0.049132

Fold Number: 4/5

Initial loss:0.047404 | Final loss:0.047404

Fold Number: 5/5

Initial loss:0.048547 | Final loss:0.048547

Mean train accuracy: 84.635417

Mean dev accuracy: 81.625000

Total time taken in 5-folds CV on given set of hyperparameters: 18.947776 seconds

--------------------------------------------------

Iteration Number: 15

Number of Epochs: 1

Learning Rate: 0.1

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.047967 | Final loss:0.047967

Fold Number: 2/5

Initial loss:0.045006 | Final loss:0.045006

Fold Number: 3/5

Initial loss:0.050267 | Final loss:0.050267

Fold Number: 4/5

Initial loss:0.048341 | Final loss:0.048341

Fold Number: 5/5

Initial loss:0.048733 | Final loss:0.048733

Mean train accuracy: 84.729167

Mean dev accuracy: 81.375000

Total time taken in 5-folds CV on given set of hyperparameters: 18.871506 seconds

--------------------------------------------------

Iteration Number: 16

Number of Epochs: 1

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.193662 | Final loss:0.193662

Fold Number: 2/5

Initial loss:0.194130 | Final loss:0.194130

Fold Number: 3/5

Initial loss:0.188523 | Final loss:0.188523

Fold Number: 4/5

Initial loss:0.195295 | Final loss:0.195295

Fold Number: 5/5

Initial loss:0.192649 | Final loss:0.192649

Mean train accuracy: 86.427083

Mean dev accuracy: 83.875000

Total time taken in 5-folds CV on given set of hyperparameters: 18.681968 seconds

--------------------------------------------------

Iteration Number: 17

Number of Epochs: 1

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.195767 | Final loss:0.195767

Fold Number: 2/5

Initial loss:0.191477 | Final loss:0.191477

Fold Number: 3/5

Initial loss:0.192583 | Final loss:0.192583

Fold Number: 4/5

Initial loss:0.187593 | Final loss:0.187593

Fold Number: 5/5

Initial loss:0.190711 | Final loss:0.190711

Mean train accuracy: 87.041667

Mean dev accuracy: 82.541667

Total time taken in 5-folds CV on given set of hyperparameters: 18.504767 seconds

--------------------------------------------------

Iteration Number: 18

Number of Epochs: 1

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.198288 | Final loss:0.198288

Fold Number: 2/5

Initial loss:0.188608 | Final loss:0.188608

Fold Number: 3/5

Initial loss:0.191898 | Final loss:0.191898

Fold Number: 4/5

Initial loss:0.194388 | Final loss:0.194388

Fold Number: 5/5

Initial loss:0.190813 | Final loss:0.190813

Mean train accuracy: 84.968750

Mean dev accuracy: 80.833333

Total time taken in 5-folds CV on given set of hyperparameters: 18.472997 seconds

--------------------------------------------------

Iteration Number: 19

Number of Epochs: 10

Learning Rate: 0.001

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.090711 | Final loss:0.083771

Fold Number: 2/5

Initial loss:0.092388 | Final loss:0.088256

Fold Number: 3/5

Initial loss:0.090054 | Final loss:0.085459

Fold Number: 4/5

Initial loss:0.092965 | Final loss:0.087129

Fold Number: 5/5

Initial loss:0.089747 | Final loss:0.082729

Mean train accuracy: 28.031250

Mean dev accuracy: 27.791667

Total time taken in 5-folds CV on given set of hyperparameters: 168.541337 seconds

--------------------------------------------------

Iteration Number: 20

Number of Epochs: 10

Learning Rate: 0.01

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.090713 | Final loss:0.085337

Fold Number: 2/5

Initial loss:0.091367 | Final loss:0.086650

Fold Number: 3/5

Initial loss:0.090200 | Final loss:0.085164

Fold Number: 4/5

Initial loss:0.092579 | Final loss:0.085216

Fold Number: 5/5

Initial loss:0.090434 | Final loss:0.085096

Mean train accuracy: 29.979167

Mean dev accuracy: 29.083333

Total time taken in 5-folds CV on given set of hyperparameters: 166.744933 seconds

--------------------------------------------------

Iteration Number: 21

Number of Epochs: 10

Learning Rate: 0.1

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.091859 | Final loss:0.086099

Fold Number: 2/5

Initial loss:0.091926 | Final loss:0.086962

Fold Number: 3/5

Initial loss:0.091690 | Final loss:0.087408

Fold Number: 4/5

Initial loss:0.092610 | Final loss:0.087948

Fold Number: 5/5

Initial loss:0.091495 | Final loss:0.086993

Mean train accuracy: 24.208333

Mean dev accuracy: 24.250000

Total time taken in 5-folds CV on given set of hyperparameters: 166.406268 seconds

--------------------------------------------------

Iteration Number: 22

Number of Epochs: 10

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.447325 | Final loss:0.289692

Fold Number: 2/5

Initial loss:0.450968 | Final loss:0.285285

Fold Number: 3/5

Initial loss:0.449127 | Final loss:0.289756

Fold Number: 4/5

Initial loss:0.437152 | Final loss:0.293050

Fold Number: 5/5

Initial loss:0.448444 | Final loss:0.287238

Mean train accuracy: 67.822917

Mean dev accuracy: 66.666667

Total time taken in 5-folds CV on given set of hyperparameters: 163.995844 seconds

--------------------------------------------------

Iteration Number: 23

Number of Epochs: 10

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.449905 | Final loss:0.275672

Fold Number: 2/5

Initial loss:0.461430 | Final loss:0.317015

Fold Number: 3/5

Initial loss:0.454640 | Final loss:0.326751

Fold Number: 4/5

Initial loss:0.442375 | Final loss:0.276670

Fold Number: 5/5

Initial loss:0.446320 | Final loss:0.292806

Mean train accuracy: 66.250000

Mean dev accuracy: 65.458333

Total time taken in 5-folds CV on given set of hyperparameters: 161.258500 seconds

--------------------------------------------------

Iteration Number: 24

Number of Epochs: 10

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.445948 | Final loss:0.307689

Fold Number: 2/5

Initial loss:0.442789 | Final loss:0.312842

Fold Number: 3/5

Initial loss:0.442895 | Final loss:0.294331

Fold Number: 4/5

Initial loss:0.433873 | Final loss:0.260888

Fold Number: 5/5

Initial loss:0.455069 | Final loss:0.307590

Mean train accuracy: 66.437500

Mean dev accuracy: 66.125000

Total time taken in 5-folds CV on given set of hyperparameters: 160.470956 seconds

--------------------------------------------------

Iteration Number: 25

Number of Epochs: 10

Learning Rate: 0.001

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.085488 | Final loss:0.034306

Fold Number: 2/5

Initial loss:0.086529 | Final loss:0.036729

Fold Number: 3/5

Initial loss:0.085166 | Final loss:0.034574

Fold Number: 4/5

Initial loss:0.086059 | Final loss:0.033066

Fold Number: 5/5

Initial loss:0.087036 | Final loss:0.034895

Mean train accuracy: 82.906250

Mean dev accuracy: 80.583333

Total time taken in 5-folds CV on given set of hyperparameters: 167.488438 seconds

--------------------------------------------------

Iteration Number: 26

Number of Epochs: 10

Learning Rate: 0.01

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.086225 | Final loss:0.037161

Fold Number: 2/5

Initial loss:0.085321 | Final loss:0.034208

Fold Number: 3/5

Initial loss:0.087323 | Final loss:0.043747

Fold Number: 4/5

Initial loss:0.087391 | Final loss:0.033151

Fold Number: 5/5

Initial loss:0.087301 | Final loss:0.038045

Mean train accuracy: 80.583333

Mean dev accuracy: 77.583333

Total time taken in 5-folds CV on given set of hyperparameters: 165.857730 seconds

--------------------------------------------------

Iteration Number: 27

Number of Epochs: 10

Learning Rate: 0.1

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.083800 | Final loss:0.035450

Fold Number: 2/5

Initial loss:0.088626 | Final loss:0.037390

Fold Number: 3/5

Initial loss:0.085564 | Final loss:0.033351

Fold Number: 4/5

Initial loss:0.088148 | Final loss:0.035110

Fold Number: 5/5

Initial loss:0.087609 | Final loss:0.037893

Mean train accuracy: 82.218750

Mean dev accuracy: 80.291667

Total time taken in 5-folds CV on given set of hyperparameters: 165.353906 seconds

--------------------------------------------------

Iteration Number: 28

Number of Epochs: 10

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.309262 | Final loss:0.063550

Fold Number: 2/5

Initial loss:0.325084 | Final loss:0.062349

Fold Number: 3/5

Initial loss:0.291381 | Final loss:0.061152

Fold Number: 4/5

Initial loss:0.318046 | Final loss:0.063188

Fold Number: 5/5

Initial loss:0.298222 | Final loss:0.062895

Mean train accuracy: 93.927083

Mean dev accuracy: 88.666667

Total time taken in 5-folds CV on given set of hyperparameters: 163.332933 seconds

--------------------------------------------------

Iteration Number: 29

Number of Epochs: 10

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.310195 | Final loss:0.060118

Fold Number: 2/5

Initial loss:0.314623 | Final loss:0.061337

Fold Number: 3/5

Initial loss:0.309724 | Final loss:0.063225

Fold Number: 4/5

Initial loss:0.298862 | Final loss:0.061782

Fold Number: 5/5

Initial loss:0.296141 | Final loss:0.063179

Mean train accuracy: 94.104167

Mean dev accuracy: 88.583333

Total time taken in 5-folds CV on given set of hyperparameters: 162.035336 seconds

--------------------------------------------------

Iteration Number: 30

Number of Epochs: 10

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.323202 | Final loss:0.062377

Fold Number: 2/5

Initial loss:0.322857 | Final loss:0.062232

Fold Number: 3/5

Initial loss:0.301145 | Final loss:0.063302

Fold Number: 4/5

Initial loss:0.322247 | Final loss:0.063759

Fold Number: 5/5

Initial loss:0.298618 | Final loss:0.062563

Mean train accuracy: 93.729167

Mean dev accuracy: 88.583333

Total time taken in 5-folds CV on given set of hyperparameters: 161.641979 seconds

--------------------------------------------------

Iteration Number: 31

Number of Epochs: 10

Learning Rate: 0.001

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.048716 | Final loss:0.008250

Fold Number: 2/5

Initial loss:0.048349 | Final loss:0.008475

Fold Number: 3/5

Initial loss:0.049554 | Final loss:0.008093

Fold Number: 4/5

Initial loss:0.050945 | Final loss:0.008084

Fold Number: 5/5

Initial loss:0.049928 | Final loss:0.008425

Mean train accuracy: 96.291667

Mean dev accuracy: 89.541667

Total time taken in 5-folds CV on given set of hyperparameters: 168.239070 seconds

--------------------------------------------------

Iteration Number: 32

Number of Epochs: 10

Learning Rate: 0.01

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.047079 | Final loss:0.007771

Fold Number: 2/5

Initial loss:0.047099 | Final loss:0.008233

Fold Number: 3/5

Initial loss:0.048829 | Final loss:0.008520

Fold Number: 4/5

Initial loss:0.052008 | Final loss:0.008691

Fold Number: 5/5

Initial loss:0.046940 | Final loss:0.008638

Mean train accuracy: 96.125000

Mean dev accuracy: 89.500000

Total time taken in 5-folds CV on given set of hyperparameters: 167.005935 seconds

--------------------------------------------------

Iteration Number: 33

Number of Epochs: 10

Learning Rate: 0.1

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.044595 | Final loss:0.008000

Fold Number: 2/5

Initial loss:0.051683 | Final loss:0.008199

Fold Number: 3/5

Initial loss:0.050208 | Final loss:0.008270

Fold Number: 4/5

Initial loss:0.048391 | Final loss:0.008408

Fold Number: 5/5

Initial loss:0.051770 | Final loss:0.008499

Mean train accuracy: 96.239583

Mean dev accuracy: 89.500000

Total time taken in 5-folds CV on given set of hyperparameters: 166.049908 seconds

--------------------------------------------------

Iteration Number: 34

Number of Epochs: 10

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.192039 | Final loss:0.010876

Fold Number: 2/5

Initial loss:0.192555 | Final loss:0.011648

Fold Number: 3/5

Initial loss:0.197304 | Final loss:0.011339

Fold Number: 4/5

Initial loss:0.192065 | Final loss:0.010159

Fold Number: 5/5

Initial loss:0.188796 | Final loss:0.012147

Mean train accuracy: 98.750000

Mean dev accuracy: 91.791667

Total time taken in 5-folds CV on given set of hyperparameters: 164.246516 seconds

--------------------------------------------------

Iteration Number: 35

Number of Epochs: 10

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.185720 | Final loss:0.010880

Fold Number: 2/5

Initial loss:0.191188 | Final loss:0.013088

Fold Number: 3/5

Initial loss:0.193368 | Final loss:0.011042

Fold Number: 4/5

Initial loss:0.192199 | Final loss:0.011842

Fold Number: 5/5

Initial loss:0.191095 | Final loss:0.011176

Mean train accuracy: 98.739583

Mean dev accuracy: 91.500000

Total time taken in 5-folds CV on given set of hyperparameters: 162.919835 seconds

--------------------------------------------------

Iteration Number: 36

Number of Epochs: 10

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.187583 | Final loss:0.009650

Fold Number: 2/5

Initial loss:0.191432 | Final loss:0.010623

Fold Number: 3/5

Initial loss:0.193160 | Final loss:0.011209

Fold Number: 4/5

Initial loss:0.196299 | Final loss:0.010835

Fold Number: 5/5

Initial loss:0.190077 | Final loss:0.012586

Mean train accuracy: 98.875000

Mean dev accuracy: 92.000000

Total time taken in 5-folds CV on given set of hyperparameters: 162.088841 seconds

--------------------------------------------------

Iteration Number: 37

Number of Epochs: 20

Learning Rate: 0.001

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.092636 | Final loss:0.084300

Fold Number: 2/5

Initial loss:0.094182 | Final loss:0.085130

Fold Number: 3/5

Initial loss:0.091956 | Final loss:0.085217

Fold Number: 4/5

Initial loss:0.092082 | Final loss:0.084438

Fold Number: 5/5

Initial loss:0.091424 | Final loss:0.084035

Mean train accuracy: 30.656250

Mean dev accuracy: 31.041667

Total time taken in 5-folds CV on given set of hyperparameters: 335.102888 seconds

--------------------------------------------------

Iteration Number: 38

Number of Epochs: 20

Learning Rate: 0.01

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.090344 | Final loss:0.082353

Fold Number: 2/5

Initial loss:0.092931 | Final loss:0.085410

Fold Number: 3/5

Initial loss:0.090679 | Final loss:0.083226

Fold Number: 4/5

Initial loss:0.091556 | Final loss:0.083163

Fold Number: 5/5

Initial loss:0.090784 | Final loss:0.083164

Mean train accuracy: 31.489583

Mean dev accuracy: 31.666667

Total time taken in 5-folds CV on given set of hyperparameters: 331.299529 seconds

--------------------------------------------------

Iteration Number: 39

Number of Epochs: 20

Learning Rate: 0.1

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.090357 | Final loss:0.080398

Fold Number: 2/5

Initial loss:0.091560 | Final loss:0.084536

Fold Number: 3/5

Initial loss:0.093416 | Final loss:0.085669

Fold Number: 4/5

Initial loss:0.090522 | Final loss:0.081221

Fold Number: 5/5

Initial loss:0.091122 | Final loss:0.082416

Mean train accuracy: 33.885417

Mean dev accuracy: 32.333333

Total time taken in 5-folds CV on given set of hyperparameters: 330.801880 seconds

--------------------------------------------------

Iteration Number: 40

Number of Epochs: 20

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.445893 | Final loss:0.213558

Fold Number: 2/5

Initial loss:0.442643 | Final loss:0.220463

Fold Number: 3/5

Initial loss:0.453265 | Final loss:0.227279

Fold Number: 4/5

Initial loss:0.449725 | Final loss:0.222526

Fold Number: 5/5

Initial loss:0.455094 | Final loss:0.233298

Mean train accuracy: 76.666667

Mean dev accuracy: 74.791667

Total time taken in 5-folds CV on given set of hyperparameters: 329.882140 seconds

--------------------------------------------------

Iteration Number: 41

Number of Epochs: 20

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.443897 | Final loss:0.222255

Fold Number: 2/5

Initial loss:0.456244 | Final loss:0.242997

Fold Number: 3/5

Initial loss:0.443657 | Final loss:0.211828

Fold Number: 4/5

Initial loss:0.447324 | Final loss:0.224877

Fold Number: 5/5

Initial loss:0.453284 | Final loss:0.221132

Mean train accuracy: 75.666667

Mean dev accuracy: 74.375000

Total time taken in 5-folds CV on given set of hyperparameters: 326.383376 seconds

--------------------------------------------------

Iteration Number: 42

Number of Epochs: 20

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.444833 | Final loss:0.220060

Fold Number: 2/5

Initial loss:0.441068 | Final loss:0.205007

Fold Number: 3/5

Initial loss:0.444044 | Final loss:0.219536

Fold Number: 4/5

Initial loss:0.441849 | Final loss:0.219550

Fold Number: 5/5

Initial loss:0.441032 | Final loss:0.211579

Mean train accuracy: 77.479167

Mean dev accuracy: 75.708333

Total time taken in 5-folds CV on given set of hyperparameters: 324.575013 seconds

--------------------------------------------------

Iteration Number: 43

Number of Epochs: 20

Learning Rate: 0.001

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.085505 | Final loss:0.025290

Fold Number: 2/5

Initial loss:0.086075 | Final loss:0.026456

Fold Number: 3/5

Initial loss:0.089088 | Final loss:0.024330

Fold Number: 4/5

Initial loss:0.086594 | Final loss:0.025087

Fold Number: 5/5

Initial loss:0.084632 | Final loss:0.024878

Mean train accuracy: 87.416667

Mean dev accuracy: 85.000000

Total time taken in 5-folds CV on given set of hyperparameters: 337.820194 seconds

--------------------------------------------------

Iteration Number: 44

Number of Epochs: 20

Learning Rate: 0.01

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.087207 | Final loss:0.026119

Fold Number: 2/5

Initial loss:0.086238 | Final loss:0.024305

Fold Number: 3/5

Initial loss:0.086814 | Final loss:0.025099

Fold Number: 4/5

Initial loss:0.087614 | Final loss:0.026430

Fold Number: 5/5

Initial loss:0.083977 | Final loss:0.025186

Mean train accuracy: 87.166667

Mean dev accuracy: 84.541667

Total time taken in 5-folds CV on given set of hyperparameters: 335.111336 seconds

--------------------------------------------------

Iteration Number: 45

Number of Epochs: 20

Learning Rate: 0.1

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.084126 | Final loss:0.024882

Fold Number: 2/5

Initial loss:0.087737 | Final loss:0.026037

Fold Number: 3/5

Initial loss:0.088209 | Final loss:0.024888

Fold Number: 4/5

Initial loss:0.086780 | Final loss:0.024789

Fold Number: 5/5

Initial loss:0.084543 | Final loss:0.024712

Mean train accuracy: 87.145833

Mean dev accuracy: 84.208333

Total time taken in 5-folds CV on given set of hyperparameters: 333.458325 seconds

--------------------------------------------------

Iteration Number: 46

Number of Epochs: 20

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.318388 | Final loss:0.046586

Fold Number: 2/5

Initial loss:0.313409 | Final loss:0.045351

Fold Number: 3/5

Initial loss:0.317689 | Final loss:0.046476

Fold Number: 4/5

Initial loss:0.325837 | Final loss:0.046659

Fold Number: 5/5

Initial loss:0.315418 | Final loss:0.046770

Mean train accuracy: 95.656250

Mean dev accuracy: 89.333333

Total time taken in 5-folds CV on given set of hyperparameters: 330.168711 seconds

--------------------------------------------------

Iteration Number: 47

Number of Epochs: 20

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.331302 | Final loss:0.045272

Fold Number: 2/5

Initial loss:0.324225 | Final loss:0.045455

Fold Number: 3/5

Initial loss:0.314395 | Final loss:0.046092

Fold Number: 4/5

Initial loss:0.310854 | Final loss:0.047143

Fold Number: 5/5

Initial loss:0.323578 | Final loss:0.047692

Mean train accuracy: 95.677083

Mean dev accuracy: 89.458333

Total time taken in 5-folds CV on given set of hyperparameters: 325.048768 seconds

--------------------------------------------------

Iteration Number: 48

Number of Epochs: 20

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.298376 | Final loss:0.044632

Fold Number: 2/5

Initial loss:0.321708 | Final loss:0.045587

Fold Number: 3/5

Initial loss:0.308578 | Final loss:0.047046

Fold Number: 4/5

Initial loss:0.338359 | Final loss:0.046885

Fold Number: 5/5

Initial loss:0.312120 | Final loss:0.047204

Mean train accuracy: 95.625000

Mean dev accuracy: 89.291667

Total time taken in 5-folds CV on given set of hyperparameters: 322.535774 seconds

--------------------------------------------------

Iteration Number: 49

Number of Epochs: 20

Learning Rate: 0.001

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.049393 | Final loss:0.004816

Fold Number: 2/5

Initial loss:0.048952 | Final loss:0.005297

Fold Number: 3/5

Initial loss:0.049412 | Final loss:0.005964

Fold Number: 4/5

Initial loss:0.048550 | Final loss:0.005375

Fold Number: 5/5

Initial loss:0.050147 | Final loss:0.005691

Mean train accuracy: 97.437500

Mean dev accuracy: 89.833333

Total time taken in 5-folds CV on given set of hyperparameters: 335.084182 seconds

--------------------------------------------------

Iteration Number: 50

Number of Epochs: 20

Learning Rate: 0.01

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.049598 | Final loss:0.005572

Fold Number: 2/5

Initial loss:0.047875 | Final loss:0.005337

Fold Number: 3/5

Initial loss:0.047878 | Final loss:0.005275

Fold Number: 4/5

Initial loss:0.049862 | Final loss:0.005381

Fold Number: 5/5

Initial loss:0.047753 | Final loss:0.005622

Mean train accuracy: 97.531250

Mean dev accuracy: 89.750000

Total time taken in 5-folds CV on given set of hyperparameters: 329.027363 seconds

--------------------------------------------------

Iteration Number: 51

Number of Epochs: 20

Learning Rate: 0.1

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.047914 | Final loss:0.004976

Fold Number: 2/5

Initial loss:0.048056 | Final loss:0.005270

Fold Number: 3/5

Initial loss:0.051879 | Final loss:0.005345

Fold Number: 4/5

Initial loss:0.048067 | Final loss:0.005171

Fold Number: 5/5

Initial loss:0.047205 | Final loss:0.005623

Mean train accuracy: 97.572917

Mean dev accuracy: 89.291667

Total time taken in 5-folds CV on given set of hyperparameters: 329.554170 seconds

--------------------------------------------------

Iteration Number: 52

Number of Epochs: 20

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.196137 | Final loss:0.007752

Fold Number: 2/5

Initial loss:0.188990 | Final loss:0.005046

Fold Number: 3/5

Initial loss:0.192969 | Final loss:0.007627

Fold Number: 4/5

Initial loss:0.190407 | Final loss:0.006497

Fold Number: 5/5

Initial loss:0.188007 | Final loss:0.007877

Mean train accuracy: 99.145833

Mean dev accuracy: 92.125000

Total time taken in 5-folds CV on given set of hyperparameters: 323.303170 seconds

--------------------------------------------------

Iteration Number: 53

Number of Epochs: 20

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.184722 | Final loss:0.006518

Fold Number: 2/5

Initial loss:0.189425 | Final loss:0.005627

Fold Number: 3/5

Initial loss:0.192760 | Final loss:0.006450

Fold Number: 4/5

Initial loss:0.196816 | Final loss:0.006210

Fold Number: 5/5

Initial loss:0.193146 | Final loss:0.004685

Mean train accuracy: 99.333333

Mean dev accuracy: 91.500000

Total time taken in 5-folds CV on given set of hyperparameters: 319.223392 seconds

--------------------------------------------------

Iteration Number: 54

Number of Epochs: 20

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.193597 | Final loss:0.006660

Fold Number: 2/5

Initial loss:0.196157 | Final loss:0.006855

Fold Number: 3/5

Initial loss:0.197749 | Final loss:0.006054

Fold Number: 4/5

Initial loss:0.187437 | Final loss:0.006769

Fold Number: 5/5

Initial loss:0.184562 | Final loss:0.007343

Mean train accuracy: 99.218750

Mean dev accuracy: 92.083333

Total time taken in 5-folds CV on given set of hyperparameters: 317.869963 seconds

--------------------------------------------------

Iteration Number: 55

Number of Epochs: 50

Learning Rate: 0.001

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.091058 | Final loss:0.077547

Fold Number: 2/5

Initial loss:0.092375 | Final loss:0.081274

Fold Number: 3/5

Initial loss:0.090794 | Final loss:0.076293

Fold Number: 4/5

Initial loss:0.090123 | Final loss:0.074309

Fold Number: 5/5

Initial loss:0.092429 | Final loss:0.079531

Mean train accuracy: 46.354167

Mean dev accuracy: 45.541667

Total time taken in 5-folds CV on given set of hyperparameters: 829.361011 seconds

--------------------------------------------------

Iteration Number: 56

Number of Epochs: 50

Learning Rate: 0.01

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.092762 | Final loss:0.080062

Fold Number: 2/5

Initial loss:0.092255 | Final loss:0.075880

Fold Number: 3/5

Initial loss:0.092387 | Final loss:0.073557

Fold Number: 4/5

Initial loss:0.091500 | Final loss:0.076067

Fold Number: 5/5

Initial loss:0.090507 | Final loss:0.075827

Mean train accuracy: 46.822917

Mean dev accuracy: 44.833333

Total time taken in 5-folds CV on given set of hyperparameters: 819.366637 seconds

--------------------------------------------------

Iteration Number: 57

Number of Epochs: 50

Learning Rate: 0.1

Cost Function used: Mse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.090643 | Final loss:0.076096

Fold Number: 2/5

Initial loss:0.090497 | Final loss:0.076849

Fold Number: 3/5

Initial loss:0.091551 | Final loss:0.074132

Fold Number: 4/5

Initial loss:0.090965 | Final loss:0.074543

Fold Number: 5/5

Initial loss:0.093309 | Final loss:0.080013

Mean train accuracy: 44.260417

Mean dev accuracy: 42.833333

Total time taken in 5-folds CV on given set of hyperparameters: 822.360053 seconds

--------------------------------------------------

Iteration Number: 58

Number of Epochs: 50

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.449144 | Final loss:0.139634

Fold Number: 2/5

Initial loss:0.437405 | Final loss:0.163928

Fold Number: 3/5

Initial loss:0.441808 | Final loss:0.136794

Fold Number: 4/5

Initial loss:0.446758 | Final loss:0.141017

Fold Number: 5/5

Initial loss:0.450937 | Final loss:0.139801

Mean train accuracy: 84.718750

Mean dev accuracy: 83.000000

Total time taken in 5-folds CV on given set of hyperparameters: 811.713399 seconds

--------------------------------------------------

Iteration Number: 59

Number of Epochs: 50

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.451829 | Final loss:0.149027

Fold Number: 2/5

Initial loss:0.449229 | Final loss:0.144982

Fold Number: 3/5

Initial loss:0.438258 | Final loss:0.143272

Fold Number: 4/5

Initial loss:0.440466 | Final loss:0.140574

Fold Number: 5/5

Initial loss:0.453115 | Final loss:0.144866

Mean train accuracy: 85.656250

Mean dev accuracy: 82.958333

Total time taken in 5-folds CV on given set of hyperparameters: 801.077369 seconds

--------------------------------------------------

Iteration Number: 60

Number of Epochs: 50

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Fold Number: 1/5

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.446523 | Final loss:0.137578

Fold Number: 2/5

Initial loss:0.445172 | Final loss:0.138242

Fold Number: 3/5

Initial loss:0.447319 | Final loss:0.141913

Fold Number: 4/5

Initial loss:0.450528 | Final loss:0.143459

Fold Number: 5/5

Initial loss:0.440060 | Final loss:0.137309

Mean train accuracy: 86.041667

Mean dev accuracy: 83.791667

Total time taken in 5-folds CV on given set of hyperparameters: 798.456491 seconds

Changing methods here

--------------------------------------------------

Iteration Number: 1

Number of Epochs: 50

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.459827 | Final loss:0.446588

Train accuracy: 14.114583

Dev accuracy: 11.250000

Total time taken in 5-folds CV on given set of hyperparameters: 172.667468 seconds

--------------------------------------------------

Iteration Number: 2

Number of Epochs: 50

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.460554 | Final loss:0.447577

Train accuracy: 15.000000

Dev accuracy: 10.625000

Total time taken in 5-folds CV on given set of hyperparameters: 165.907405 seconds

--------------------------------------------------

Iteration Number: 3

Number of Epochs: 50

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 50

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.460644 | Final loss:0.446887

Train accuracy: 16.354167

Dev accuracy: 11.041667

Total time taken in 5-folds CV on given set of hyperparameters: 169.892315 seconds

--------------------------------------------------

Iteration Number: 4

Number of Epochs: 50

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.457932 | Final loss:0.445334

Train accuracy: 16.093750

Dev accuracy: 10.000000

Total time taken in 5-folds CV on given set of hyperparameters: 165.945735 seconds

--------------------------------------------------

Iteration Number: 5

Number of Epochs: 50

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.455016 | Final loss:0.406463

Train accuracy: 37.604167

Dev accuracy: 8.333333

Total time taken in 5-folds CV on given set of hyperparameters: 173.109801 seconds

--------------------------------------------------

Iteration Number: 6

Number of Epochs: 50

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.453991 | Final loss:0.408500

Train accuracy: 37.500000

Dev accuracy: 8.333333

Total time taken in 5-folds CV on given set of hyperparameters: 166.901206 seconds

--------------------------------------------------

Iteration Number: 7

Number of Epochs: 50

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 50

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.454560 | Final loss:0.404301

Train accuracy: 37.239583

Dev accuracy: 12.916667

Total time taken in 5-folds CV on given set of hyperparameters: 170.298896 seconds

--------------------------------------------------

Iteration Number: 8

Number of Epochs: 50

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.453337 | Final loss:0.404764

Train accuracy: 36.927083

Dev accuracy: 9.791667

Total time taken in 5-folds CV on given set of hyperparameters: 165.664031 seconds

--------------------------------------------------

Iteration Number: 9

Number of Epochs: 50

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 1

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.453116 | Final loss:0.071524

Train accuracy: 90.052083

Dev accuracy: 8.958333

Total time taken in 5-folds CV on given set of hyperparameters: 172.120519 seconds

--------------------------------------------------

Iteration Number: 10

Number of Epochs: 50

Learning Rate: 0.001

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 10

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.453340 | Final loss:0.069505

Train accuracy: 89.375000

Dev accuracy: 10.833333

Total time taken in 5-folds CV on given set of hyperparameters: 165.738162 seconds

--------------------------------------------------

Iteration Number: 11

Number of Epochs: 50

Learning Rate: 0.01

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 50

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.452857 | Final loss:0.076003

Train accuracy: 88.906250

Dev accuracy: 7.083333

Total time taken in 5-folds CV on given set of hyperparameters: 169.543934 seconds

--------------------------------------------------

Iteration Number: 12

Number of Epochs: 50

Learning Rate: 0.1

Cost Function used: Sse

Learning Rate is being updated?: Yes

Early Stopping Regularization being used?: Yes

Batch Size: 100

Layer type: Fc

Input Dimension: 784 Output Dimension: 128

Layer type: Activation

Activation Function: Relu

Layer type: Fc

Input Dimension: 128 Output Dimension: 10

Layer type: Softmax

Input Dimension: 10

Training Size: (1920, 784)

Validation Size: (480, 784)

Initial loss:0.453919 | Final loss:0.069019

Train accuracy: 89.479167

Dev accuracy: 9.791667

Total time taken in 5-folds CV on given set of hyperparameters: 165.746456 seconds

From above, the biggest difference in accuracies came because of the cost function/loss function that I was using.  
SSE performed significantly better than MSE as indicated. This might be because 1/N times the SSE cost function is the MSE cost function. So, the gradient descent update rule with small gradients is

θk+1 ← θk − α⋅∇θ *L*(θ)

Therefore, as N grows, the SSE cost function grows too, and so does its gradient. Conversely, while the MSE cost function *may* also grow in value as N increases, it does not grow as fast as the SSE cost function, and so its gradients are also growing slowly. This is probably the reason why SSE outperforms MSE in my case. This is validated by seeing how much the loss decreases with time

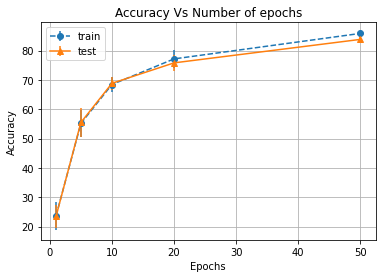
Cross entropy gives much better performance in general as I saw in the case of a standard library. I could not use it even if there is an option, probably because of my initialization of weights. Because of those, some probabilities were so low or so high that the cost/error kept hitting NaN and therefore, I could not use it in my implementation.

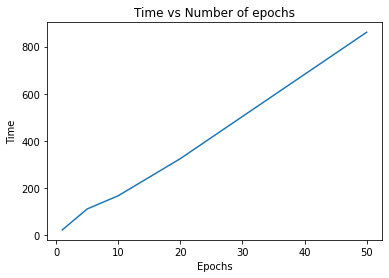
For comparisons, I have plotted for 1 hyperparameter at a time as there were too many variables for simultaneous plotting (i.e., 6D plots etc.)

Among all these models, the best performing one was the 52nd iteration with a dev set accuracy of 92.125%. Therefore, for plotting trends for various hyperparameters, I will keep rest of the values constant to:  
  
Number of Epochs: 20   
Learning Rate: 0.001  
Cost Function used: SSE  
Learning Rate is being updated? Yes   
Early Stopping Regularization being used? Yes   
Batch Size: 1

For number of epochs, the following results are produced. The training accuracy increases with epochs but after some time, the test accuracy starts decreasing i.e., as epochs increase, the model goes from underfitting to overfitting.

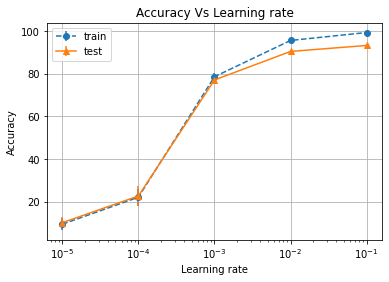
As we can see in the plot below it, the time taken by the model to train, as expected, increases with number of epochs. This is because we need to go over the training data that many times. This is the largest increase among the plots

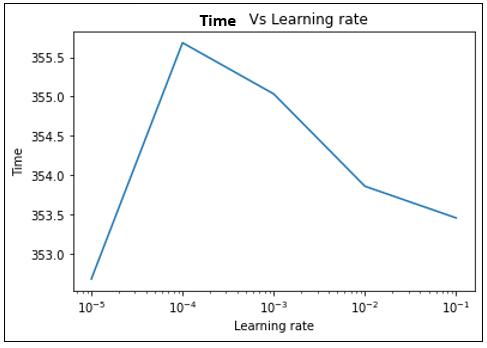




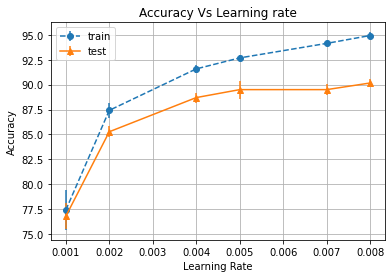
For learning rate, the accuracy trend is similar as before, it goes from underfitting to overfitting. The learning rate affects how quickly our model can converge to a local-minima. Increasing this may affect the training in a way that it converges to a local-minima, instead of a global-minima.

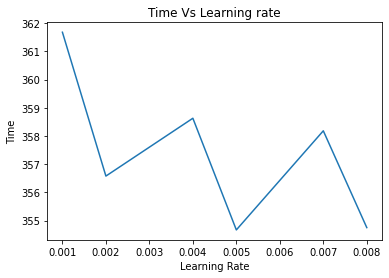
Further, for the time plot, we don’t see any clear trend. This could partially be by the virtue of the dynamic learning rate policy that I had adopted. After a few iterations, the initial learning rate has lower bearing on the training time





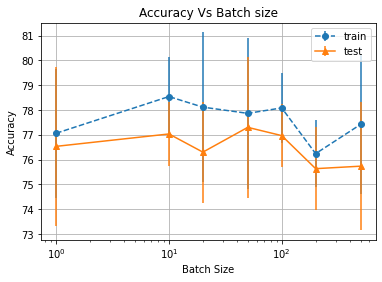
Here is another set of plots for the learning rate but on a linear scale. Too small learning rate make learning slow and under fitting and too big learning rate leads to divergence.

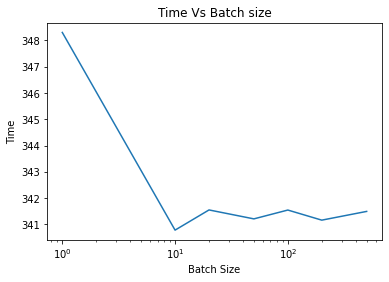




Finally, I have plots against batch size. One peculiar observation here is that there is a lot of error among the 5-folds that I have used. Even though the mean value is more or less constant, the range they covered on being trained on different parts of data varies a lot.

For time, there isn’t much difference either as the max difference is of mere 7 seconds





For the next experiment with **2 layers**, I don’t use 5-folds CV as it would naturally take a lot of time. However, I have used them for the trends.  
I am using tanh here but in my setting, it isn’t advisable because it is known to cause vanishing gradient problem at times. That coupled with my dynamic learning rate might make learning way worse than we want. I tried with both tanh and relu and found relu to perform better.

Layer type: Fully Connected   
Input Dimension: 784 Output Dimension: 256  
Layer type: Activation Layer  
Activation Function: ReLU  
Layer type: Fully Connected  
Input Dimension: 256 Output Dimension: 64  
Layer type: Activation Layer  
Activation Function: ReLU  
Layer type: Fully Connected  
Input Dimension: 64 Output Dimension: 10  
Layer type: Softmax   
Input Dimension: 10

Finally, I repeated the experiment that I did earlier and found better accuracy than before. <Results are skipped>

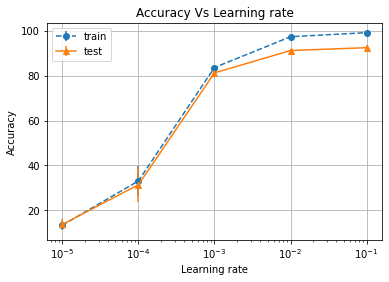
I used the following hyperparameters which gave 90.14% accuracy on test set and 97.98% on training set:

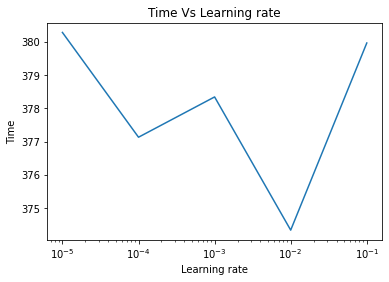
ReLU layer

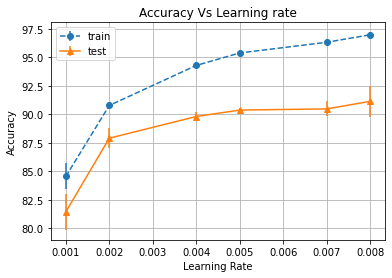
Number of Epochs: 50   
Learning Rate: 0.1  
Cost Function used: SSE  
Learning Rate is being updated? Yes   
Early Stopping Regularization being used? Yes   
Batch Size: 30

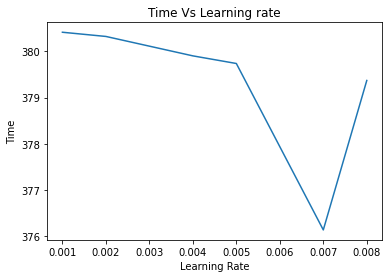
Here is the analysis of individual hyperparameters:

For learning rate (log and linear), we see a similar trend as before. Note that the test accuracy isn’t dropping from an earlier value, but constantly rising. Implying that there is no overfitting happening here. For time taken, it is somewhat irregular but max difference is of 1.5% and is therefore insignificant.

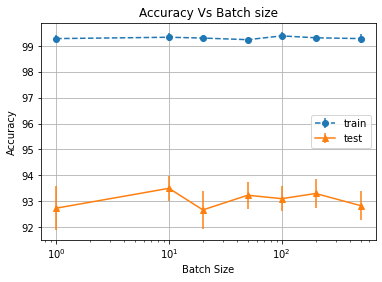
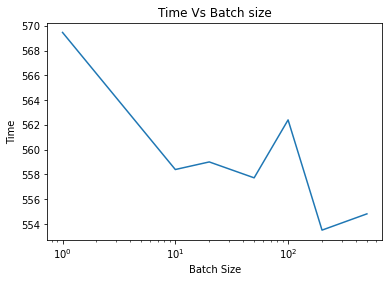




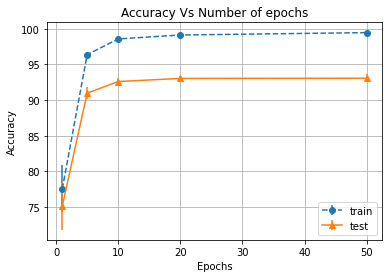
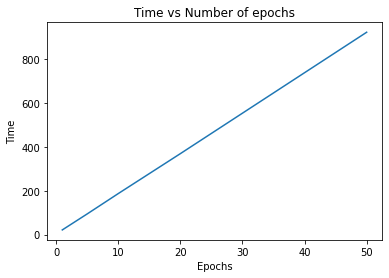




For batch size, we see that the accuracy is more or less independent of the choice of batch size, however time decreases as batch size increases, which is also as expected

For number of epochs, the trend start with underfitting and gradually moves to rightly fit. We see that the test accuracy doesn’t increase after a point, but that is still not overfitting as it is not harming the performance.

The most significant observation here was that the accuracy with the right hyperparameters increased.

One thing that I noticed from the trends was that there wasn’t much dip in test set accuracies, which means that at least in my models there was not much problem of overfitting and that L2 regularization will most probably not improve the performance by much.

For the error analysis and standard library part, I went with the model with best test set accuracy on hyperparameters. The following model was used:

keras.layers.Dense(256, activation='relu'),

keras.layers.Dense(64, activation='relu'),

keras.layers.Dense(10, activation='softmax')

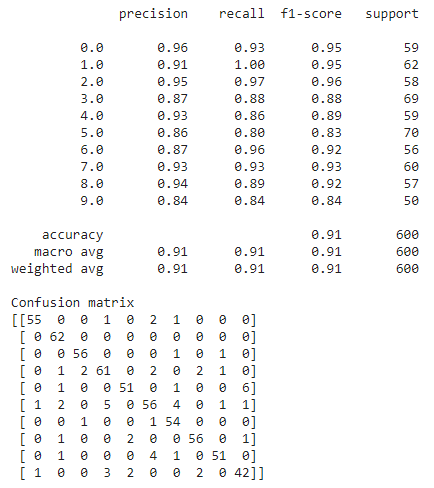
Along with this, an SGD optimizer was used with a learning rate of 0.1, and categorical\_crossentropy loss function as it is known to perform better. Finally, I used mean\_squared\_error metrics.

The results were quite astonishing. The training time was a mere 9 seconds which was must faster than my implementation. One prime reason can be that keras now works with tensorflow which is created by Google, and offers GPU support for that on Google Colab by itself, along with a tensorflow processing unit (TPU) of its own. These speed up the process by a lot. Another can be in-built parallelization in keras.

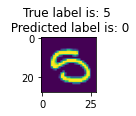
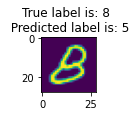
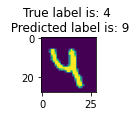
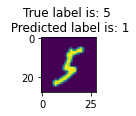
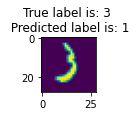
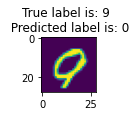
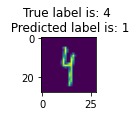
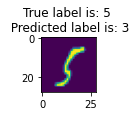
Finally, we get a slightly better results than my own implementation (possibly because of the loss function) with a training accuracy of 98.1% and test accuracy of ~91%  
Some more possible reasons can be better initialisation of weight matrix and using a modified version of ReLU

**ERROR ANALYSIS**

Finally, I will do bit of an error analysis

****

We see that all the labels are almost equally distributed with label 5 having highest frequency, therefore we can be assured that our model works uniformly well on all the labels and not on specific few.  
Next, from the confusion matrix, I see that the main points of confusion for the model is the number 9. 8 times a number was wrongly predicted as 9 and 8 times 9 was predicted as some other number. Here are some incorrectly classified images:



A lot of images are understandably misclassified. For the others, a possible reason can be that many of the images when rotated, will look a lot like the predicted labels

Finally, let’s look at some outputs of hidden layers:

**  
  
  
  
  
**

To most of us, these look like random noises, but for the model, these are some really high space representation of the input data that helps it classify these. We see that the input layer is distinguishable to the human eye, which after going through a tanh layer results in something that is not understandable to us. But after the ReLU layer, it seems to have started assigning each of them some separate identity that helps it distinguish between them. And then it outputs the labels as we require it to.

**References:**

* <https://stats.stackexchange.com/questions/539137/why-use-mse-instead-of-sse-as-cost-function-in-linear-regression>
* <https://towardsdatascience.com/math-neural-network-from-scratch-in-python-d6da9f29ce65>
* <https://colab.research.google.com/github/casperbh96/Neural-Network-From-Scratch/blob/master/NN_From_Scratch.ipynb#scrollTo=5IucezTTXiSM>
* <https://github.com/sar-gupta/neural-network-from-scratch/blob/master/neuralnetwork.py>
* <https://hackernoon.com/dl03-gradient-descent-719aff91c7d6>
* <https://scikit-learn.org/stable/auto_examples/semi_supervised/plot_label_propagation_digits_active_learning.html#sphx-glr-auto-examples-semi-supervised-plot-label-propagation-digits-active-learning-py>