(a) I have implemented the DNN using SGD. This algorithm works for any number of hidden layers (Even with zero!). Just edit the hyper parameters as you wish from the **main** function. If you want to tune the hyper parameters, add each parameter to the corresponding list in the **main** function, and change the key word argument of **dnn** class to **True**. If the tune parameter is set to false, the network chooses a random hyperparameter set.

(b) At 3 Hidden layers and 64 hidden units per each hidden layer, check_grad function returns the following output:

Check grad: 8.48675979004169e-07

Hyper Parameters used are:
Batch Size = 254,
Epsilon = 0.09,
Epochs = 300,
Alpha = 0.0025,
Hidden layers= 7,
Hidden units at each hidden layer = [512, 512, 512, 512, 512, 512, 512]

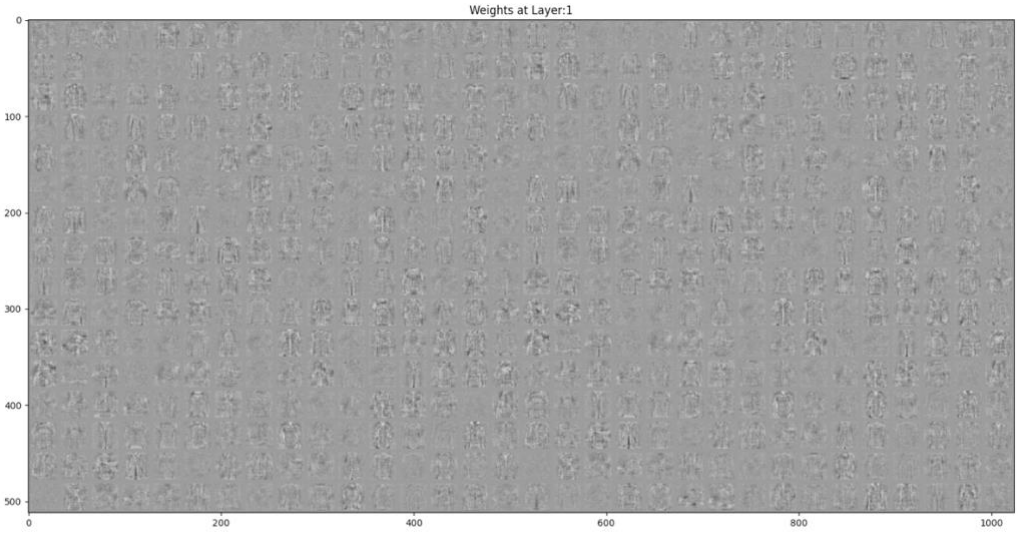
The following is the loss for last 20 epochs, and final test loss and accuracy

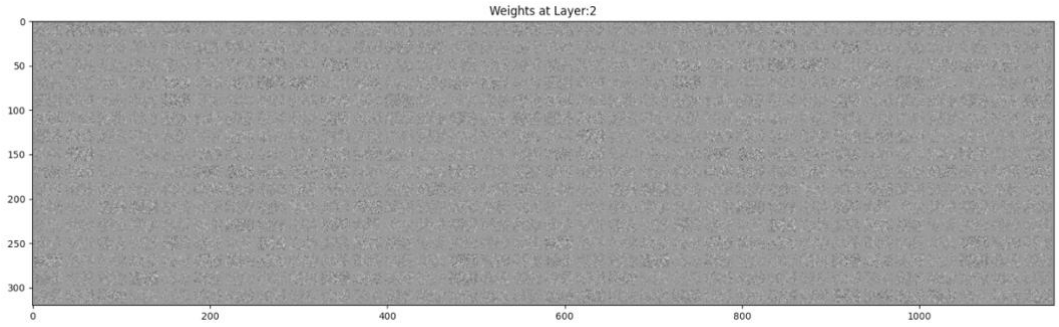
Epoch 280 , and loss 0.040836453339212084
Epoch 281 , and loss 0.04080345183386274
Epoch 282 , and loss 0.04077048268356996
Epoch 283 , and loss 0.04073753906553885
Epoch 284 , and loss 0.04070462817376836
Epoch 285 , and loss 0.040671748133887256
Epoch 286 , and loss 0.04063889139504536
Epoch 287 , and loss 0.04060606970019438
Epoch 288 , and loss 0.04057327370535337
Epoch 289 , and loss 0.040540507913879496
Epoch 290 , and loss 0.04050777167114746
Epoch 291 , and loss 0.040475065230115015
Epoch 292 , and loss 0.040442393311240075
Epoch 293 , and loss 0.0404097444967561
Epoch 294 , and loss 0.04037712921089954
Epoch 295 , and loss 0.04034454127191848
Epoch 296 , and loss 0.040311985156857875
Epoch 297 , and loss 0.04027945925601887
Epoch 298 , and loss 0.04024696299719237
Epoch 299 , and loss 0.040214497047978345
Epoch 300 , and loss 0.04018205912911439
Training Error is 0.04018205912911439
Test Error is 0.8894396867771397
Test accuracy 90.02

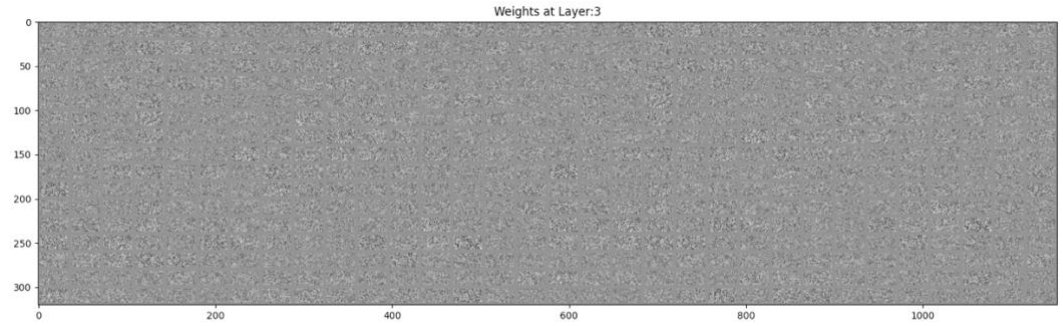
Screenshot:

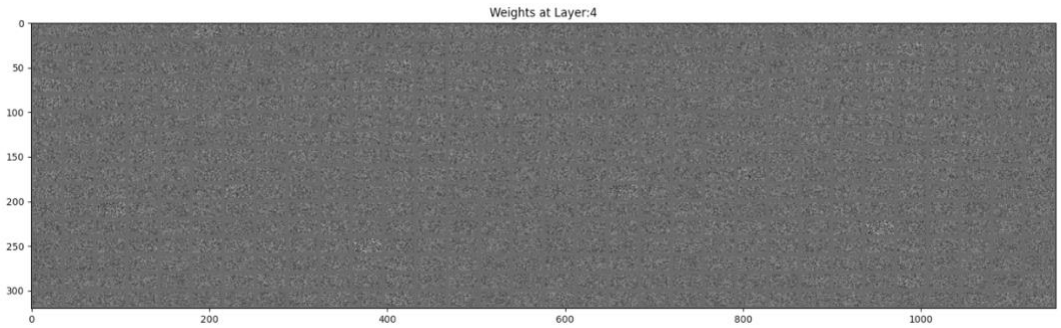
```
Epoch 274 , and loss 0.04103506288119193
Epoch 275 , and loss 0.0410018869452682
Epoch 276 , and loss 0.040968740898962366
Epoch 277 , and loss 0.04093562644495439
Epoch 278 , and loss 0.040902539721332606
Epoch 279 , and loss 0.04086948036897053
Epoch 280 , and loss 0.040836453339212084
Epoch 281 , and loss 0.04080345183386274
Epoch 282 , and loss 0.04077048268356996
Epoch 283 , and loss 0.04073753906553885
Epoch 284 , and loss 0.04070462817376836
Epoch 285 , and loss 0.040671748133887256
Epoch 286 , and loss 0.04063889139504536
Epoch 287 , and loss 0.04060606970019438
Epoch 288 , and loss 0.04057327370535337
Epoch 289 , and loss 0.040540507913879496
Epoch 290 , and loss 0.04050777167114746
Epoch 291 , and loss 0.040475065230115015
Epoch 292 , and loss 0.040442393311240075
Epoch 293 , and loss 0.0404097444967561
Epoch 294 , and loss 0.04037712921089954
Epoch 295 , and loss 0.04034454127191848
Epoch 296 , and loss 0.040311985156857875
Epoch 297 , and loss 0.04027945925601887
Epoch 298 , and loss 0.04024696299719237
Epoch 299 , and loss 0.040214497047978345
Epoch 300 , and loss 0.04018205912911439
Training Error is  0.04018205912911439
Test Error is  0.8894396867771397
Test accuracy 90.02
```

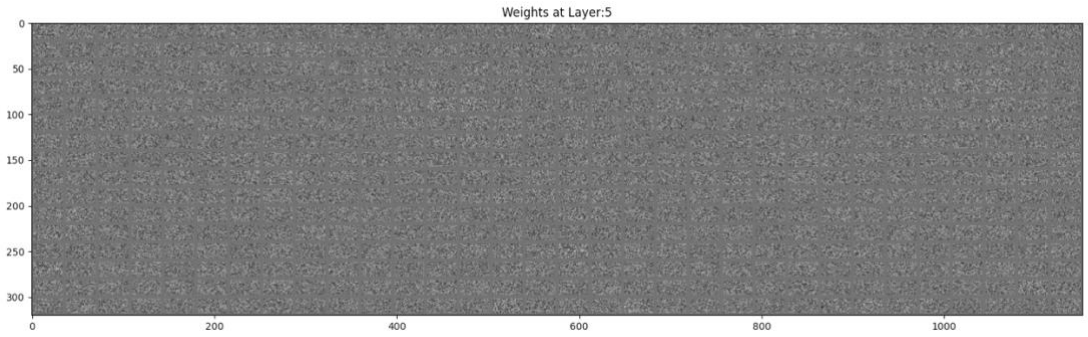
Weight Plots for each layer at these hyper parameters:

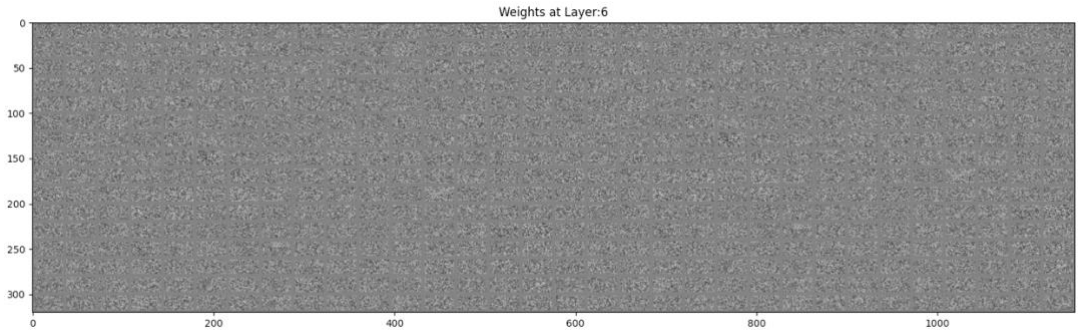


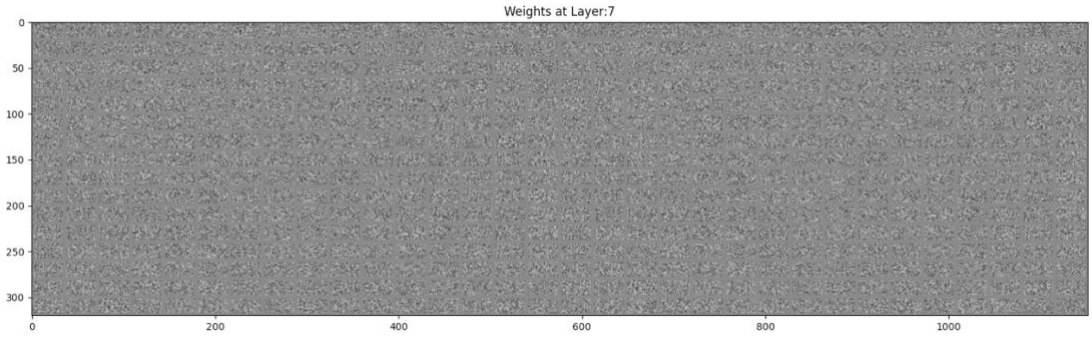

Weights at Layer:2
0
50
100
150
200
250
300
0
200
400
600
800
1000

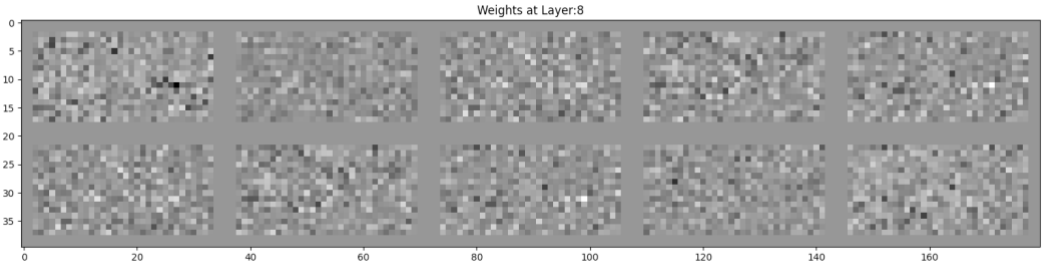

Weights at Layer:3
0
50
100
150
200
250
300
0
200
400
600
800
1000


Weights at Layer:4
0
50
100
150
200
250
300
0
200
400
600
800
1000


Weights at Layer:5
0
50
100
150
200
250
300
0
200
400
600
800
1000


Weights at Layer:6
0
50
100
150
200
250
300
0
200
400
600
800
1000





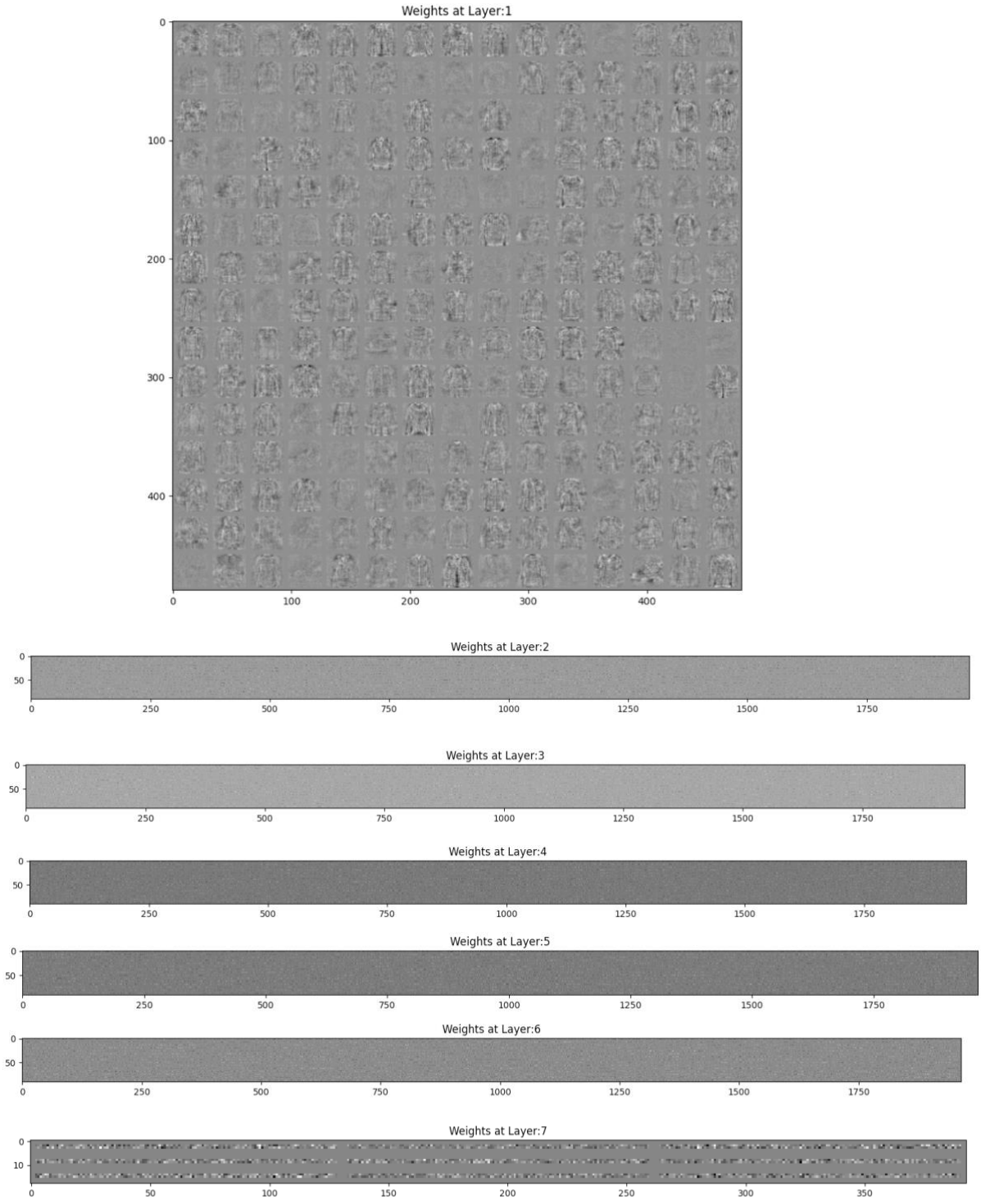
Hyper Parameters used are:
Batch Size = 254,
Epsilon = 0.09,
Epochs = 200,
Alpha = 0.0025,
Hidden layers= 6,
Hidden units at each hidden layer = [254, 254, 254, 254, 254, 254]

The following is the loss for last 20 epochs, and final test loss and accuracy

Epoch 181 , and loss 0.03079890284608955
Epoch 182 , and loss 0.030778589952512498
Epoch 183 , and loss 0.03075770897291907
Epoch 184 , and loss 0.03073550271534558
Epoch 185 , and loss 0.03071373218922566
Epoch 186 , and loss 0.03069290763156216
Epoch 187 , and loss 0.03067186744579173
Epoch 188 , and loss 0.030650605454243765
Epoch 189 , and loss 0.030629358867598048
Epoch 190 , and loss 0.030608186227203838
Epoch 191 , and loss 0.030586724371946556
Epoch 192 , and loss 0.03056553952436401
Epoch 193 , and loss 0.030543902117510735
Epoch 194 , and loss 0.030522602382308132
Epoch 195 , and loss 0.030501143977252572
Epoch 196 , and loss 0.030479403414290156
Epoch 197 , and loss 0.03045764993283975
Epoch 198 , and loss 0.030435960465143828
Epoch 199 , and loss 0.030414091530481933

Epoch 200 , and loss 0.030392264417619132
Training Error is <u>0.030392264417619132</u>
Test Error is <u>1.0013279776909778</u>
Test accuracy <u>89.28999999999999</u>

Weight Plots for each layer at these hyper parameters:



Hyper Parameters used are:
Batch Size = 254,
Epsilon = 0.09,
Epochs = 1,
Alpha = 0.0025,
Hidden layers= 0,
Hidden units at each hidden layer = []

The following is the loss for last epoch, and final test loss and accuracy

Epoch 1 , and loss 0.5853625684133672
Training Error is 0.5853625684133672
Test Error is 0.6099911842736919
Test accuracy 78.96999999999997

Weight Plots for each layer at these hyper parameters: