Machine Learning
Assignment 6 Report

Submitted By:

- Anurag Saraswat (M20CS066)

For this assignment we use Stochastic Gradient Descent as our optimization methodology. Learning rate is kept at 0.1 and batch size fixed at 100. Crossentrophy is used. MNIST dataset is used. Data has 784 features.60,000 samples are used for training and 10,000 samples are used for testing. For validation 10 percent of training data is used.

1. Varying the number of hidden layers from 0 to 2. Choose the number of neurons in the hidden layers such that the total number of parameters in the network remains the same.

0 Hidden Layer:

In this no hidden layer is used. Only Input and Output layer is present. Detail of model is as below

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_3 (InputLayer)         [(None, 784)]             0
_________________________________________________________________
Output_Layer (Dense)         (None, 10)                7850
=================================================================
Total params: 7,850
Trainable params: 7,850
Non-trainable params: 0
_________________________________________________________________
```

For Validation we use 10 percent of training data. Validation accuracy and Validation using 0 hidden layer error at epoch is listed below :

```
Epoch 1/20
540/540 - 1s - loss: 1.2198 - accuracy: 0.7224 - val_loss: 0.7141 - val_accuracy: 0.8702
Epoch 2/20
540/540 - 1s - loss: 0.6819 - accuracy: 0.8458 - val_loss: 0.5265 - val_accuracy: 0.8927
Epoch 3/20
540/540 - 1s - loss: 0.5657 - accuracy: 0.8631 - val_loss: 0.4530 - val_accuracy: 0.8992
Epoch 4/20
540/540 - 1s - loss: 0.5098 - accuracy: 0.8718 - val_loss: 0.4126 - val_accuracy: 0.9047
Epoch 5/20
540/540 - 1s - loss: 0.4756 - accuracy: 0.8771 - val_loss: 0.3869 - val_accuracy: 0.9070
Epoch 6/20
540/540 - 1s - loss: 0.4521 - accuracy: 0.8807 - val_loss: 0.3687 - val_accuracy: 0.9088
Epoch 7/20
540/540 - 1s - loss: 0.4345 - accuracy: 0.8849 - val_loss: 0.3550 - val_accuracy: 0.9110
Epoch 8/20
540/540 - 1s - loss: 0.4207 - accuracy: 0.8875 - val_loss: 0.3439 - val_accuracy: 0.9130
Epoch 9/20
540/540 - 1s - loss: 0.4096 - accuracy: 0.8897 - val_loss: 0.3354 - val_accuracy: 0.9153
Epoch 10/20
540/540 - 1s - loss: 0.4003 - accuracy: 0.8918 - val_loss: 0.3281 - val_accuracy: 0.9155
Epoch 11/20
540/540 - 1s - loss: 0.3925 - accuracy: 0.8934 - val_loss: 0.3218 - val_accuracy: 0.9175
Epoch 12/20
540/540 - 1s - loss: 0.3857 - accuracy: 0.8953 - val_loss: 0.3165 - val_accuracy: 0.9167
Epoch 13/20
540/540 - 1s - loss: 0.3797 - accuracy: 0.8965 - val_loss: 0.3119 - val_accuracy: 0.9182
Epoch 14/20
540/540 - 1s - loss: 0.3745 - accuracy: 0.8977 - val_loss: 0.3079 - val_accuracy: 0.9193
Epoch 15/20
540/540 - 1s - loss: 0.3698 - accuracy: 0.8985 - val_loss: 0.3041 - val_accuracy: 0.9193
Epoch 16/20
540/540 - 1s - loss: 0.3656 - accuracy: 0.8994 - val_loss: 0.3011 - val_accuracy: 0.9187
Epoch 17/20
540/540 - 1s - loss: 0.3617 - accuracy: 0.9001 - val_loss: 0.2979 - val_accuracy: 0.9210
Epoch 18/20
540/540 - 1s - loss: 0.3582 - accuracy: 0.9008 - val_loss: 0.2951 - val_accuracy: 0.9212
Epoch 19/20
540/540 - 1s - loss: 0.3550 - accuracy: 0.9016 - val_loss: 0.2929 - val_accuracy: 0.9222
Epoch 20/20
540/540 - 1s - loss: 0.3520 - accuracy: 0.9020 - val_loss: 0.2904 - val_accuracy: 0.9223
100/100 [==============================] - 0s 1ms/step - loss: 0.3280 - accuracy: 0.9108
```

1 Hidden Layer:

Input, output and one hidden layer with relu activation function is present. Detail of model is as below:

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         [(None, 784)]             0
_________________________________________________________________
Hidden_Layer_1 (Dense)       (None, 100)               78500
_________________________________________________________________
Output_Layer (Dense)         (None, 10)                1010
=================================================================
Total params: 79,510
Trainable params: 79,510
Non-trainable params: 0
_________________________________________________________________
```

For Validation we use 10 percent of training data. Validation accuracy and Validation using 1 hidden layer error at epoch is listed below :

```
Epoch 1/20
540/540 - 1s - loss: 1.1633 - accuracy: 0.7163 - val_loss: 0.5610 - val_accuracy: 0.8862
Epoch 2/20
540/540 - 1s - loss: 0.5317 - accuracy: 0.8676 - val_loss: 0.3871 - val_accuracy: 0.9090
Epoch 3/20
540/540 - 1s - loss: 0.4285 - accuracy: 0.8859 - val_loss: 0.3301 - val_accuracy: 0.9158
Epoch 4/20
540/540 - 1s - loss: 0.3831 - accuracy: 0.8949 - val_loss: 0.3012 - val_accuracy: 0.9205
Epoch 5/20
540/540 - 1s - loss: 0.3559 - accuracy: 0.9006 - val_loss: 0.2827 - val_accuracy: 0.9247
Epoch 6/20
540/540 - 1s - loss: 0.3365 - accuracy: 0.9061 - val_loss: 0.2701 - val_accuracy: 0.9260
Epoch 7/20
540/540 - 1s - loss: 0.3215 - accuracy: 0.9098 - val_loss: 0.2583 - val_accuracy: 0.9275
Epoch 8/20
540/540 - 1s - loss: 0.3090 - accuracy: 0.9131 - val_loss: 0.2511 - val_accuracy: 0.9308
Epoch 9/20
540/540 - 1s - loss: 0.2984 - accuracy: 0.9164 - val_loss: 0.2417 - val_accuracy: 0.9335
Epoch 10/20
540/540 - 1s - loss: 0.2887 - accuracy: 0.9190 - val_loss: 0.2342 - val_accuracy: 0.9355
Epoch 11/20
540/540 - 1s - loss: 0.2798 - accuracy: 0.9214 - val_loss: 0.2288 - val_accuracy: 0.9370
Epoch 12/20
540/540 - 1s - loss: 0.2719 - accuracy: 0.9236 - val_loss: 0.2209 - val_accuracy: 0.9380
Epoch 13/20
540/540 - 1s - loss: 0.2645 - accuracy: 0.9262 - val_loss: 0.2157 - val_accuracy: 0.9400
Epoch 14/20
540/540 - 1s - loss: 0.2574 - accuracy: 0.9279 - val_loss: 0.2099 - val_accuracy: 0.9407
Epoch 15/20
540/540 - 1s - loss: 0.2509 - accuracy: 0.9300 - val_loss: 0.2040 - val_accuracy: 0.9432
Epoch 16/20
540/540 - 1s - loss: 0.2445 - accuracy: 0.9315 - val_loss: 0.1995 - val_accuracy: 0.9438
Epoch 17/20
540/540 - 1s - loss: 0.2387 - accuracy: 0.9334 - val_loss: 0.1952 - val_accuracy: 0.9460
Epoch 18/20
540/540 - 1s - loss: 0.2330 - accuracy: 0.9351 - val_loss: 0.1905 - val_accuracy: 0.9465
Epoch 19/20
540/540 - 1s - loss: 0.2279 - accuracy: 0.9361 - val_loss: 0.1877 - val_accuracy: 0.9490
Epoch 20/20
540/540 - 1s - loss: 0.2230 - accuracy: 0.9376 - val_loss: 0.1833 - val_accuracy: 0.9517
100/100 [==============================] - 0s 1ms/step - loss: 0.2143 - accuracy: 0.9401
```

2 Hidden Layer:

Input, output and two hidden layers with relu activation function is present. Detail of model is as below:

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 784)]             0
_________________________________________________________________
Hidden_Layer_1 (Dense)       (None, 100)               78500
_________________________________________________________________
Hidden_Layer_2 (Dense)       (None, 100)               10100
_________________________________________________________________
Output_Layer (Dense)         (None, 10)                1010
=================================================================
Total params: 89,610
Trainable params: 89,610
Non-trainable params: 0
```

For Validation we use 10 percent of training data. Validation accuracy and Validation using 1 hidden layer error at epoch is listed below :

```
Epoch 1/20
540/540 - 2s - loss: 1.2561 - accuracy: 0.6706 - val_loss: 0.5389 - val_accuracy: 0.8770
Epoch 2/20
540/540 - 1s - loss: 0.4946 - accuracy: 0.8692 - val_loss: 0.3458 - val_accuracy: 0.9113
Epoch 3/20
540/540 - 1s - loss: 0.3837 - accuracy: 0.8937 - val_loss: 0.2916 - val_accuracy: 0.9205
Epoch 4/20
540/540 - 1s - loss: 0.3392 - accuracy: 0.9042 - val_loss: 0.2625 - val_accuracy: 0.9262
Epoch 5/20
540/540 - 1s - loss: 0.3116 - accuracy: 0.9111 - val_loss: 0.2446 - val_accuracy: 0.9297
Epoch 6/20
540/540 - 1s - loss: 0.2904 - accuracy: 0.9176 - val_loss: 0.2311 - val_accuracy: 0.9338
Epoch 7/20
540/540 - 1s - loss: 0.2737 - accuracy: 0.9218 - val_loss: 0.2171 - val_accuracy: 0.9400
Epoch 8/20
540/540 - 1s - loss: 0.2592 - accuracy: 0.9264 - val_loss: 0.2073 - val_accuracy: 0.9393
Epoch 9/20
540/540 - 1s - loss: 0.2468 - accuracy: 0.9295 - val_loss: 0.1996 - val_accuracy: 0.9437
Epoch 10/20
540/540 - 1s - loss: 0.2356 - accuracy: 0.9329 - val_loss: 0.1901 - val_accuracy: 0.9457
Epoch 11/20
540/540 - 1s - loss: 0.2259 - accuracy: 0.9355 - val_loss: 0.1831 - val_accuracy: 0.9477
Epoch 12/20
540/540 - 1s - loss: 0.2165 - accuracy: 0.9381 - val_loss: 0.1769 - val_accuracy: 0.9510
Epoch 13/20
540/540 - 1s - loss: 0.2085 - accuracy: 0.9407 - val_loss: 0.1697 - val_accuracy: 0.9543
Epoch 14/20
540/540 - 1s - loss: 0.2008 - accuracy: 0.9425 - val_loss: 0.1649 - val_accuracy: 0.9553
Epoch 15/20
540/540 - 1s - loss: 0.1936 - accuracy: 0.9446 - val_loss: 0.1600 - val_accuracy: 0.9588
Epoch 16/20
540/540 - 1s - loss: 0.1870 - accuracy: 0.9464 - val_loss: 0.1567 - val_accuracy: 0.9577
Epoch 17/20
540/540 - 1s - loss: 0.1808 - accuracy: 0.9481 - val_loss: 0.1508 - val_accuracy: 0.9597
Epoch 18/20
540/540 - 1s - loss: 0.1748 - accuracy: 0.9505 - val_loss: 0.1479 - val_accuracy: 0.9610
Epoch 19/20
540/540 - 1s - loss: 0.1694 - accuracy: 0.9514 - val_loss: 0.1444 - val_accuracy: 0.9632
Epoch 20/20
540/540 - 1s - loss: 0.1642 - accuracy: 0.9528 - val_loss: 0.1418 - val_accuracy: 0.9630
100/100 [==============================] - 0s 2ms/step - loss: 0.1672 - accuracy: 0.9510
```

Accuracy Comparison of above cases:

It is clearly observed that as number of hidden layer increased from 0 to 2 accuracy increases

Accuracy achieved Using 0 Hidden Layer: 91.08
Accuracy achieved Using 1 Hidden Layer: 94.01
Accuracy achieved Using 2 Hidden Layer: 95.1

2. Trying sigmoid and relu activation functions for the hidden layer nodes.

In the first case we use Input, output and two hidden layers with relu activation function. In the Second case we use Input, output and two hidden layers with sigmoid activation function. 20 epochs of batch size 100 is performed. Accuracy obtained in both the cases are listed below:

- Accuracy achieved Using Relu activation function for Hidden Layer: 95.43
- Accuracy achieved Using Sigmoid activation function for Hidden Layer: 86.89

It is observed that the relu activation function for the hidden layer gives better accuracy than sigmoid.

3. Not using any nonlinearity in the network

In this case we use Input, output and two hidden layers with linear activation function. 20 epochs of batch size 100 is performed. Accuracy obtained in this cases is listed below:

- Accuracy achieved Using Linear activation function for Hidden Layer: 92.15