10/29/24

Neural Network - HW6

Gokhulnath
Thirumaran
675086474

1) a) Given:

$$f_i = \frac{e^{z_i}}{\sum_{i'} e^{z_{i'}}}$$

$$l(y, f) = \sum_i y_i \log \frac{1}{f_i} \quad -①$$

where y is non negative, $\sum y_i = 1$

Computing the gradient wrt f_i in ①

$$\frac{\partial l}{\partial f_i} = -\frac{y_i}{f_i} \quad -②$$

b) Backpropagation equation from $\nabla_f l$ to $\nabla_z l$.

Since we know that $f_i = \frac{e^{z_i}}{\sum_{i'} e^{z_{i'}}}$

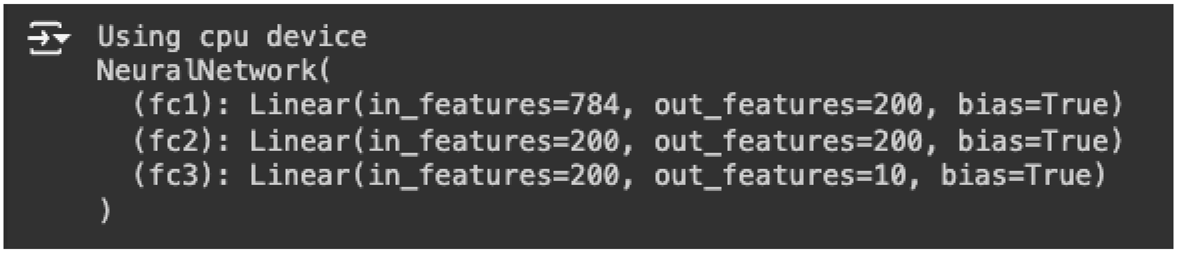
Using chainrule we can derive that

$$\frac{\partial l}{\partial z_i} = \sum_j \frac{\partial l}{\partial f_i} \cdot \frac{\partial f_i}{\partial z_i} \quad -③$$

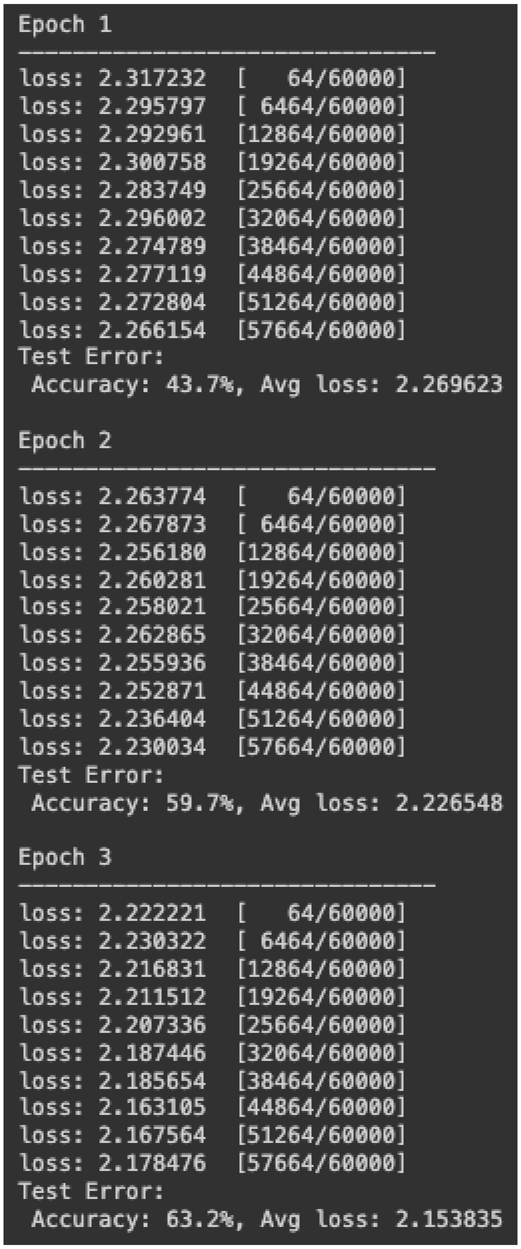
c) Subs. $\textcircled{2}$ & $\textcircled{3}$.

$$\nabla_{z} l = \frac{\partial l}{\partial z_i} = f_i - y_i$$

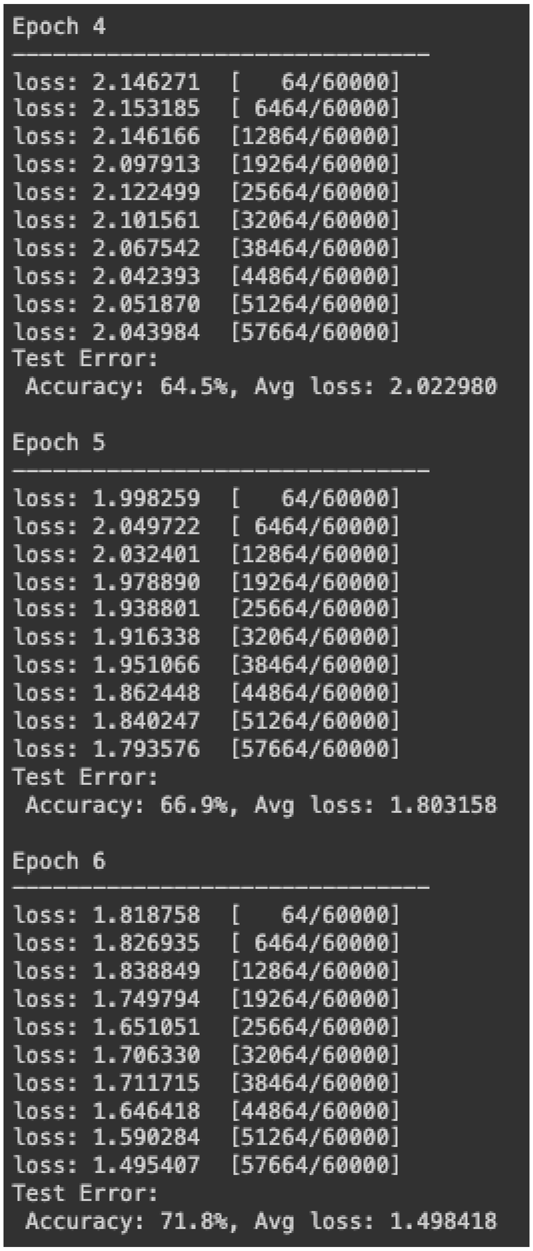
2. b)



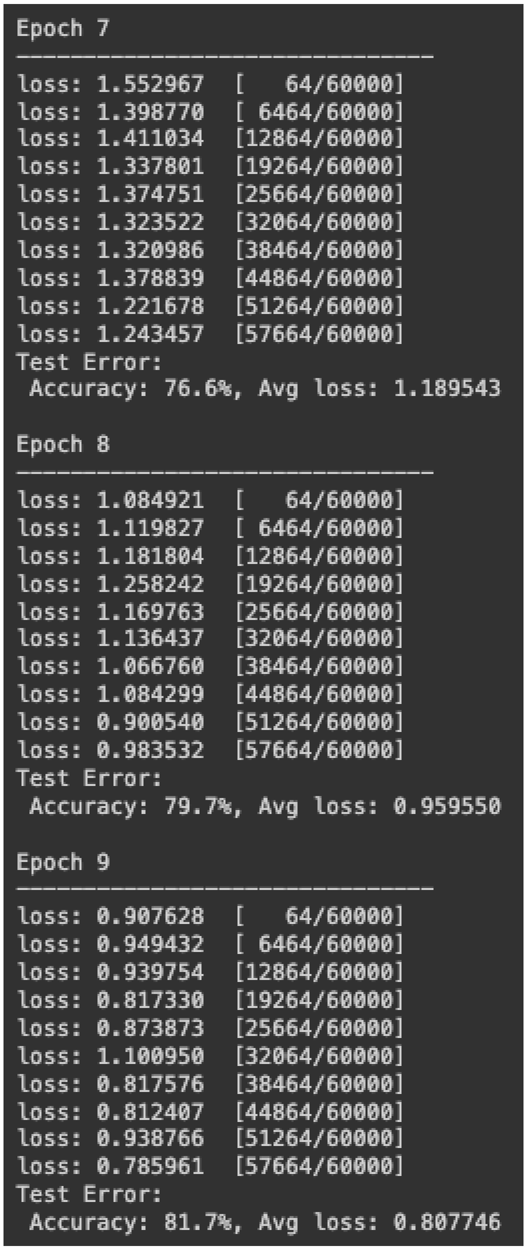
```
Using cpu device
NeuralNetwork(
  (fc1): Linear(in_features=784, out_features=200, bias=True)
  (fc2): Linear(in_features=200, out_features=200, bias=True)
  (fc3): Linear(in_features=200, out_features=10, bias=True)
)
```



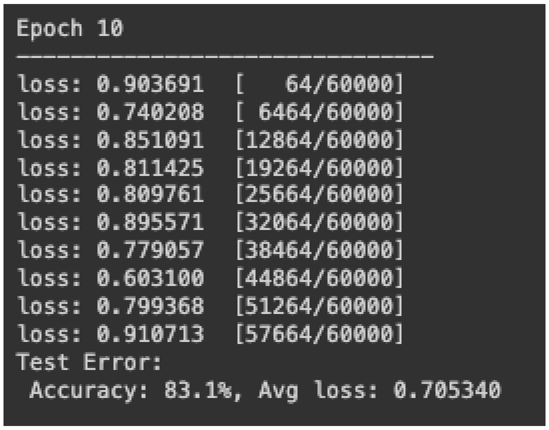
```
Epoch 1
-------------------------------
loss: 2.317232  [   64/60000]
loss: 2.295797  [ 6464/60000]
loss: 2.292961  [12864/60000]
loss: 2.300758  [19264/60000]
loss: 2.283749  [25664/60000]
loss: 2.296002  [32064/60000]
loss: 2.274789  [38464/60000]
loss: 2.277119  [44864/60000]
loss: 2.272804  [51264/60000]
loss: 2.266154  [57664/60000]
Test Error:
 Accuracy: 43.7%, Avg loss: 2.269623

Epoch 2
-------------------------------
loss: 2.263774  [   64/60000]
loss: 2.267873  [ 6464/60000]
loss: 2.256180  [12864/60000]
loss: 2.260281  [19264/60000]
loss: 2.258021  [25664/60000]
loss: 2.262865  [32064/60000]
loss: 2.255936  [38464/60000]
loss: 2.252871  [44864/60000]
loss: 2.236404  [51264/60000]
loss: 2.230034  [57664/60000]
Test Error:
 Accuracy: 59.7%, Avg loss: 2.226548

Epoch 3
-------------------------------
loss: 2.222221  [   64/60000]
loss: 2.230322  [ 6464/60000]
loss: 2.216831  [12864/60000]
loss: 2.211512  [19264/60000]
loss: 2.207336  [25664/60000]
loss: 2.187446  [32064/60000]
loss: 2.185654  [38464/60000]
loss: 2.163105  [44864/60000]
loss: 2.167564  [51264/60000]
loss: 2.178476  [57664/60000]
Test Error:
 Accuracy: 63.2%, Avg loss: 2.153835
```



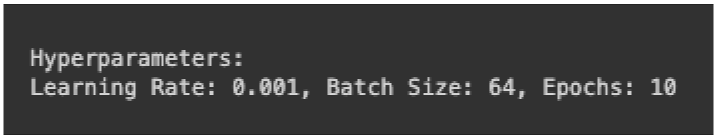
```
Epoch 4
-------------------------------
loss: 2.146271  [   64/60000]
loss: 2.153185  [ 6464/60000]
loss: 2.146166  [12864/60000]
loss: 2.097913  [19264/60000]
loss: 2.122499  [25664/60000]
loss: 2.101561  [32064/60000]
loss: 2.067542  [38464/60000]
loss: 2.042393  [44864/60000]
loss: 2.051870  [51264/60000]
loss: 2.043984  [57664/60000]
Test Error:
 Accuracy: 64.5%, Avg loss: 2.022980

Epoch 5
-------------------------------
loss: 1.998259  [   64/60000]
loss: 2.049722  [ 6464/60000]
loss: 2.032401  [12864/60000]
loss: 1.978890  [19264/60000]
loss: 1.938801  [25664/60000]
loss: 1.916338  [32064/60000]
loss: 1.951066  [38464/60000]
loss: 1.862448  [44864/60000]
loss: 1.840247  [51264/60000]
loss: 1.793576  [57664/60000]
Test Error:
 Accuracy: 66.9%, Avg loss: 1.803158

Epoch 6
-------------------------------
loss: 1.818758  [   64/60000]
loss: 1.826935  [ 6464/60000]
loss: 1.838849  [12864/60000]
loss: 1.749794  [19264/60000]
loss: 1.651051  [25664/60000]
loss: 1.706330  [32064/60000]
loss: 1.711715  [38464/60000]
loss: 1.646418  [44864/60000]
loss: 1.590284  [51264/60000]
loss: 1.495407  [57664/60000]
Test Error:
 Accuracy: 71.8%, Avg loss: 1.498418
```



```
Epoch 7
-------------------------------
loss: 1.552967  [   64/60000]
loss: 1.398770  [ 6464/60000]
loss: 1.411034  [12864/60000]
loss: 1.337801  [19264/60000]
loss: 1.374751  [25664/60000]
loss: 1.323522  [32064/60000]
loss: 1.320986  [38464/60000]
loss: 1.378839  [44864/60000]
loss: 1.221678  [51264/60000]
loss: 1.243457  [57664/60000]
Test Error:
 Accuracy: 76.6%, Avg loss: 1.189543

Epoch 8
-------------------------------
loss: 1.084921  [   64/60000]
loss: 1.119827  [ 6464/60000]
loss: 1.181804  [12864/60000]
loss: 1.258242  [19264/60000]
loss: 1.169763  [25664/60000]
loss: 1.136437  [32064/60000]
loss: 1.066760  [38464/60000]
loss: 1.084299  [44864/60000]
loss: 0.900540  [51264/60000]
loss: 0.983532  [57664/60000]
Test Error:
 Accuracy: 79.7%, Avg loss: 0.959550

Epoch 9
-------------------------------
loss: 0.907628  [   64/60000]
loss: 0.949432  [ 6464/60000]
loss: 0.939754  [12864/60000]
loss: 0.817330  [19264/60000]
loss: 0.873873  [25664/60000]
loss: 1.100950  [32064/60000]
loss: 0.817576  [38464/60000]
loss: 0.812407  [44864/60000]
loss: 0.938766  [51264/60000]
loss: 0.785961  [57664/60000]
Test Error:
 Accuracy: 81.7%, Avg loss: 0.807746
```

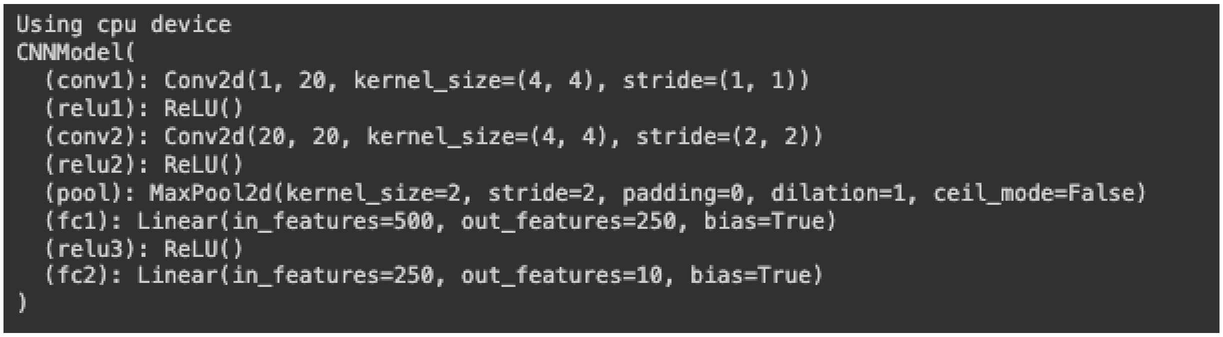


```
Epoch 10
-------------------------------
loss: 0.903691  [   64/60000]
loss: 0.740208  [ 6464/60000]
loss: 0.851091  [12864/60000]
loss: 0.811425  [19264/60000]
loss: 0.809761  [25664/60000]
loss: 0.895571  [32064/60000]
loss: 0.779057  [38464/60000]
loss: 0.603100  [44864/60000]
loss: 0.799368  [51264/60000]
loss: 0.910713  [57664/60000]
Test Error:
 Accuracy: 83.1%, Avg loss: 0.705340
```

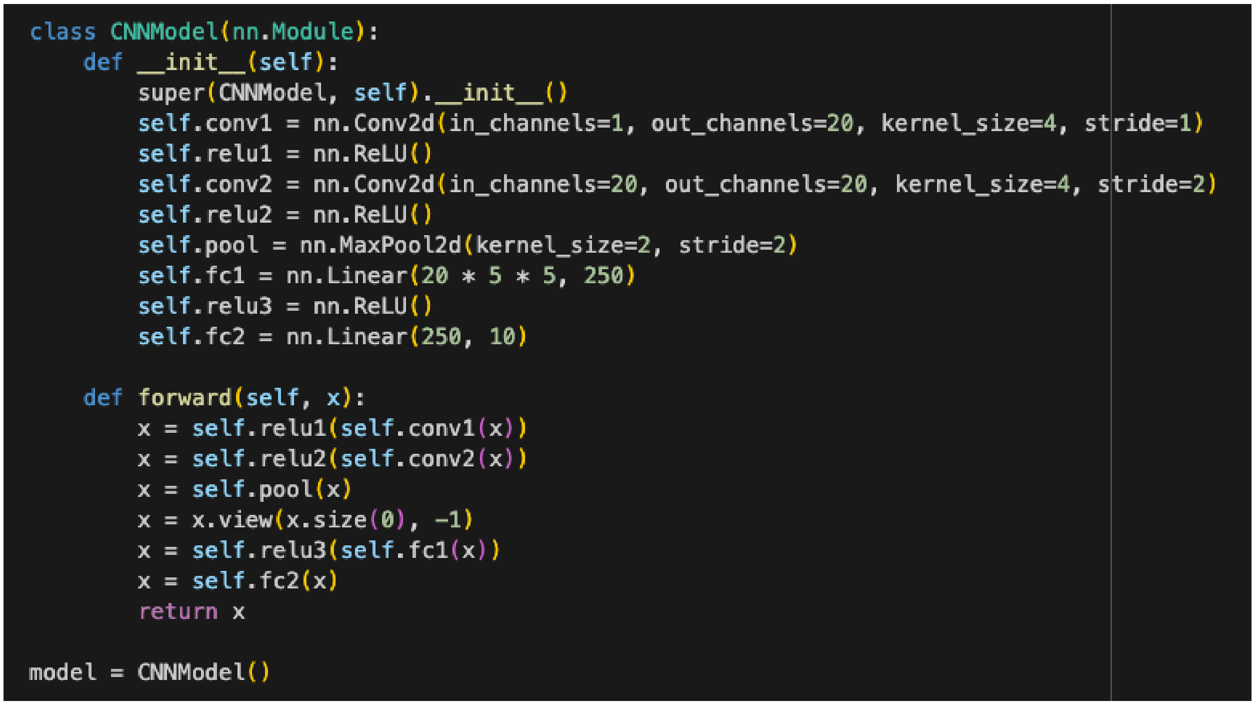


```
 Hyperparameters:
 Learning Rate: 0.001, Batch Size: 64, Epochs: 10
```

3. a)

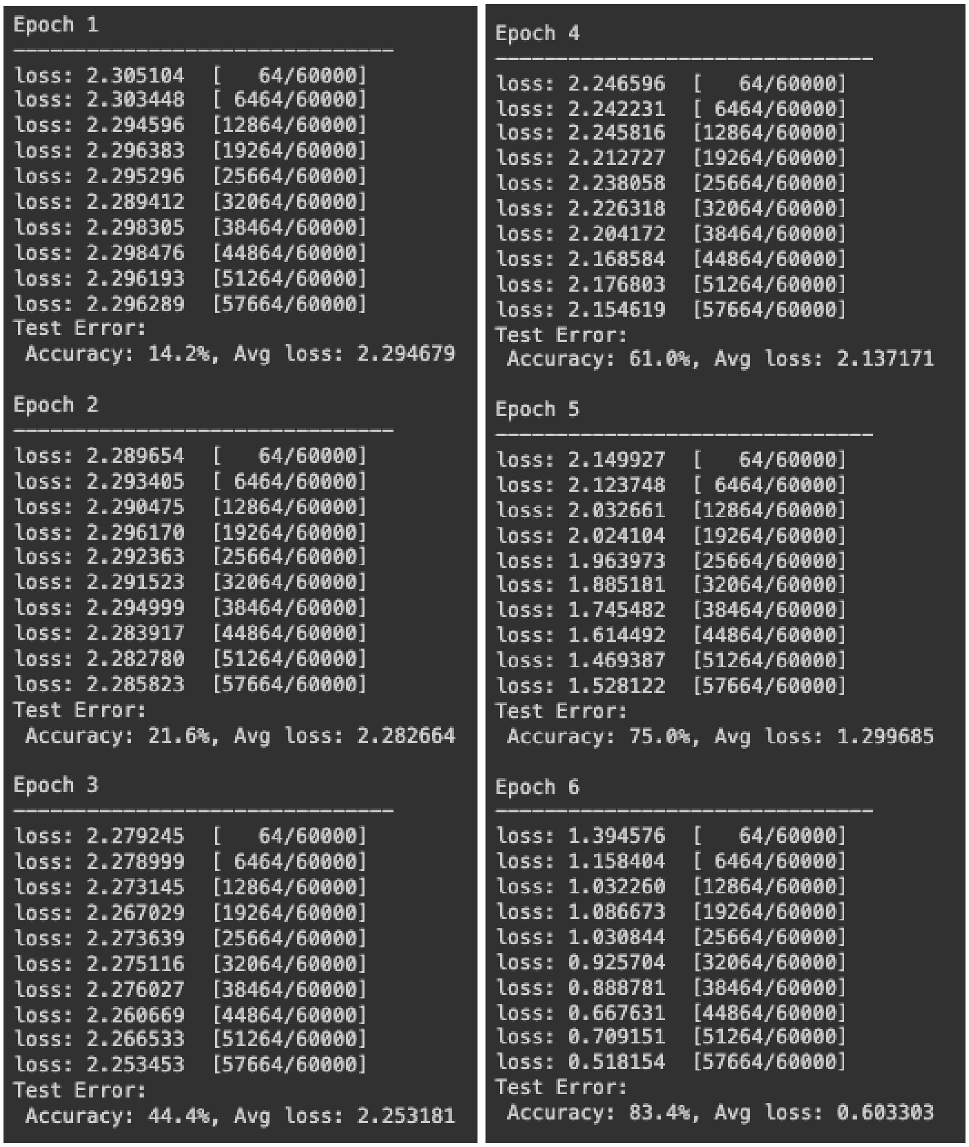


```
Using cpu device
CNNModel(
  (conv1): Conv2d(1, 20, kernel_size=(4, 4), stride=(1, 1))
  (relu1): ReLU()
  (conv2): Conv2d(20, 20, kernel_size=(4, 4), stride=(2, 2))
  (relu2): ReLU()
  (pool): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (fc1): Linear(in_features=500, out_features=250, bias=True)
  (relu3): ReLU()
  (fc2): Linear(in_features=250, out_features=10, bias=True)
)
```



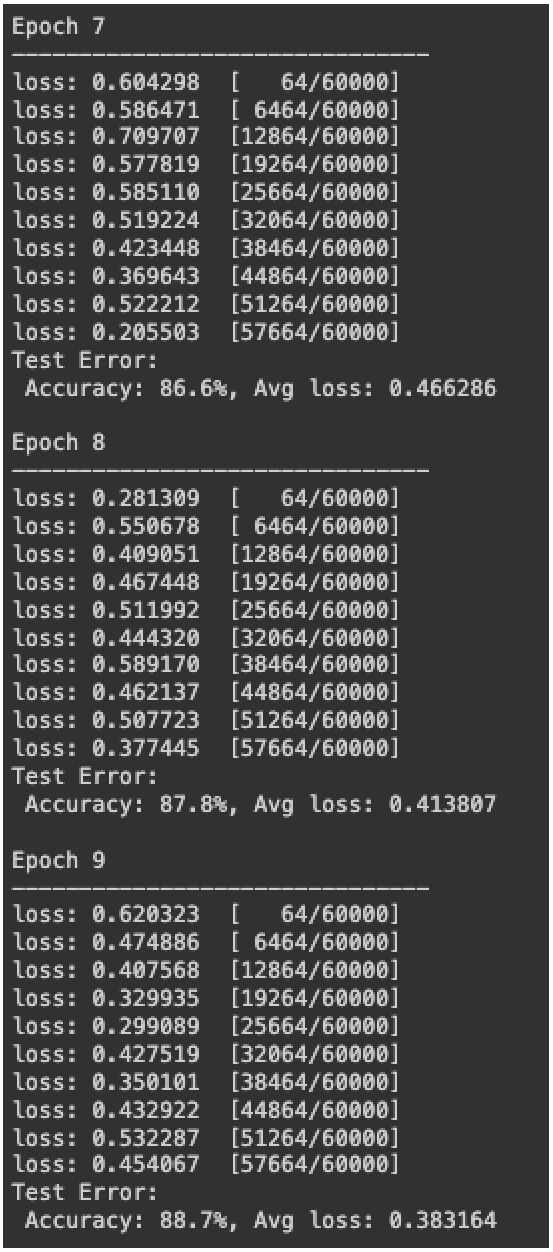
```
class CNNModel(nn.Module):
    def __init__(self):
        super(CNNModel, self).__init__()
        self.conv1 = nn.Conv2d(in_channels=1, out_channels=20, kernel_size=4, stride=1)
        self.relu1 = nn.ReLU()
        self.conv2 = nn.Conv2d(in_channels=20, out_channels=20, kernel_size=4, stride=2)
        self.relu2 = nn.ReLU()
        self.pool = nn.MaxPool2d(kernel_size=2, stride=2)
        self.fc1 = nn.Linear(20 * 5 * 5, 250)
        self.relu3 = nn.ReLU()
        self.fc2 = nn.Linear(250, 10)

    def forward(self, x):
        x = self.relu1(self.conv1(x))
        x = self.relu2(self.conv2(x))
        x = self.pool(x)
        x = x.view(x.size(0), -1)
        x = self.relu3(self.fc1(x))
        x = self.fc2(x)
        return x

model = CNNModel()
```

b)

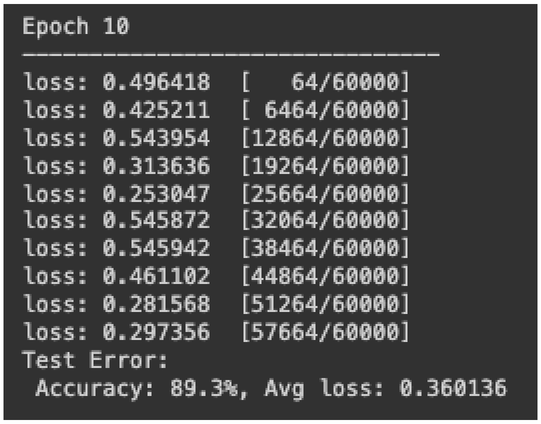


```
Epoch 1
-------------------------------
loss: 2.305104  [   64/60000]
loss: 2.303448  [ 6464/60000]
loss: 2.294596  [12864/60000]
loss: 2.296383  [19264/60000]
loss: 2.295296  [25664/60000]
loss: 2.289412  [32064/60000]
loss: 2.298305  [38464/60000]
loss: 2.298476  [44864/60000]
loss: 2.296193  [51264/60000]
loss: 2.296289  [57664/60000]
Test Error:
 Accuracy: 14.2%, Avg loss: 2.294679

Epoch 2
-------------------------------
loss: 2.289654  [   64/60000]
loss: 2.293405  [ 6464/60000]
loss: 2.290475  [12864/60000]
loss: 2.296170  [19264/60000]
loss: 2.292363  [25664/60000]
loss: 2.291523  [32064/60000]
loss: 2.294999  [38464/60000]
loss: 2.283917  [44864/60000]
loss: 2.282780  [51264/60000]
loss: 2.285823  [57664/60000]
Test Error:
 Accuracy: 21.6%, Avg loss: 2.282664

Epoch 3
-------------------------------
loss: 2.279245  [   64/60000]
loss: 2.278999  [ 6464/60000]
loss: 2.273145  [12864/60000]
loss: 2.267029  [19264/60000]
loss: 2.273639  [25664/60000]
loss: 2.275116  [32064/60000]
loss: 2.276027  [38464/60000]
loss: 2.260669  [44864/60000]
loss: 2.266533  [51264/60000]
loss: 2.253453  [57664/60000]
Test Error:
 Accuracy: 44.4%, Avg loss: 2.253181
```

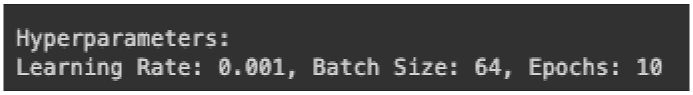
```
Epoch 4
-------------------------------
loss: 2.246596  [   64/60000]
loss: 2.242231  [ 6464/60000]
loss: 2.245816  [12864/60000]
loss: 2.212727  [19264/60000]
loss: 2.238058  [25664/60000]
loss: 2.226318  [32064/60000]
loss: 2.204172  [38464/60000]
loss: 2.168584  [44864/60000]
loss: 2.176803  [51264/60000]
loss: 2.154619  [57664/60000]
Test Error:
 Accuracy: 61.0%, Avg loss: 2.137171

Epoch 5
-------------------------------
loss: 2.149927  [   64/60000]
loss: 2.123748  [ 6464/60000]
loss: 2.032661  [12864/60000]
loss: 2.024104  [19264/60000]
loss: 1.963973  [25664/60000]
loss: 1.885181  [32064/60000]
loss: 1.745482  [38464/60000]
loss: 1.614492  [44864/60000]
loss: 1.469387  [51264/60000]
loss: 1.528122  [57664/60000]
Test Error:
 Accuracy: 75.0%, Avg loss: 1.299685

Epoch 6
-------------------------------
loss: 1.394576  [   64/60000]
loss: 1.158404  [ 6464/60000]
loss: 1.032260  [12864/60000]
loss: 1.086673  [19264/60000]
loss: 1.030844  [25664/60000]
loss: 0.925704  [32064/60000]
loss: 0.888781  [38464/60000]
loss: 0.667631  [44864/60000]
loss: 0.709151  [51264/60000]
loss: 0.518154  [57664/60000]
Test Error:
 Accuracy: 83.4%, Avg loss: 0.603303
```



```
Epoch 7
-------------------------------
loss: 0.604298  [   64/60000]
loss: 0.586471  [ 6464/60000]
loss: 0.709707  [12864/60000]
loss: 0.577819  [19264/60000]
loss: 0.585110  [25664/60000]
loss: 0.519224  [32064/60000]
loss: 0.423448  [38464/60000]
loss: 0.369643  [44864/60000]
loss: 0.522212  [51264/60000]
loss: 0.205503  [57664/60000]
Test Error:
 Accuracy: 86.6%, Avg loss: 0.466286

Epoch 8
-------------------------------
loss: 0.281309  [   64/60000]
loss: 0.550678  [ 6464/60000]
loss: 0.409051  [12864/60000]
loss: 0.467448  [19264/60000]
loss: 0.511992  [25664/60000]
loss: 0.444320  [32064/60000]
loss: 0.589170  [38464/60000]
loss: 0.462137  [44864/60000]
loss: 0.507723  [51264/60000]
loss: 0.377445  [57664/60000]
Test Error:
 Accuracy: 87.8%, Avg loss: 0.413807

Epoch 9
-------------------------------
loss: 0.620323  [   64/60000]
loss: 0.474886  [ 6464/60000]
loss: 0.407568  [12864/60000]
loss: 0.329935  [19264/60000]
loss: 0.299089  [25664/60000]
loss: 0.427519  [32064/60000]
loss: 0.350101  [38464/60000]
loss: 0.432922  [44864/60000]
loss: 0.532287  [51264/60000]
loss: 0.454067  [57664/60000]
Test Error:
 Accuracy: 88.7%, Avg loss: 0.383164
```

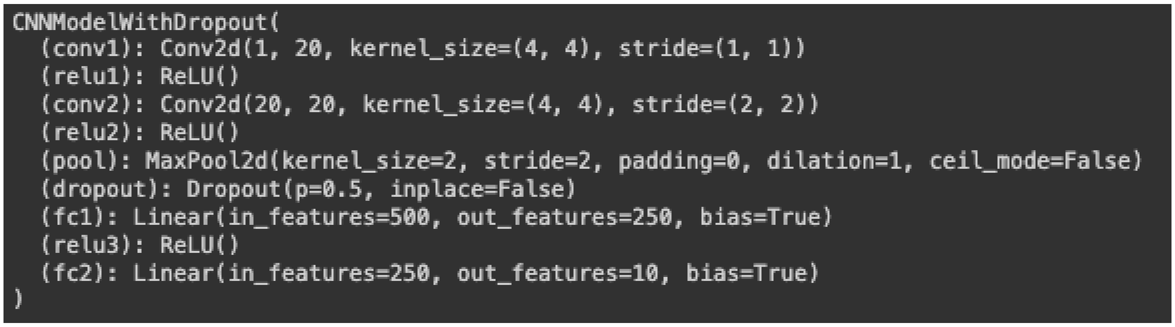


```
Epoch 10
-------------------------------
loss: 0.496418  [   64/60000]
loss: 0.425211  [ 6464/60000]
loss: 0.543954  [12864/60000]
loss: 0.313636  [19264/60000]
loss: 0.253047  [25664/60000]
loss: 0.545872  [32064/60000]
loss: 0.545942  [38464/60000]
loss: 0.461102  [44864/60000]
loss: 0.281568  [51264/60000]
loss: 0.297356  [57664/60000]
Test Error:
 Accuracy: 89.3%, Avg loss: 0.360136
```

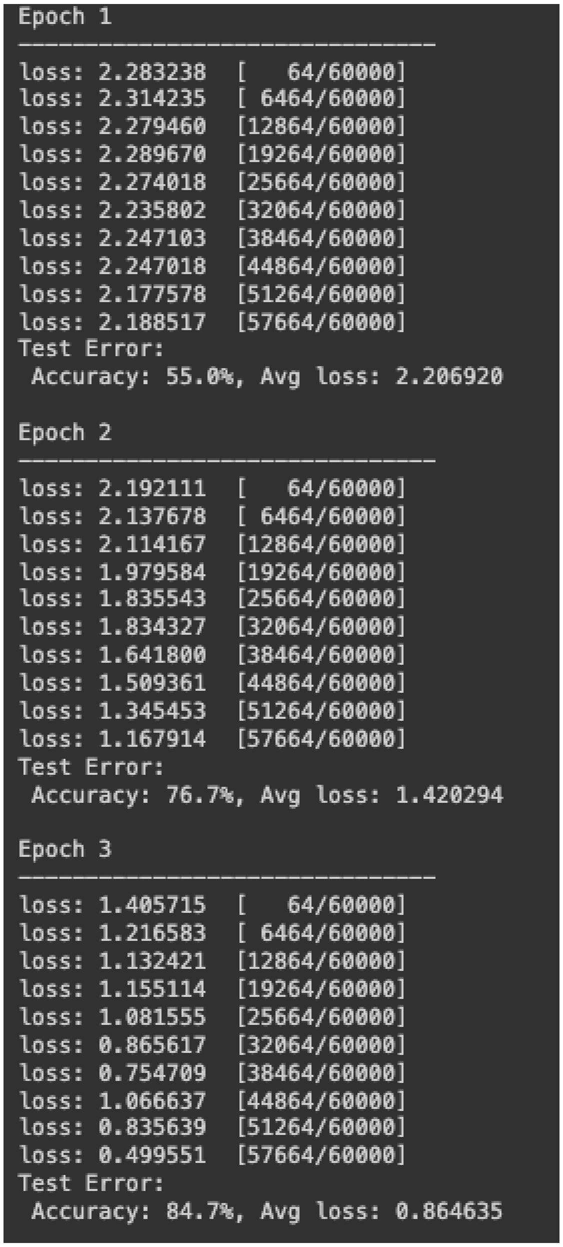


```
Hyperparameters:
Learning Rate: 0.001, Batch Size: 64, Epochs: 10
```

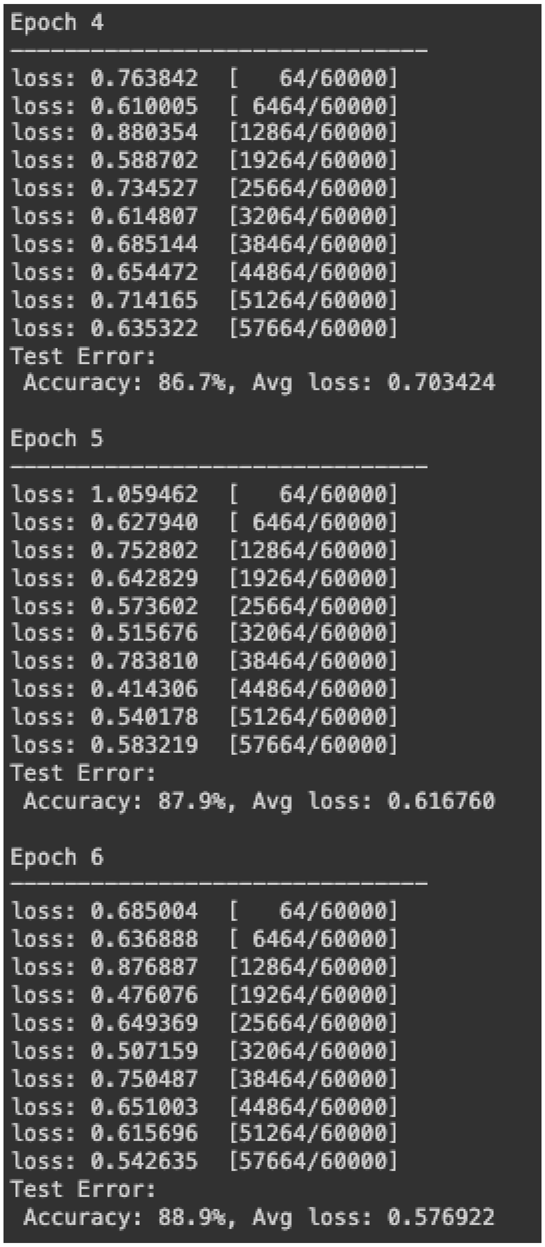
c)



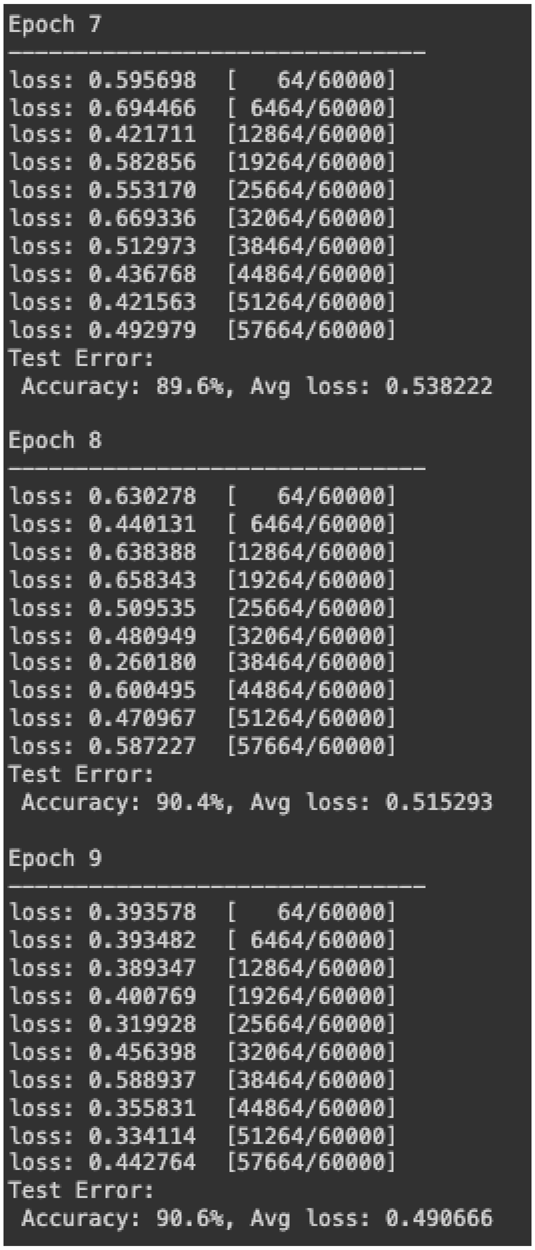
```
CNNModelWithDropout(
  (conv1): Conv2d(1, 20, kernel_size=(4, 4), stride=(1, 1))
  (relu1): ReLU()
  (conv2): Conv2d(20, 20, kernel_size=(4, 4), stride=(2, 2))
  (relu2): ReLU()
  (pool): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (dropout): Dropout(p=0.5, inplace=False)
  (fc1): Linear(in_features=500, out_features=250, bias=True)
  (relu3): ReLU()
  (fc2): Linear(in_features=250, out_features=10, bias=True)
)
```



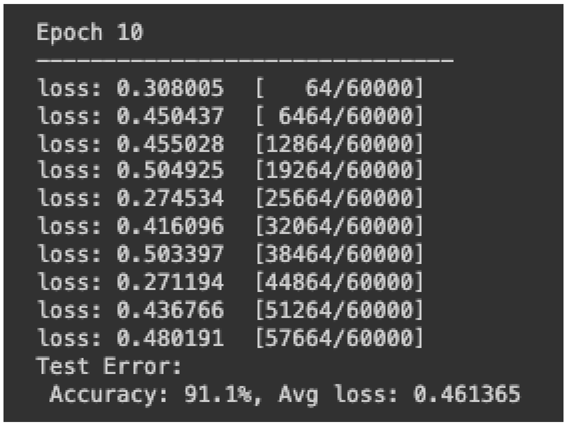
```
Epoch 1
-------------------------------
loss: 2.283238  [   64/60000]
loss: 2.314235  [ 6464/60000]
loss: 2.279460  [12864/60000]
loss: 2.289670  [19264/60000]
loss: 2.274018  [25664/60000]
loss: 2.235802  [32064/60000]
loss: 2.247103  [38464/60000]
loss: 2.247018  [44864/60000]
loss: 2.177578  [51264/60000]
loss: 2.188517  [57664/60000]
Test Error:
 Accuracy: 55.0%, Avg loss: 2.206920

Epoch 2
-------------------------------
loss: 2.192111  [   64/60000]
loss: 2.137678  [ 6464/60000]
loss: 2.114167  [12864/60000]
loss: 1.979584  [19264/60000]
loss: 1.835543  [25664/60000]
loss: 1.834327  [32064/60000]
loss: 1.641800  [38464/60000]
loss: 1.509361  [44864/60000]
loss: 1.345453  [51264/60000]
loss: 1.167914  [57664/60000]
Test Error:
 Accuracy: 76.7%, Avg loss: 1.420294

Epoch 3
-------------------------------
loss: 1.405715  [   64/60000]
loss: 1.216583  [ 6464/60000]
loss: 1.132421  [12864/60000]
loss: 1.155114  [19264/60000]
loss: 1.081555  [25664/60000]
loss: 0.865617  [32064/60000]
loss: 0.754709  [38464/60000]
loss: 1.066637  [44864/60000]
loss: 0.835639  [51264/60000]
loss: 0.499551  [57664/60000]
Test Error:
 Accuracy: 84.7%, Avg loss: 0.864635
```



```
Epoch 4
-------------------------------
loss: 0.763842  [   64/60000]
loss: 0.610005  [ 6464/60000]
loss: 0.880354  [12864/60000]
loss: 0.588702  [19264/60000]
loss: 0.734527  [25664/60000]
loss: 0.614807  [32064/60000]
loss: 0.685144  [38464/60000]
loss: 0.654472  [44864/60000]
loss: 0.714165  [51264/60000]
loss: 0.635322  [57664/60000]
Test Error:
 Accuracy: 86.7%, Avg loss: 0.703424

Epoch 5
-------------------------------
loss: 1.059462  [   64/60000]
loss: 0.627940  [ 6464/60000]
loss: 0.752802  [12864/60000]
loss: 0.642829  [19264/60000]
loss: 0.573602  [25664/60000]
loss: 0.515676  [32064/60000]
loss: 0.783810  [38464/60000]
loss: 0.414306  [44864/60000]
loss: 0.540178  [51264/60000]
loss: 0.583219  [57664/60000]
Test Error:
 Accuracy: 87.9%, Avg loss: 0.616760

Epoch 6
-------------------------------
loss: 0.685004  [   64/60000]
loss: 0.636888  [ 6464/60000]
loss: 0.876887  [12864/60000]
loss: 0.476076  [19264/60000]
loss: 0.649369  [25664/60000]
loss: 0.507159  [32064/60000]
loss: 0.750487  [38464/60000]
loss: 0.651003  [44864/60000]
loss: 0.615696  [51264/60000]
loss: 0.542635  [57664/60000]
Test Error:
 Accuracy: 88.9%, Avg loss: 0.576922
```



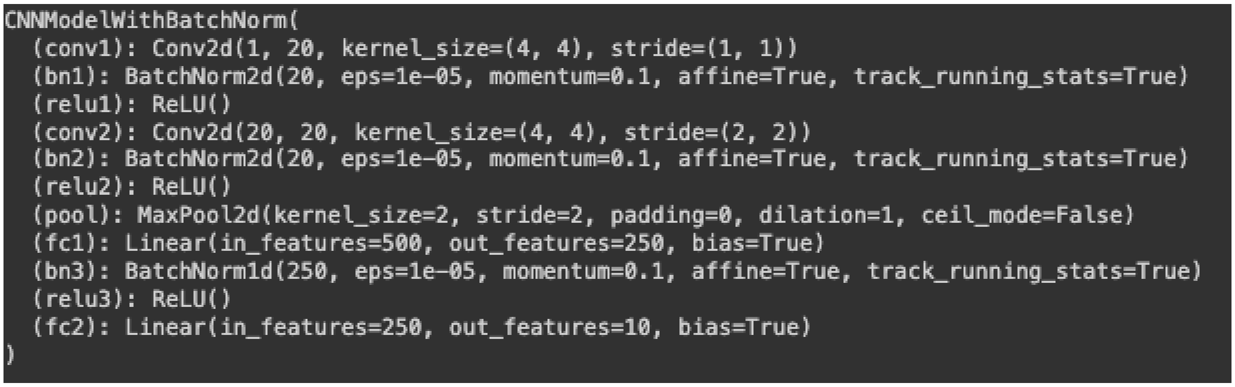
```
Epoch 7
-------------------------------
loss: 0.595698  [   64/60000]
loss: 0.694466  [ 6464/60000]
loss: 0.421711  [12864/60000]
loss: 0.582856  [19264/60000]
loss: 0.553170  [25664/60000]
loss: 0.669336  [32064/60000]
loss: 0.512973  [38464/60000]
loss: 0.436768  [44864/60000]
loss: 0.421563  [51264/60000]
loss: 0.492979  [57664/60000]
Test Error:
 Accuracy: 89.6%, Avg loss: 0.538222

Epoch 8
-------------------------------
loss: 0.630278  [   64/60000]
loss: 0.440131  [ 6464/60000]
loss: 0.638388  [12864/60000]
loss: 0.658343  [19264/60000]
loss: 0.509535  [25664/60000]
loss: 0.480949  [32064/60000]
loss: 0.260180  [38464/60000]
loss: 0.600495  [44864/60000]
loss: 0.470967  [51264/60000]
loss: 0.587227  [57664/60000]
Test Error:
 Accuracy: 90.4%, Avg loss: 0.515293

Epoch 9
-------------------------------
loss: 0.393578  [   64/60000]
loss: 0.393482  [ 6464/60000]
loss: 0.389347  [12864/60000]
loss: 0.400769  [19264/60000]
loss: 0.319928  [25664/60000]
loss: 0.456398  [32064/60000]
loss: 0.588937  [38464/60000]
loss: 0.355831  [44864/60000]
loss: 0.334114  [51264/60000]
loss: 0.442764  [57664/60000]
Test Error:
 Accuracy: 90.6%, Avg loss: 0.490666
```



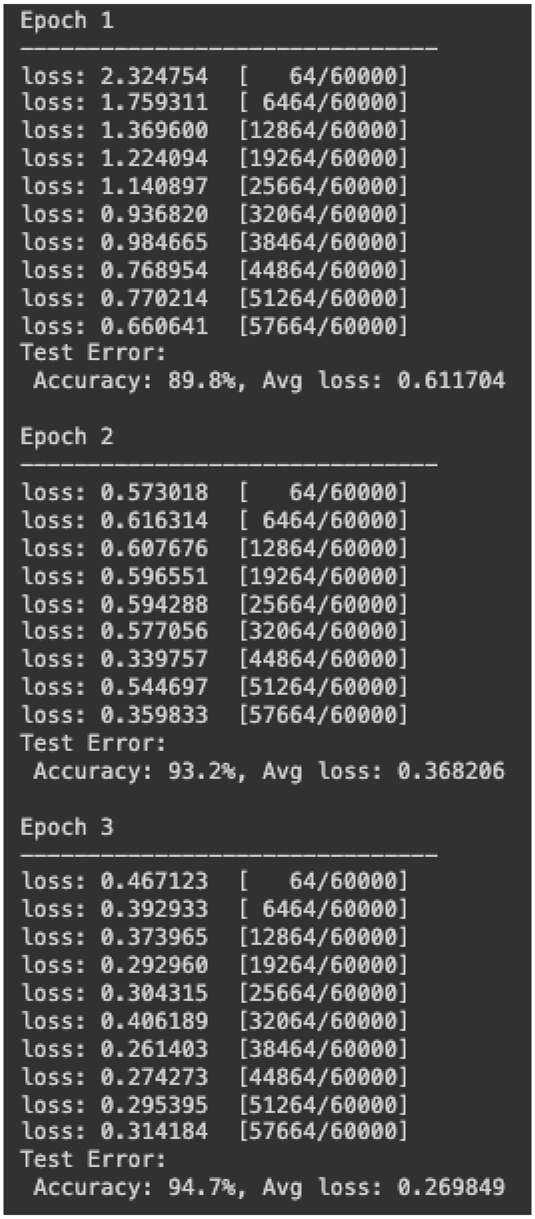
```
Epoch 10
-------------------------------
loss: 0.308005  [   64/60000]
loss: 0.450437  [ 6464/60000]
loss: 0.455028  [12864/60000]
loss: 0.504925  [19264/60000]
loss: 0.274534  [25664/60000]
loss: 0.416096  [32064/60000]
loss: 0.503397  [38464/60000]
loss: 0.271194  [44864/60000]
loss: 0.436766  [51264/60000]
loss: 0.480191  [57664/60000]
Test Error:
 Accuracy: 91.1%, Avg loss: 0.461365
```

```
Hyperparameters:
Learning Rate: 0.001, Batch Size: 64, Epochs: 10, Weight Decay: 0.0005
```

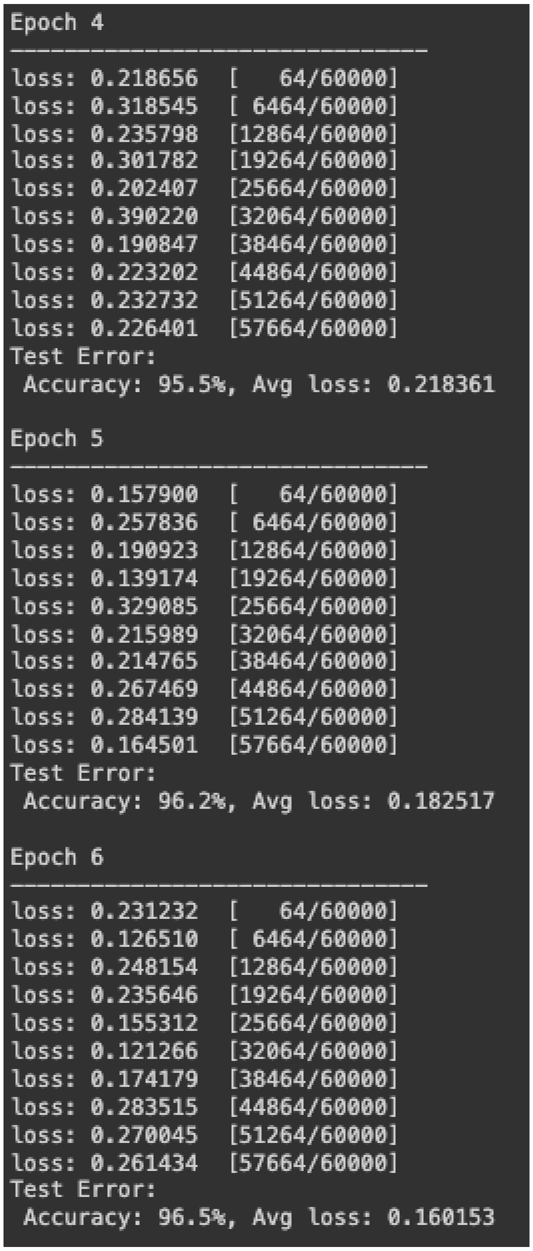
d)



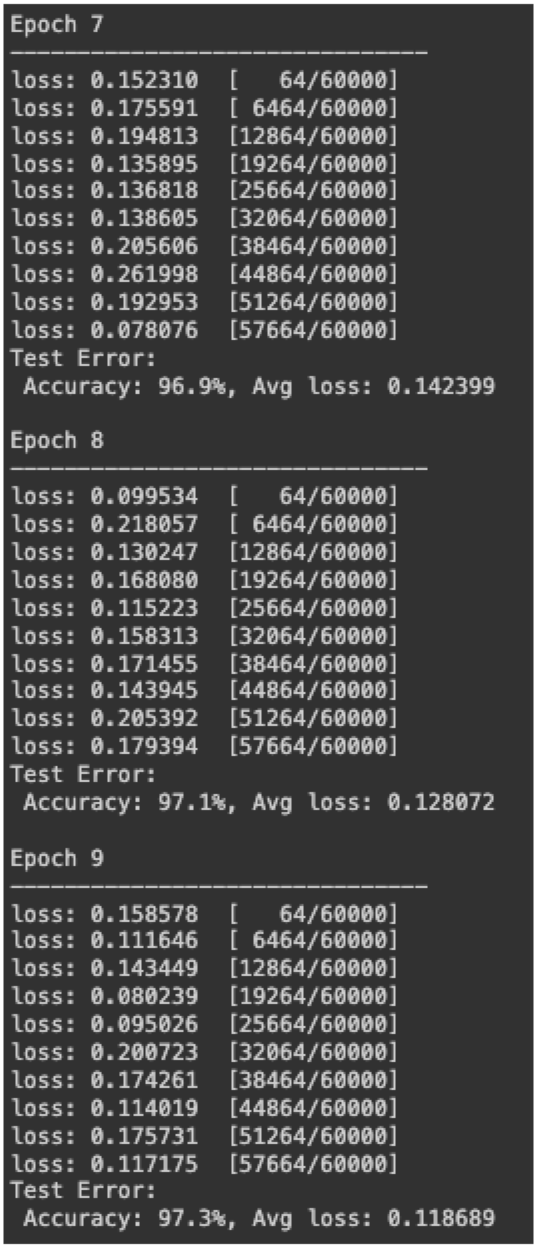
```
CNNModelWithBatchNorm(
  (conv1): Conv2d(1, 20, kernel_size=(4, 4), stride=(1, 1))
  (bn1): BatchNorm2d(20, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (relu1): ReLU()
  (conv2): Conv2d(20, 20, kernel_size=(4, 4), stride=(2, 2))
  (bn2): BatchNorm2d(20, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (relu2): ReLU()
  (pool): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (fc1): Linear(in_features=500, out_features=250, bias=True)
  (bn3): BatchNorm1d(250, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (relu3): ReLU()
  (fc2): Linear(in_features=250, out_features=10, bias=True)
)
```



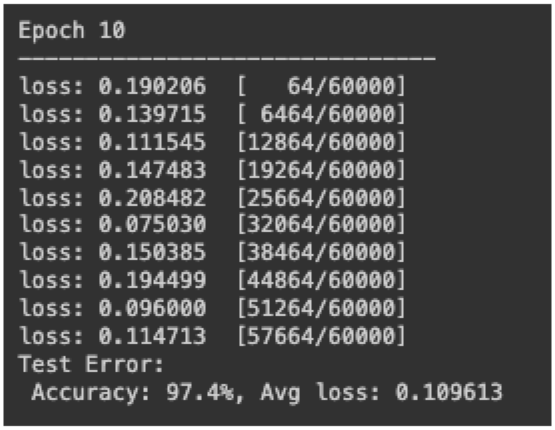
```
Epoch 1
-------------------------------
loss: 2.324754  [   64/60000]
loss: 1.759311  [ 6464/60000]
loss: 1.369600  [12864/60000]
loss: 1.224094  [19264/60000]
loss: 1.140897  [25664/60000]
loss: 0.936820  [32064/60000]
loss: 0.984665  [38464/60000]
loss: 0.768954  [44864/60000]
loss: 0.770214  [51264/60000]
loss: 0.660641  [57664/60000]
Test Error:
 Accuracy: 89.8%, Avg loss: 0.611704

Epoch 2
-------------------------------
loss: 0.573018  [   64/60000]
loss: 0.616314  [ 6464/60000]
loss: 0.607676  [12864/60000]
loss: 0.596551  [19264/60000]
loss: 0.594288  [25664/60000]
loss: 0.577056  [32064/60000]
loss: 0.339757  [44864/60000]
loss: 0.544697  [51264/60000]
loss: 0.359833  [57664/60000]
Test Error:
 Accuracy: 93.2%, Avg loss: 0.368206

Epoch 3
-------------------------------
loss: 0.467123  [   64/60000]
loss: 0.392933  [ 6464/60000]
loss: 0.373965  [12864/60000]
loss: 0.292960  [19264/60000]
loss: 0.304315  [25664/60000]
loss: 0.406189  [32064/60000]
loss: 0.261403  [38464/60000]
loss: 0.274273  [44864/60000]
loss: 0.295395  [51264/60000]
loss: 0.314184  [57664/60000]
Test Error:
 Accuracy: 94.7%, Avg loss: 0.269849
```



```
Epoch 4
-------------------------------
loss: 0.218656  [   64/60000]
loss: 0.318545  [ 6464/60000]
loss: 0.235798  [12864/60000]
loss: 0.301782  [19264/60000]
loss: 0.202407  [25664/60000]
loss: 0.390220  [32064/60000]
loss: 0.190847  [38464/60000]
loss: 0.223202  [44864/60000]
loss: 0.232732  [51264/60000]
loss: 0.226401  [57664/60000]
Test Error:
 Accuracy: 95.5%, Avg loss: 0.218361

Epoch 5
-------------------------------
loss: 0.157900  [   64/60000]
loss: 0.257836  [ 6464/60000]
loss: 0.190923  [12864/60000]
loss: 0.139174  [19264/60000]
loss: 0.329085  [25664/60000]
loss: 0.215989  [32064/60000]
loss: 0.214765  [38464/60000]
loss: 0.267469  [44864/60000]
loss: 0.284139  [51264/60000]
loss: 0.164501  [57664/60000]
Test Error:
 Accuracy: 96.2%, Avg loss: 0.182517

Epoch 6
-------------------------------
loss: 0.231232  [   64/60000]
loss: 0.126510  [ 6464/60000]
loss: 0.248154  [12864/60000]
loss: 0.235646  [19264/60000]
loss: 0.155312  [25664/60000]
loss: 0.121266  [32064/60000]
loss: 0.174179  [38464/60000]
loss: 0.283515  [44864/60000]
loss: 0.270045  [51264/60000]
loss: 0.261434  [57664/60000]
Test Error:
 Accuracy: 96.5%, Avg loss: 0.160153
```



```
Epoch 7
-------------------------------
loss: 0.152310  [   64/60000]
loss: 0.175591  [ 6464/60000]
loss: 0.194813  [12864/60000]
loss: 0.135895  [19264/60000]
loss: 0.136818  [25664/60000]
loss: 0.138605  [32064/60000]
loss: 0.205606  [38464/60000]
loss: 0.261998  [44864/60000]
loss: 0.192953  [51264/60000]
loss: 0.078076  [57664/60000]
Test Error:
 Accuracy: 96.9%, Avg loss: 0.142399

Epoch 8
-------------------------------
loss: 0.099534  [   64/60000]
loss: 0.218057  [ 6464/60000]
loss: 0.130247  [12864/60000]
loss: 0.168080  [19264/60000]
loss: 0.115223  [25664/60000]
loss: 0.158313  [32064/60000]
loss: 0.171455  [38464/60000]
loss: 0.143945  [44864/60000]
loss: 0.205392  [51264/60000]
loss: 0.179394  [57664/60000]
Test Error:
 Accuracy: 97.1%, Avg loss: 0.128072

Epoch 9
-------------------------------
loss: 0.158578  [   64/60000]
loss: 0.111646  [ 6464/60000]
loss: 0.143449  [12864/60000]
loss: 0.080239  [19264/60000]
loss: 0.095026  [25664/60000]
loss: 0.200723  [32064/60000]
loss: 0.174261  [38464/60000]
loss: 0.114019  [44864/60000]
loss: 0.175731  [51264/60000]
loss: 0.117175  [57664/60000]
Test Error:
 Accuracy: 97.3%, Avg loss: 0.118689
```



```
Epoch 10
-------------------------------
loss: 0.190206  [   64/60000]
loss: 0.139715  [ 6464/60000]
loss: 0.111545  [12864/60000]
loss: 0.147483  [19264/60000]
loss: 0.208482  [25664/60000]
loss: 0.075030  [32064/60000]
loss: 0.150385  [38464/60000]
loss: 0.194499  [44864/60000]
loss: 0.096000  [51264/60000]
loss: 0.114713  [57664/60000]
Test Error:
 Accuracy: 97.4%, Avg loss: 0.109613
```

```
Hyperparameters:
Learning Rate: 0.001, Batch Size: 64, Epochs: 10, Weight Decay: 0.0005
```