```
In [1]: import numpy as np
        def relu(x):
            return np.maximum(0, x)
        def relu deriv(x):
            return (x > 0).astype(float)
        def sigmoid(x):
           return 1 / (1 + np.exp(-x))
        def sigmoid deriv(x):
           s = sigmoid(x)
            return s * (1 - s)
        def forward(x, parameters, activation='relu'):
           caches = []
           A = x
           L = len(parameters) // 2
            for l in range(1, L + 1):
                W = parameters["W" + str(l)]
                b = parameters["b" + str(l)]
                Z = np.dot(W, A) + b
                caches.append((A, Z))
                if activation == 'relu':
                   A = relu(Z)
                elif activation == 'sigmoid':
                    A = sigmoid(Z)
                    raise ValueError("Unsupported activation function")
            AL = A
            return AL, caches
        def backward(dAL, caches, parameters, activation='relu'):
            grads = {}
            L = len(caches)
            A prev, Z = caches[-1]
            if activation == 'relu':
                dZ = dAL * relu deriv(Z)
            elif activation == 'sigmoid':
                dZ = dAL * sigmoid deriv(Z)
                raise ValueError("Unsupported activation function")
            grads["dW" + str(L)] = np.einsum('ik, jk->ij', dZ, A prev)
            grads["db" + str(L)] = np.sum(dZ, axis=1, keepdims=True)
            dA prev = np.dot(parameters["W" + str(L)].T, dZ)
            for 1 in range(L - 1, 0, -1):
                A prev, Z = caches[1 - 1]
                if activation == 'relu':
                    dZ = dA prev * relu deriv(Z)
                elif activation == 'sigmoid':
                    dZ = dA_prev * sigmoid deriv(Z)
                grads["dW" + str(l)] = np.einsum('ik,jk->ij', dZ, A prev)
                grads["db" + str(l)] = np.sum(dZ, axis=1, keepdims=True)
                if 1 > 1:
                    dA prev = np.dot(parameters["W" + str(1)].T, dZ)
```

return grads

```
A): Warm up:
                                                        5 0 0 0
             0
             0
                                  0
                                  0
             0
                                                                    0
            O
                                                                    0
                 L<sub>2</sub> L<sub>3</sub> L<sub>4</sub>
           Li
                                                                    Te
 0: from raw data to 21:
                                                                    0---0
       L(= Wo-1 Lo + 80-)1
                                                                    200
                                                                    0
L_2 = W_{52} \cdot L_1 + S_{102}
Dim: 4x6 4
                                                                    03
(2) L3 = W2-13 L2 + 82-3
Dim: 344 3
(2) L4 = W3-4 L3 + 83-4
        243
```

```
In [2]: | # --- Example Usage ---
        # Define network architecture: input 6, then layers with 6, 4, 3, and 2 neurons.
        layer dims = [6, 6, 4, 3, 2]
       parameters = {}
        np.random.seed(42)
        L = len(layer dims) - 1
        for l in range(1, L + 1):
            parameters["W" + str(1)] = np.random.randn(layer dims[1], layer dims[1 - 1]) * 0.01
            parameters["b" + str(l)] = np.zeros((layer dims[l], 1))
        x = np.random.randn(6, 1)
        # Forward pass
       AL, caches = forward(x, parameters, activation='sigmoid')
        print("Forward propagation output (AL):\n", AL)
        # Assume a dummy derivative from the loss (e.g., from mean squared error)
        # For our output layer with 2 neurons, dAL has shape (2, 1)
        dAL = np.random.randn(2, 1)
        # Backward pass to compute gradients
        grads = backward(dAL, caches, parameters, activation='sigmoid')
        # Print gradients for each layer
        for l in range(1, L + 1):
            print(f"\nGradients for Layer {1}:")
            print("dW" + str(l) + ":\n", grads["dW" + str(l)])
            print("db" + str(l) + ":\n", grads["db" + str(l)])
```

Forward propagation output (AL):

```
[0.50075413]]
Gradients for Layer 1:
dW1:
[ 1.26488930e-09 -2.73979226e-08 -3.02809797e-09 4.92266836e-09
  2.03722379e-08 -7.14416856e-091
 [ 4.16995430e-10 -9.03225959e-09 -9.98271557e-10 1.62285364e-09
  6.71610557e-09 -2.35521451e-09]
 [ 1.06951551e-09 -2.31660615e-08 -2.56038038e-09 4.16231695e-09
   1.72255583e-08 -6.04068600e-09]
 [ 3.36203080e-10 -7.28227044e-09 -8.04857674e-10 1.30842775e-09
  5.41486840e-09 -1.89889460e-09]
 [ 4.95193484e-10 -1.07260554e-08 -1.18547479e-09 1.92718311e-09
   7.97555915e-09 -2.79688168e-091
 [1.07490250e-11 -2.32827455e-10 -2.57327664e-11  4.18328192e-11]
  1.73123209e-10 -6.07111204e-11]]
db1:
 [[1.37846403e-08]
 [4.54437555e-09]
 [1.16554758e-08]
 [3.66390840e-09]
 [5.39657032e-09]
 [1.17141827e-10]]
Gradients for Layer 2:
dW2:
[-4.64793566e-08 -4.60100809e-08 -4.69072904e-08 -4.70964132e-08
 -4.62203650e-08 -4.58394542e-081
 [-2.48009333e-06 -2.45505324e-06 -2.50292747e-06 -2.51301888e-06]
 -2.46627379e-06 -2.44594877e-06]
 1.04210658e-06 1.03351838e-06]
 [ 2.06265479e-06  2.04182933e-06  2.08164558e-06  2.09003844e-06
  2.05116129e-06 2.03425728e-06]]
db2:
[[-9.27281473e-08]
 [-4.94788390e-06]
 [ 2.09069341e-06]
 [ 4.11507755e-06]]
Gradients for Layer 3:
dW3:
 [[-0.00011922 -0.0001192 -0.00011935 -0.00012044]
 [-0.00040692 -0.00040683 -0.00040736 -0.00041107]
[ 0.00070526  0.00070511  0.00070603  0.00071246]]
db3:
 [[-0.00023963]
 [-0.00081787]
 [ 0.00141751]]
Gradients for Layer 4:
[-0.10031159 -0.10184285 -0.10147129]
 [-0.06225442 -0.06320474 -0.06297415]]
db4:
[[-0.2021219]
[-0.12543898]
```

Task 2

[[0.49863851]

B): CNN Using a bxb hemel, 3 chan output

28 × 28 - 5 3 × 23 × 23 -> 3 × 12 × 12

CNN flatter) 64

Poedding = 2.

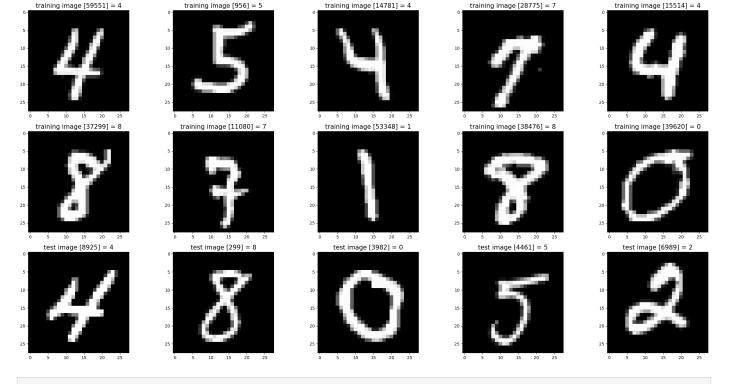
```
In [ ]:
        import numpy as np
        def relu(x):
           return np.maximum(0, x)
        def relu deriv(x):
            return (x > 0).astype(float)
        class CNN:
            def init (self, kernel size=3, num filters=8, pool size=2, num classes=10):
                Initializes a simple CNN for MNIST.
                - Convolutional layer: 3x3 filters, 'num_filters' filters.
                - Pooling layer: 2x2 max pooling.
                - Fully connected layer: flattens and outputs 'num classes' scores.
                The input image is assumed to be 28x28x1.
                self.kernel size = kernel size
                self.num filters = num filters
                self.pool size = pool size
                self.num classes = num classes
                # Initialize convolutional layer parameters.
                # Weight shape: (kernel size, kernel size, input channels, num filters)
                # For MNIST, input channels = 1.
                self.W conv = np.random.randn(kernel size, kernel size, 1, num filters) * 0.01
                self.b conv = np.zeros((1, 1, 1, num filters))
                # After convolution:
                # Output height/width = 28 - 3 + 1 = 26.
                # After 2x2 pooling with stride 2:
                # Output height/width = 26 // 2 = 13.
                self.conv out height = 26
```

```
self.conv out width = 26
    self.pool out height = self.conv out height // pool size
    self.pool out width = self.conv out width // pool size
    # Fully connected layer parameters.
    self.fc input dim = self.pool out height * self.pool out width * num filters
    self.W fc = np.random.randn(num classes, self.fc input dim) * 0.01
    self.b fc = np.zeros((num classes, 1))
def forward(self, X):
    Forward propagation for the CNN.
   Arguments:
    - X: Input image, a numpy array of shape (28, 28, 1).
   Returns:
    - out fc: Output of the fully connected layer (before any softmax),
             shape (num classes, 1).
   Also stores intermediate values needed for backpropagation.
    # Store input for backpropagation.
    self.X = X # shape: (28,28,1)
    # ---- Convolutional Layer Forward ----
    H, W, C = X.shape # (28, 28, 1)
    out conv = np.zeros((H - self.kernel size + 1, W - self.kernel size + 1, self.nu
    # Naive convolution (no padding, stride=1)
    for i in range(out conv.shape[0]):
        for j in range(out conv.shape[1]):
            # Extract the current 3x3 patch from input.
           patch = X[i:i+self.kernel size, j:j+self.kernel size, :] # shape: (3,3,
            for f in range(self.num filters):
                conv_sum = np.sum(patch * self.W_conv[:, :, :, f]) + self.b conv[0, f]
                out conv[i, j, f] = conv sum
    # Apply ReLU activation
    self.conv out = relu(out conv) # shape: (26,26,num filters)
    # ---- Pooling Layer Forward (Max Pooling) ----
    H conv, W conv, F = self.conv out.shape # (26,26,num filters)
    H pool = H conv // self.pool size # 13
    W pool = W conv // self.pool size # 13
    out pool = np.zeros((H pool, W pool, F))
    # To backpropagate through max pooling, record the mask.
    self.pool mask = np.zeros like(self.conv out)
    for f in range(F):
        for i in range(H pool):
            for j in range(W pool):
               h start = i * self.pool size
               h end = h start + self.pool size
                w start = j * self.pool size
                w end = w start + self.pool size
               patch = self.conv out[h start:h end, w start:w end, f]
               max val = np.max(patch)
                out pool[i, j, f] = max val
                mask = (patch == max val)
                self.pool mask[h start:h end, w start:w end, f] = mask
    self.pool out = out pool
    self.flat = self.pool out.reshape(-1, 1)
    Z fc = np.dot(self.W fc, self.flat) + self.b fc
    self.fc out = relu(Z fc)
    return self.fc out
def backward(self, dZ fc, learning rate=0.01):
```

```
Backward propagation for the CNN.
        Arguments:
        - dZ fc: Gradient of the loss with respect to the FC layer output (shape: (num c
        - learning rate: Learning rate for parameter updates.
        Returns:
        - Gradients for parameters (for illustration).
        dW fc = np.dot(dZ fc, self.flat.T)
        db fc = dZ fc
        dflat = np.dot(self.W fc.T, dZ fc)
        dpool = dflat.reshape(self.pool out.shape)
        dconv out = np.zeros like(self.conv out) # shape: (26,26,num filters)
        H pool, W pool, F = self.pool out.shape
        for f in range(F):
            for i in range(H pool):
                for j in range(W pool):
                    h start = i * self.pool size
                    h end = h start + self.pool size
                    w start = j * self.pool size
                    w end = w start + self.pool size
                    dconv out[h start:h end, w start:w end, f] += self.pool mask[h start
        dconv out[self.conv out <= 0] = 0
        dW conv = np.zeros like(self.W conv)
        db conv = np.zeros like(self.b conv)
        H, W, C = self.X.shape
        H conv, W conv, F = self.conv out.shape # (26,26,num filters)
        for f in range(F):
            for i in range(H conv):
                for j in range(W conv):
                    patch = self.X[i:i+self.kernel size, j:j+self.kernel size, :] # sha
                    dW_{conv}[:, :, :, f] += patch * dconv out[i, j, f]
                    db conv[0, 0, 0, f] += dconv out[i, j, f]
        # ---- Parameter Update ----
        self.W fc -= learning rate * dW fc
        self.b fc -= learning rate * db fc
        self.W conv -= learning rate * dW conv
        self.b conv -= learning rate * db conv
        return dW fc, db fc, dW conv, db conv
np.random.seed(42)
cnn = CNN(kernel size=3, num filters=8, pool size=2, num classes=10)
X = np.random.randn(28, 28, 1)
out = cnn.forward(X)
print("CNN forward output (FC layer activations):\n", out)
dZ fc = np.random.randn(10, 1)
grads = cnn.backward(dZ fc, learning rate=0.01)
print("\nParameter gradients computed during backward propagation:")
print("dW fc shape:", grads[0].shape)
print("dW conv shape:", grads[2].shape)
```

```
In [4]: import random
       import matplotlib.pyplot as plt
       import kagglehub
       import numpy as np
       import struct
       from array import array
       from os.path import join
       import torch
       import torch.nn as nn
       import torch.optim as optim
       from torch.utils.data import TensorDataset, DataLoader
       from torchvision import transforms
       from tqdm import tqdm # Standard tqdm package
       import numpy as np
       from matplotlib import pyplot as plt
       from sklearn.metrics import confusion matrix
## Install the training material
       path = kagglehub.dataset download("hojjatk/mnist-dataset")
       print("Path to dataset files:", path)
       data set location = r"C:\Users\Eric\.cache\kagglehub\datasets\hojjatk\mnist-dataset\vers
       Downloading from https://www.kaggle.com/api/v1/datasets/download/hojjatk/mnist-dataset?d
       ataset version number=1...
       100%| 22.0M/22.0M [00:00<00:00, 65.0MB/s]
       Extracting files...
       Path to dataset files: C:\Users\Eric\.cache\kagglehub\datasets\hojjatk\mnist-dataset\ver
       sions \1
## Define the loading function for the training material, obtained from the example give
       class MnistDataloader(object):
          def init (self, training images filepath, training labels filepath,
                      test images filepath, test labels filepath):
              self.training images filepath = training images filepath
              self.training labels filepath = training labels filepath
              self.test images filepath = test images filepath
              self.test labels filepath = test labels filepath
          def read images labels (self, images filepath, labels filepath):
              labels = []
              with open(labels filepath, 'rb') as file:
                 magic, size = struct.unpack(">II", file.read(8))
                 if magic != 2049:
                     raise ValueError ('Magic number mismatch, expected 2049, got {}'.format(m
                  labels = array("B", file.read())
              with open(images filepath, 'rb') as file:
                 magic, size, rows, cols = struct.unpack(">IIII", file.read(16))
                  if magic != 2051:
                     raise ValueError('Magic number mismatch, expected 2051, got {}'.format(m
                 image data = array("B", file.read())
              images = []
              for i in range(size):
                 images.append([0] * rows * cols)
              for i in range(size):
                 img = np.array(image data[i * rows * cols:(i + 1) * rows * cols])
                  img = img.reshape(28, 28)
                  images[i][:] = img
              return images, labels
```

```
def load data(self):
       x train, y train = self.read images labels(self.training images filepath, self.t
       x test, y test = self.read images labels(self.test images filepath, self.test la
       return (x train, y train), (x test, y test)
## Example loading
input path = r"C:\Users\Eric\.cache\kagglehub\datasets\hojjatk\mnist-dataset\versions\1"
training images filepath = join(input path, 'train-images-idx3-ubyte/train-images-idx3-u
training labels filepath = join(input path, 'train-labels-idx1-ubyte/train-labels-idx1-u
test images filepath = join(input_path, 't10k-images-idx3-ubyte/t10k-images-idx3-ubyte')
test labels filepath = join(input path, 't10k-labels-idx1-ubyte/t10k-labels-idx1-ubyte')
def show images (images, title texts):
   cols = 5
   rows = int(len(images)/cols) + 1
   plt.figure(figsize=(30,20))
   for x in zip(images, title texts):
      image = x[0]
      title text = x[1]
      plt.subplot(rows, cols, index)
       plt.imshow(image, cmap=plt.cm.gray)
       if (title text != ''):
          plt.title(title text, fontsize = 15);
       index += 1
mnist dataloader = MnistDataloader(training images filepath, training labels filepath, t
(x train, y train), (x test, y test) = mnist dataloader.load data()
images 2 show = []
titles 2 show = []
for i in range (0, 10):
   r = random.randint(1, 60000)
   images 2 show.append(x train[r])
   titles 2 show.append('training image [' + str(r) + '] = ' + str(y train[r]))
for i in range (0, 5):
   r = random.randint(1, 10000)
   images 2 show.append(x test[r])
   titles 2 show.append('test image [' + str(r) + '] = ' + str(y test[r]))
show images (images 2 show, titles 2 show)
```



```
In [ ]: | # -----
                 # Set up file paths for the MNIST dataset (adjust these paths as needed)
                 input path = r"C:\Users\Eric\.cache\kagglehub\datasets\hojjatk\mnist-dataset\versions\1"
                training images filepath = join(input path, 'train-images-idx3-ubyte/train-images-idx3-u
                training labels filepath = join(input path, 'train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-labels-idx1-ubyte/train-lab
                test images filepath = join(input path, 't10k-images-idx3-ubyte/t10k-images-idx3-ubyte')
                test labels filepath = join(input path, 't10k-labels-idx1-ubyte/t10k-labels-idx1-ubyte')
                 # Load the MNIST data using your loader
                mnist dataloader = MnistDataloader(training images filepath, training labels filepath,
                                                                                     test images filepath, test labels filepath)
                 (x_train, y_train), (x_test, y_test) = mnist dataloader.load data()
                print("MNIST data loaded.")
                # Convert the loaded images and labels to NumPy arrays
                x train = np.array(x train) # shape: (num train, 28, 28)
                x test = np.array(x test) # shape: (num test, 28, 28)
                y train = np.array(y train) # shape: (num train,)
                y test = np.array(y test) # shape: (num test,)
                 # Convert the data to PyTorch tensors (and add a channel dimension)
                train_images_tensor = torch.tensor(x_train, dtype=torch.float32).unsqueeze(1) # shape:
                train labels tensor = torch.tensor(y train, dtype=torch.long)
                test images tensor = torch.tensor(x test, dtype=torch.float32).unsqueeze(1) # shape:
                test labels tensor = torch.tensor(y test, dtype=torch.long)
                 # Create a TensorDataset for training
                train dataset = TensorDataset(train images tensor, train labels tensor)
                 # Define the CNN Model (using PyTorch modules)
                class CNN (nn.Module):
                       def init__(self):
                               super(CNN, self). init ()
                                self.conv = nn.Conv2d(in channels=1, out channels=3, kernel size=6)
```

```
self.relu = nn.ReLU()
       self.pool = nn.MaxPool2d(kernel size=2, padding=1)
       self.fc1 = nn.Linear(3 * 12 * 12, 60)
       self.sigmoid = nn.Sigmoid()
       self.fc2 = nn.Linear(60, 10)
       self.softmax = nn.Softmax(dim=1)
    def forward(self, x):
      x = self.conv(x)
       x = self.relu(x)
       x = self.pool(x)
       x = torch.flatten(x, 1)
       x = self.fcl(x)
       x = self.sigmoid(x)
       x = self.fc2(x)
       x = self.softmax(x)
       return x
    def train all(self, dataset, batch size, lr, epochs):
       loader = DataLoader(dataset, batch size=batch size, shuffle=True)
       crit = nn.CrossEntropyLoss()
       opt = optim.Adam(self.parameters(), lr=lr)
       losses = []
       total steps = len(loader) * epochs
       with tqdm(total=total steps, desc="CNN Training") as pbar:
            for epoch in range(epochs):
               self.train()
                epoch loss = 0.0
                for images, labels in loader:
                    images, labels = images.to(device), labels.to(device)
                    opt.zero grad()
                    outputs = self(images)
                    loss = crit(outputs, labels)
                    loss.backward()
                    opt.step()
                    epoch loss += loss.item()
                    pbar.update(1)
                losses.append(epoch loss)
                #print(f"Epoch {epoch+1}/{epochs}: Loss = {epoch loss:.4f}")
       return losses
# Define the MLP Model (using PyTorch modules)
class MLP(nn.Module):
   def init (self):
       super(MLP, self). init ()
       self.fc1 = nn.Linear(28 * 28, 500)
       self.relu = nn.ReLU()
       self.fc2 = nn.Linear(500, 300)
       self.sigmoid = nn.Sigmoid()
       self.fc3 = nn.Linear(300, 10)
       self.softmax = nn.Softmax(dim=1)
   def forward(self, x):
       x = torch.flatten(x, 1)
       x = self.fcl(x)
       x = self.relu(x)
       x = self.fc2(x)
       x = self.sigmoid(x)
       x = self.fc3(x)
       x = self.softmax(x)
       return x
    def train all(self, dataset, batch size, lr, epochs):
```

```
loader = DataLoader(dataset, batch size=batch size, shuffle=True)
        crit = nn.CrossEntropyLoss()
        opt = optim.Adam(self.parameters(), lr=lr)
        losses = []
        total steps = len(loader) * epochs
        with tqdm(total=total steps, desc="MLP Training") as pbar:
            for epoch in range(epochs):
                self.train()
                epoch loss = 0.0
                for images, labels in loader:
                    images, labels = images.to(device), labels.to(device)
                    opt.zero grad()
                    outputs = self(images)
                    loss = crit(outputs, labels)
                    loss.backward()
                    opt.step()
                    epoch loss += loss.item()
                    pbar.update(1)
                losses.append(epoch loss)
        return losses
def conf matrix(model):
   model.eval()
   with torch.no grad():
        outputs = model(test images tensor.to(device))
        preds = torch.argmax(outputs, dim=1).cpu().numpy()
    conf = confusion matrix(y test, preds)
    fig, ax = plt.subplots()
    im = ax.imshow(conf, cmap='Reds')
   ax.set xticks(np.arange(conf.shape[1]))
   ax.set yticks(np.arange(conf.shape[0]))
   ax.set xticklabels(range(10))
    ax.set yticklabels(range(10))
   plt.setp(ax.get xticklabels(), rotation=45, ha="right", rotation mode="anchor")
    thresh = conf.max() / 2.0
    for i in range(conf.shape[0]):
        for j in range(conf.shape[1]):
            color = "white" if conf[i, j] > thresh else "black"
            ax.text(j, i, f"{conf[i, j]:d}", ha="center", va="center", color=color)
   ax.set xlabel('Predicted')
   ax.set ylabel('Actual')
   plt.title("Confusion Matrix")
   plt.colorbar(im)
   plt.show()
```

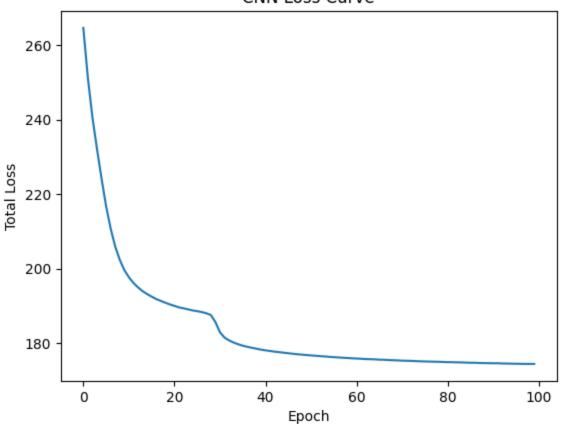
MNIST data loaded.

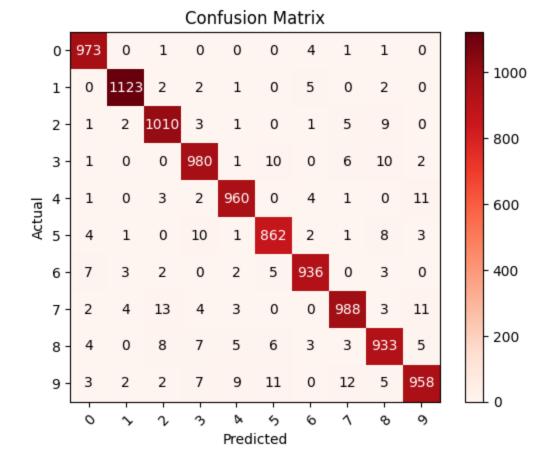
```
print(mlp.fc1.weight.shape)
print(mlp.fc1.bias.shape)

torch.Size([500, 784])
torch.Size([500])

In []: cnn_losses = cnn.train_all(train_dataset, batch_size=batch, lr=iteration_step, epochs=ep
plt.plot(cnn_losses)
plt.title("CNN Loss Curve")
plt.xlabel("Epoch")
plt.ylabel("Total Loss")
plt.show()
conf_matrix(cnn)
CNN Training: 100%| 11800/11800 [01:01<00:00, 192.43it/s]
```

CNN Loss Curve





```
In []: mlp_losses = mlp.train_all(train_dataset, batch_size=batch, lr=iteration_step, epochs=ep
plt.plot(mlp_losses)
plt.title("MLP Loss Curve")
plt.xlabel("Epoch")
plt.ylabel("Total Loss")
plt.show()
conf_matrix(mlp)
MLP Training: 100%| 11800/11800 [00:56<00:00, 210.57it/s]
```

