hw4

March 3, 2023

[1]:
```
%matplotlib inline
import matplotlib.pyplot as plt
import numpy as np
import torch
import torchvision
import torchvision.transforms as transforms
import torch.nn as nn
import torch.nn.functional as F
import torch.optim as optim
```

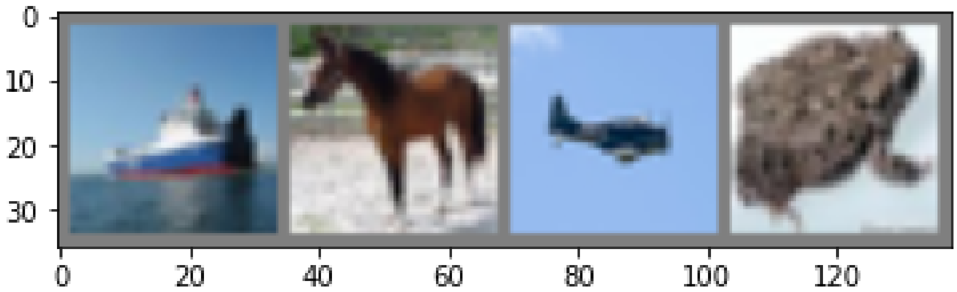
Prepare for Dataset

[2]:
```
transform = transforms.Compose(
    [transforms.ToTensor(),
     transforms.Normalize((0.5, 0.5, 0.5), (0.5, 0.5, 0.5))])

trainset = torchvision.datasets.CIFAR10(root='./data', train=True,
                                        download=True, transform=transform)
trainloader = torch.utils.data.DataLoader(trainset, batch_size=4,
                                          shuffle=True, num_workers=2)

testset = torchvision.datasets.CIFAR10(root='./data', train=False,
                                       download=True, transform=transform)
testloader = torch.utils.data.DataLoader(testset, batch_size=4,
                                         shuffle=False, num_workers=2)

classes = ('plane', 'car', 'bird', 'cat',
           'deer', 'dog', 'frog', 'horse', 'ship', 'truck')
```

```
Files already downloaded and verified
Files already downloaded and verified
```

[3]:
```
# The function to show an image.
def imshow(img):
    img = img / 2 + 0.5     # Unnormalize.
    npimg = img.numpy()
    plt.imshow(np.transpose(npimg, (1, 2, 0)))
    plt.show()
```

```python
# Get some random training images.
dataiter = iter(trainloader)
images, labels = next(dataiter)
# Show images.
imshow(torchvision.utils.make_grid(images))
# Print labels.
print(' '.join('%5s' % classes[labels[j]] for j in range(4)))
```



```
 ship horse plane  frog
```

Choose a Device

[4]:
```python
# If there are GPUs, choose the first one for computing. Otherwise use CPU.
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
print(device)
# If 'cuda:0' is printed, it means GPU is available.
```

```
cpu
```

Network Definition

[5]:
```python
class Net(nn.Module):
    def __init__(self):
        super(Net, self).__init__()
        ###### Fill the blank here ######
        self.conv1 = nn.Conv2d(3, 10, kernel_size=3, padding=1)

        self.relu1 = nn.ReLU()

        self.pool1 = nn.AvgPool2d(kernel_size=2, stride=2)

        self.conv2 = nn.Conv2d(10, 20, kernel_size=3, padding=1)

        self.relu2 = nn.ReLU()

        self.pool2 = nn.AvgPool2d(kernel_size=2, stride=2)
```

```python
        self.conv3 = nn.Conv2d(20, 30, kernel_size=3, padding=1)

        self.relu3 = nn.ReLU()

        self.pool3 = nn.AvgPool2d(kernel_size=2, stride=2)

        self.fc1 = nn.Linear(30 * 4 * 4, 100)

        self.relu4 = nn.ReLU()

        self.fc2 = nn.Linear(100, 10)


    def forward(self, x):
        ###### Fill the blank here ######
        x = self.pool1(self.relu1(self.conv1(x)))
        x = self.pool2(self.relu2(self.conv2(x)))
        x = self.pool3(self.relu3(self.conv3(x)))
        x = x.view(-1, 30 * 4 * 4)
        x = self.relu4(self.fc1(x))
        x = self.fc2(x)
        return x



net = Net()      # Create the network instance.
net.to(device)  # Move the network parameters to the specified device.
```

[5]:
```
Net(
  (conv1): Conv2d(3, 10, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
  (relu1): ReLU()
  (pool1): AvgPool2d(kernel_size=2, stride=2, padding=0)
  (conv2): Conv2d(10, 20, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
  (relu2): ReLU()
  (pool2): AvgPool2d(kernel_size=2, stride=2, padding=0)
  (conv3): Conv2d(20, 30, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
  (relu3): ReLU()
  (pool3): AvgPool2d(kernel_size=2, stride=2, padding=0)
  (fc1): Linear(in_features=480, out_features=100, bias=True)
  (relu4): ReLU()
  (fc2): Linear(in_features=100, out_features=10, bias=True)
)
```

Optimizer and Loss Function

[15]:

```python
# We use cross-entropy as loss function.
loss_func = nn.CrossEntropyLoss()
# We use stochastic gradient descent (SGD) as optimizer.
opt = optim.SGD(net.parameters(), lr=0.0005, momentum=0.9)
```

Training Procedure

[16]:

```python
avg_losses = []   # Avg. losses.
epochs = 10       # Total epochs.
print_freq = 100  # Print frequency.

for epoch in range(epochs):  # Loop over the dataset multiple times.
    running_loss = 0.0       # Initialize running loss.
    for i, data in enumerate(trainloader, 0):
        # Get the inputs.
        inputs, labels = data

        # Move the inputs to the specified device.
        inputs, labels = inputs.to(device), labels.to(device)

        # Zero the parameter gradients.
        opt.zero_grad()

        # Forward step.
        outputs = net(inputs)
        loss = loss_func(outputs, labels)

        # Backward step.
        loss.backward()

        # Optimization step (update the parameters).
        opt.step()

        # Print statistics.
        running_loss += loss.item()
        if i % print_freq == print_freq - 1: # Print every several mini-batches.
            avg_loss = running_loss / print_freq
            print('[epoch: {}, i: {:5d}] avg mini-batch loss: {:.3f}'.format(
                epoch, i, avg_loss))
            avg_losses.append(avg_loss)
            running_loss = 0.0

print('Finished Training.')
```

```
[epoch: 0, i:    99] avg mini-batch loss: 0.763
[epoch: 0, i:   199] avg mini-batch loss: 0.700
[epoch: 0, i:   299] avg mini-batch loss: 0.722
[epoch: 0, i:   399] avg mini-batch loss: 0.708
```

```
[epoch: 0, i:   499] avg mini-batch loss: 0.574
[epoch: 0, i:   599] avg mini-batch loss: 0.657
[epoch: 0, i:   699] avg mini-batch loss: 0.647
[epoch: 0, i:   799] avg mini-batch loss: 0.600
[epoch: 0, i:   899] avg mini-batch loss: 0.599
[epoch: 0, i:   999] avg mini-batch loss: 0.617
[epoch: 0, i:  1099] avg mini-batch loss: 0.579
[epoch: 0, i:  1199] avg mini-batch loss: 0.628
[epoch: 0, i:  1299] avg mini-batch loss: 0.580
[epoch: 0, i:  1399] avg mini-batch loss: 0.573
[epoch: 0, i:  1499] avg mini-batch loss: 0.565
[epoch: 0, i:  1599] avg mini-batch loss: 0.601
[epoch: 0, i:  1699] avg mini-batch loss: 0.582
[epoch: 0, i:  1799] avg mini-batch loss: 0.654
[epoch: 0, i:  1899] avg mini-batch loss: 0.560
[epoch: 0, i:  1999] avg mini-batch loss: 0.498
[epoch: 0, i:  2099] avg mini-batch loss: 0.497
[epoch: 0, i:  2199] avg mini-batch loss: 0.503
[epoch: 0, i:  2299] avg mini-batch loss: 0.639
[epoch: 0, i:  2399] avg mini-batch loss: 0.587
[epoch: 0, i:  2499] avg mini-batch loss: 0.606
[epoch: 0, i:  2599] avg mini-batch loss: 0.571
[epoch: 0, i:  2699] avg mini-batch loss: 0.604
[epoch: 0, i:  2799] avg mini-batch loss: 0.546
[epoch: 0, i:  2899] avg mini-batch loss: 0.614
[epoch: 0, i:  2999] avg mini-batch loss: 0.625
[epoch: 0, i:  3099] avg mini-batch loss: 0.468
[epoch: 0, i:  3199] avg mini-batch loss: 0.580
[epoch: 0, i:  3299] avg mini-batch loss: 0.526
[epoch: 0, i:  3399] avg mini-batch loss: 0.623
[epoch: 0, i:  3499] avg mini-batch loss: 0.562
[epoch: 0, i:  3599] avg mini-batch loss: 0.650
[epoch: 0, i:  3699] avg mini-batch loss: 0.443
[epoch: 0, i:  3799] avg mini-batch loss: 0.564
[epoch: 0, i:  3899] avg mini-batch loss: 0.518
[epoch: 0, i:  3999] avg mini-batch loss: 0.536
[epoch: 0, i:  4099] avg mini-batch loss: 0.483
[epoch: 0, i:  4199] avg mini-batch loss: 0.515
[epoch: 0, i:  4299] avg mini-batch loss: 0.599
[epoch: 0, i:  4399] avg mini-batch loss: 0.488
[epoch: 0, i:  4499] avg mini-batch loss: 0.524
[epoch: 0, i:  4599] avg mini-batch loss: 0.421
[epoch: 0, i:  4699] avg mini-batch loss: 0.534
[epoch: 0, i:  4799] avg mini-batch loss: 0.455
[epoch: 0, i:  4899] avg mini-batch loss: 0.476
[epoch: 0, i:  4999] avg mini-batch loss: 0.492
[epoch: 0, i:  5099] avg mini-batch loss: 0.461
[epoch: 0, i:  5199] avg mini-batch loss: 0.530
```

```
[epoch: 0, i:  5299] avg mini-batch loss: 0.566
[epoch: 0, i:  5399] avg mini-batch loss: 0.472
[epoch: 0, i:  5499] avg mini-batch loss: 0.534
[epoch: 0, i:  5599] avg mini-batch loss: 0.500
[epoch: 0, i:  5699] avg mini-batch loss: 0.491
[epoch: 0, i:  5799] avg mini-batch loss: 0.486
[epoch: 0, i:  5899] avg mini-batch loss: 0.517
[epoch: 0, i:  5999] avg mini-batch loss: 0.566
[epoch: 0, i:  6099] avg mini-batch loss: 0.537
[epoch: 0, i:  6199] avg mini-batch loss: 0.558
[epoch: 0, i:  6299] avg mini-batch loss: 0.540
[epoch: 0, i:  6399] avg mini-batch loss: 0.533
[epoch: 0, i:  6499] avg mini-batch loss: 0.513
[epoch: 0, i:  6599] avg mini-batch loss: 0.586
[epoch: 0, i:  6699] avg mini-batch loss: 0.435
[epoch: 0, i:  6799] avg mini-batch loss: 0.538
[epoch: 0, i:  6899] avg mini-batch loss: 0.524
[epoch: 0, i:  6999] avg mini-batch loss: 0.517
[epoch: 0, i:  7099] avg mini-batch loss: 0.484
[epoch: 0, i:  7199] avg mini-batch loss: 0.554
[epoch: 0, i:  7299] avg mini-batch loss: 0.563
[epoch: 0, i:  7399] avg mini-batch loss: 0.564
[epoch: 0, i:  7499] avg mini-batch loss: 0.530
[epoch: 0, i:  7599] avg mini-batch loss: 0.491
[epoch: 0, i:  7699] avg mini-batch loss: 0.480
[epoch: 0, i:  7799] avg mini-batch loss: 0.474
[epoch: 0, i:  7899] avg mini-batch loss: 0.548
[epoch: 0, i:  7999] avg mini-batch loss: 0.529
[epoch: 0, i:  8099] avg mini-batch loss: 0.553
[epoch: 0, i:  8199] avg mini-batch loss: 0.502
[epoch: 0, i:  8299] avg mini-batch loss: 0.592
[epoch: 0, i:  8399] avg mini-batch loss: 0.483
[epoch: 0, i:  8499] avg mini-batch loss: 0.501
[epoch: 0, i:  8599] avg mini-batch loss: 0.558
[epoch: 0, i:  8699] avg mini-batch loss: 0.439
[epoch: 0, i:  8799] avg mini-batch loss: 0.483
[epoch: 0, i:  8899] avg mini-batch loss: 0.524
[epoch: 0, i:  8999] avg mini-batch loss: 0.451
[epoch: 0, i:  9099] avg mini-batch loss: 0.539
[epoch: 0, i:  9199] avg mini-batch loss: 0.504
[epoch: 0, i:  9299] avg mini-batch loss: 0.553
[epoch: 0, i:  9399] avg mini-batch loss: 0.494
[epoch: 0, i:  9499] avg mini-batch loss: 0.475
[epoch: 0, i:  9599] avg mini-batch loss: 0.508
[epoch: 0, i:  9699] avg mini-batch loss: 0.623
[epoch: 0, i:  9799] avg mini-batch loss: 0.628
[epoch: 0, i:  9899] avg mini-batch loss: 0.463
[epoch: 0, i:  9999] avg mini-batch loss: 0.594
```

```
[epoch: 0, i: 10099] avg mini-batch loss: 0.448
[epoch: 0, i: 10199] avg mini-batch loss: 0.433
[epoch: 0, i: 10299] avg mini-batch loss: 0.495
[epoch: 0, i: 10399] avg mini-batch loss: 0.489
[epoch: 0, i: 10499] avg mini-batch loss: 0.453
[epoch: 0, i: 10599] avg mini-batch loss: 0.511
[epoch: 0, i: 10699] avg mini-batch loss: 0.465
[epoch: 0, i: 10799] avg mini-batch loss: 0.560
[epoch: 0, i: 10899] avg mini-batch loss: 0.518
[epoch: 0, i: 10999] avg mini-batch loss: 0.527
[epoch: 0, i: 11099] avg mini-batch loss: 0.610
[epoch: 0, i: 11199] avg mini-batch loss: 0.557
[epoch: 0, i: 11299] avg mini-batch loss: 0.514
[epoch: 0, i: 11399] avg mini-batch loss: 0.455
[epoch: 0, i: 11499] avg mini-batch loss: 0.458
[epoch: 0, i: 11599] avg mini-batch loss: 0.559
[epoch: 0, i: 11699] avg mini-batch loss: 0.496
[epoch: 0, i: 11799] avg mini-batch loss: 0.495
[epoch: 0, i: 11899] avg mini-batch loss: 0.492
[epoch: 0, i: 11999] avg mini-batch loss: 0.586
[epoch: 0, i: 12099] avg mini-batch loss: 0.512
[epoch: 0, i: 12199] avg mini-batch loss: 0.464
[epoch: 0, i: 12299] avg mini-batch loss: 0.492
[epoch: 0, i: 12399] avg mini-batch loss: 0.518
[epoch: 0, i: 12499] avg mini-batch loss: 0.515
[epoch: 1, i:    99] avg mini-batch loss: 0.433
[epoch: 1, i:   199] avg mini-batch loss: 0.523
[epoch: 1, i:   299] avg mini-batch loss: 0.410
[epoch: 1, i:   399] avg mini-batch loss: 0.492
[epoch: 1, i:   499] avg mini-batch loss: 0.407
[epoch: 1, i:   599] avg mini-batch loss: 0.457
[epoch: 1, i:   699] avg mini-batch loss: 0.417
[epoch: 1, i:   799] avg mini-batch loss: 0.508
[epoch: 1, i:   899] avg mini-batch loss: 0.452
[epoch: 1, i:   999] avg mini-batch loss: 0.519
[epoch: 1, i:  1099] avg mini-batch loss: 0.509
[epoch: 1, i:  1199] avg mini-batch loss: 0.524
[epoch: 1, i:  1299] avg mini-batch loss: 0.397
[epoch: 1, i:  1399] avg mini-batch loss: 0.447
[epoch: 1, i:  1499] avg mini-batch loss: 0.464
[epoch: 1, i:  1599] avg mini-batch loss: 0.480
[epoch: 1, i:  1699] avg mini-batch loss: 0.500
[epoch: 1, i:  1799] avg mini-batch loss: 0.405
[epoch: 1, i:  1899] avg mini-batch loss: 0.380
[epoch: 1, i:  1999] avg mini-batch loss: 0.454
[epoch: 1, i:  2099] avg mini-batch loss: 0.449
[epoch: 1, i:  2199] avg mini-batch loss: 0.407
[epoch: 1, i:  2299] avg mini-batch loss: 0.495
```

```
[epoch: 1, i:  2399] avg mini-batch loss: 0.415
[epoch: 1, i:  2499] avg mini-batch loss: 0.485
[epoch: 1, i:  2599] avg mini-batch loss: 0.384
[epoch: 1, i:  2699] avg mini-batch loss: 0.459
[epoch: 1, i:  2799] avg mini-batch loss: 0.437
[epoch: 1, i:  2899] avg mini-batch loss: 0.440
[epoch: 1, i:  2999] avg mini-batch loss: 0.510
[epoch: 1, i:  3099] avg mini-batch loss: 0.482
[epoch: 1, i:  3199] avg mini-batch loss: 0.445
[epoch: 1, i:  3299] avg mini-batch loss: 0.477
[epoch: 1, i:  3399] avg mini-batch loss: 0.571
[epoch: 1, i:  3499] avg mini-batch loss: 0.481
[epoch: 1, i:  3599] avg mini-batch loss: 0.493
[epoch: 1, i:  3699] avg mini-batch loss: 0.518
[epoch: 1, i:  3799] avg mini-batch loss: 0.445
[epoch: 1, i:  3899] avg mini-batch loss: 0.380
[epoch: 1, i:  3999] avg mini-batch loss: 0.374
[epoch: 1, i:  4099] avg mini-batch loss: 0.472
[epoch: 1, i:  4199] avg mini-batch loss: 0.446
[epoch: 1, i:  4299] avg mini-batch loss: 0.484
[epoch: 1, i:  4399] avg mini-batch loss: 0.517
[epoch: 1, i:  4499] avg mini-batch loss: 0.524
[epoch: 1, i:  4599] avg mini-batch loss: 0.423
[epoch: 1, i:  4699] avg mini-batch loss: 0.452
[epoch: 1, i:  4799] avg mini-batch loss: 0.462
[epoch: 1, i:  4899] avg mini-batch loss: 0.504
[epoch: 1, i:  4999] avg mini-batch loss: 0.422
[epoch: 1, i:  5099] avg mini-batch loss: 0.481
[epoch: 1, i:  5199] avg mini-batch loss: 0.454
[epoch: 1, i:  5299] avg mini-batch loss: 0.469
[epoch: 1, i:  5399] avg mini-batch loss: 0.522
[epoch: 1, i:  5499] avg mini-batch loss: 0.429
[epoch: 1, i:  5599] avg mini-batch loss: 0.496
[epoch: 1, i:  5699] avg mini-batch loss: 0.386
[epoch: 1, i:  5799] avg mini-batch loss: 0.423
[epoch: 1, i:  5899] avg mini-batch loss: 0.545
[epoch: 1, i:  5999] avg mini-batch loss: 0.510
[epoch: 1, i:  6099] avg mini-batch loss: 0.445
[epoch: 1, i:  6199] avg mini-batch loss: 0.542
[epoch: 1, i:  6299] avg mini-batch loss: 0.460
[epoch: 1, i:  6399] avg mini-batch loss: 0.513
[epoch: 1, i:  6499] avg mini-batch loss: 0.505
[epoch: 1, i:  6599] avg mini-batch loss: 0.513
[epoch: 1, i:  6699] avg mini-batch loss: 0.500
[epoch: 1, i:  6799] avg mini-batch loss: 0.396
[epoch: 1, i:  6899] avg mini-batch loss: 0.463
[epoch: 1, i:  6999] avg mini-batch loss: 0.466
[epoch: 1, i:  7099] avg mini-batch loss: 0.488
```

```
[epoch: 1, i:  7199] avg mini-batch loss: 0.468
[epoch: 1, i:  7299] avg mini-batch loss: 0.513
[epoch: 1, i:  7399] avg mini-batch loss: 0.405
[epoch: 1, i:  7499] avg mini-batch loss: 0.477
[epoch: 1, i:  7599] avg mini-batch loss: 0.549
[epoch: 1, i:  7699] avg mini-batch loss: 0.439
[epoch: 1, i:  7799] avg mini-batch loss: 0.526
[epoch: 1, i:  7899] avg mini-batch loss: 0.384
[epoch: 1, i:  7999] avg mini-batch loss: 0.388
[epoch: 1, i:  8099] avg mini-batch loss: 0.434
[epoch: 1, i:  8199] avg mini-batch loss: 0.428
[epoch: 1, i:  8299] avg mini-batch loss: 0.373
[epoch: 1, i:  8399] avg mini-batch loss: 0.430
[epoch: 1, i:  8499] avg mini-batch loss: 0.444
[epoch: 1, i:  8599] avg mini-batch loss: 0.530
[epoch: 1, i:  8699] avg mini-batch loss: 0.470
[epoch: 1, i:  8799] avg mini-batch loss: 0.519
[epoch: 1, i:  8899] avg mini-batch loss: 0.421
[epoch: 1, i:  8999] avg mini-batch loss: 0.439
[epoch: 1, i:  9099] avg mini-batch loss: 0.498
[epoch: 1, i:  9199] avg mini-batch loss: 0.530
[epoch: 1, i:  9299] avg mini-batch loss: 0.463
[epoch: 1, i:  9399] avg mini-batch loss: 0.440
[epoch: 1, i:  9499] avg mini-batch loss: 0.442
[epoch: 1, i:  9599] avg mini-batch loss: 0.429
[epoch: 1, i:  9699] avg mini-batch loss: 0.375
[epoch: 1, i:  9799] avg mini-batch loss: 0.498
[epoch: 1, i:  9899] avg mini-batch loss: 0.473
[epoch: 1, i:  9999] avg mini-batch loss: 0.492
[epoch: 1, i: 10099] avg mini-batch loss: 0.459
[epoch: 1, i: 10199] avg mini-batch loss: 0.405
[epoch: 1, i: 10299] avg mini-batch loss: 0.475
[epoch: 1, i: 10399] avg mini-batch loss: 0.481
[epoch: 1, i: 10499] avg mini-batch loss: 0.499
[epoch: 1, i: 10599] avg mini-batch loss: 0.410
[epoch: 1, i: 10699] avg mini-batch loss: 0.435
[epoch: 1, i: 10799] avg mini-batch loss: 0.545
[epoch: 1, i: 10899] avg mini-batch loss: 0.432
[epoch: 1, i: 10999] avg mini-batch loss: 0.508
[epoch: 1, i: 11099] avg mini-batch loss: 0.484
[epoch: 1, i: 11199] avg mini-batch loss: 0.448
[epoch: 1, i: 11299] avg mini-batch loss: 0.459
[epoch: 1, i: 11399] avg mini-batch loss: 0.405
[epoch: 1, i: 11499] avg mini-batch loss: 0.425
[epoch: 1, i: 11599] avg mini-batch loss: 0.452
[epoch: 1, i: 11699] avg mini-batch loss: 0.498
[epoch: 1, i: 11799] avg mini-batch loss: 0.407
[epoch: 1, i: 11899] avg mini-batch loss: 0.438
```

```
[epoch: 1, i: 11999] avg mini-batch loss: 0.484
[epoch: 1, i: 12099] avg mini-batch loss: 0.549
[epoch: 1, i: 12199] avg mini-batch loss: 0.530
[epoch: 1, i: 12299] avg mini-batch loss: 0.505
[epoch: 1, i: 12399] avg mini-batch loss: 0.453
[epoch: 1, i: 12499] avg mini-batch loss: 0.495
[epoch: 2, i:    99] avg mini-batch loss: 0.443
[epoch: 2, i:   199] avg mini-batch loss: 0.410
[epoch: 2, i:   299] avg mini-batch loss: 0.463
[epoch: 2, i:   399] avg mini-batch loss: 0.403
[epoch: 2, i:   499] avg mini-batch loss: 0.361
[epoch: 2, i:   599] avg mini-batch loss: 0.443
[epoch: 2, i:   699] avg mini-batch loss: 0.486
[epoch: 2, i:   799] avg mini-batch loss: 0.406
[epoch: 2, i:   899] avg mini-batch loss: 0.425
[epoch: 2, i:   999] avg mini-batch loss: 0.438
[epoch: 2, i:  1099] avg mini-batch loss: 0.404
[epoch: 2, i:  1199] avg mini-batch loss: 0.381
[epoch: 2, i:  1299] avg mini-batch loss: 0.405
[epoch: 2, i:  1399] avg mini-batch loss: 0.369
[epoch: 2, i:  1499] avg mini-batch loss: 0.440
[epoch: 2, i:  1599] avg mini-batch loss: 0.430
[epoch: 2, i:  1699] avg mini-batch loss: 0.428
[epoch: 2, i:  1799] avg mini-batch loss: 0.405
[epoch: 2, i:  1899] avg mini-batch loss: 0.418
[epoch: 2, i:  1999] avg mini-batch loss: 0.379
[epoch: 2, i:  2099] avg mini-batch loss: 0.424
[epoch: 2, i:  2199] avg mini-batch loss: 0.370
[epoch: 2, i:  2299] avg mini-batch loss: 0.375
[epoch: 2, i:  2399] avg mini-batch loss: 0.421
[epoch: 2, i:  2499] avg mini-batch loss: 0.413
[epoch: 2, i:  2599] avg mini-batch loss: 0.430
[epoch: 2, i:  2699] avg mini-batch loss: 0.479
[epoch: 2, i:  2799] avg mini-batch loss: 0.375
[epoch: 2, i:  2899] avg mini-batch loss: 0.380
[epoch: 2, i:  2999] avg mini-batch loss: 0.314
[epoch: 2, i:  3099] avg mini-batch loss: 0.368
[epoch: 2, i:  3199] avg mini-batch loss: 0.479
[epoch: 2, i:  3299] avg mini-batch loss: 0.402
[epoch: 2, i:  3399] avg mini-batch loss: 0.518
[epoch: 2, i:  3499] avg mini-batch loss: 0.418
[epoch: 2, i:  3599] avg mini-batch loss: 0.379
[epoch: 2, i:  3699] avg mini-batch loss: 0.451
[epoch: 2, i:  3799] avg mini-batch loss: 0.453
[epoch: 2, i:  3899] avg mini-batch loss: 0.476
[epoch: 2, i:  3999] avg mini-batch loss: 0.374
[epoch: 2, i:  4099] avg mini-batch loss: 0.373
[epoch: 2, i:  4199] avg mini-batch loss: 0.471
```

```
[epoch: 2, i:  4299] avg mini-batch loss: 0.419
[epoch: 2, i:  4399] avg mini-batch loss: 0.457
[epoch: 2, i:  4499] avg mini-batch loss: 0.414
[epoch: 2, i:  4599] avg mini-batch loss: 0.467
[epoch: 2, i:  4699] avg mini-batch loss: 0.415
[epoch: 2, i:  4799] avg mini-batch loss: 0.469
[epoch: 2, i:  4899] avg mini-batch loss: 0.397
[epoch: 2, i:  4999] avg mini-batch loss: 0.419
[epoch: 2, i:  5099] avg mini-batch loss: 0.419
[epoch: 2, i:  5199] avg mini-batch loss: 0.438
[epoch: 2, i:  5299] avg mini-batch loss: 0.370
[epoch: 2, i:  5399] avg mini-batch loss: 0.400
[epoch: 2, i:  5499] avg mini-batch loss: 0.504
[epoch: 2, i:  5599] avg mini-batch loss: 0.386
[epoch: 2, i:  5699] avg mini-batch loss: 0.388
[epoch: 2, i:  5799] avg mini-batch loss: 0.388
[epoch: 2, i:  5899] avg mini-batch loss: 0.387
[epoch: 2, i:  5999] avg mini-batch loss: 0.468
[epoch: 2, i:  6099] avg mini-batch loss: 0.421
[epoch: 2, i:  6199] avg mini-batch loss: 0.440
[epoch: 2, i:  6299] avg mini-batch loss: 0.399
[epoch: 2, i:  6399] avg mini-batch loss: 0.424
[epoch: 2, i:  6499] avg mini-batch loss: 0.454
[epoch: 2, i:  6599] avg mini-batch loss: 0.450
[epoch: 2, i:  6699] avg mini-batch loss: 0.450
[epoch: 2, i:  6799] avg mini-batch loss: 0.497
[epoch: 2, i:  6899] avg mini-batch loss: 0.443
[epoch: 2, i:  6999] avg mini-batch loss: 0.494
[epoch: 2, i:  7099] avg mini-batch loss: 0.424
[epoch: 2, i:  7199] avg mini-batch loss: 0.441
[epoch: 2, i:  7299] avg mini-batch loss: 0.511
[epoch: 2, i:  7399] avg mini-batch loss: 0.424
[epoch: 2, i:  7499] avg mini-batch loss: 0.435
[epoch: 2, i:  7599] avg mini-batch loss: 0.372
[epoch: 2, i:  7799] avg mini-batch loss: 0.395
[epoch: 2, i:  7899] avg mini-batch loss: 0.441
[epoch: 2, i:  7999] avg mini-batch loss: 0.411
[epoch: 2, i:  8099] avg mini-batch loss: 0.497
[epoch: 2, i:  8199] avg mini-batch loss: 0.402
[epoch: 2, i:  8299] avg mini-batch loss: 0.379
[epoch: 2, i:  8399] avg mini-batch loss: 0.481
[epoch: 2, i:  8499] avg mini-batch loss: 0.388
[epoch: 2, i:  8599] avg mini-batch loss: 0.463
[epoch: 2, i:  8699] avg mini-batch loss: 0.542
[epoch: 2, i:  8799] avg mini-batch loss: 0.409
[epoch: 2, i:  8899] avg mini-batch loss: 0.405
[epoch: 2, i:  8999] avg mini-batch loss: 0.470
[epoch: 2, i:  9099] avg mini-batch loss: 0.460
```

```
[epoch: 2, i:  9199] avg mini-batch loss: 0.453
[epoch: 2, i:  9299] avg mini-batch loss: 0.419
[epoch: 2, i:  9399] avg mini-batch loss: 0.470
[epoch: 2, i:  9499] avg mini-batch loss: 0.442
[epoch: 2, i:  9599] avg mini-batch loss: 0.480
[epoch: 2, i:  9699] avg mini-batch loss: 0.481
[epoch: 2, i:  9799] avg mini-batch loss: 0.506
[epoch: 2, i:  9899] avg mini-batch loss: 0.432
[epoch: 2, i:  9999] avg mini-batch loss: 0.467
[epoch: 2, i: 10099] avg mini-batch loss: 0.463
[epoch: 2, i: 10199] avg mini-batch loss: 0.374
[epoch: 2, i: 10299] avg mini-batch loss: 0.441
[epoch: 2, i: 10399] avg mini-batch loss: 0.387
[epoch: 2, i: 10499] avg mini-batch loss: 0.469
[epoch: 2, i: 10599] avg mini-batch loss: 0.426
[epoch: 2, i: 10699] avg mini-batch loss: 0.413
[epoch: 2, i: 10799] avg mini-batch loss: 0.515
[epoch: 2, i: 10899] avg mini-batch loss: 0.411
[epoch: 2, i: 10999] avg mini-batch loss: 0.477
[epoch: 2, i: 11099] avg mini-batch loss: 0.411
[epoch: 2, i: 11199] avg mini-batch loss: 0.469
[epoch: 2, i: 11299] avg mini-batch loss: 0.401
[epoch: 2, i: 11399] avg mini-batch loss: 0.479
[epoch: 2, i: 11499] avg mini-batch loss: 0.400
[epoch: 2, i: 11599] avg mini-batch loss: 0.524
[epoch: 2, i: 11699] avg mini-batch loss: 0.459
[epoch: 2, i: 11799] avg mini-batch loss: 0.467
[epoch: 2, i: 11899] avg mini-batch loss: 0.386
[epoch: 2, i: 11999] avg mini-batch loss: 0.472
[epoch: 2, i: 12099] avg mini-batch loss: 0.419
[epoch: 2, i: 12199] avg mini-batch loss: 0.429
[epoch: 2, i: 12299] avg mini-batch loss: 0.479
[epoch: 2, i: 12399] avg mini-batch loss: 0.350
[epoch: 2, i: 12499] avg mini-batch loss: 0.448
[epoch: 3, i:    99] avg mini-batch loss: 0.367
[epoch: 3, i:   199] avg mini-batch loss: 0.394
[epoch: 3, i:   299] avg mini-batch loss: 0.350
[epoch: 3, i:   399] avg mini-batch loss: 0.396
[epoch: 3, i:   499] avg mini-batch loss: 0.339
[epoch: 3, i:   599] avg mini-batch loss: 0.415
[epoch: 3, i:   699] avg mini-batch loss: 0.385
[epoch: 3, i:   799] avg mini-batch loss: 0.446
[epoch: 3, i:   899] avg mini-batch loss: 0.352
[epoch: 3, i:   999] avg mini-batch loss: 0.464
[epoch: 3, i:  1099] avg mini-batch loss: 0.345
[epoch: 3, i:  1199] avg mini-batch loss: 0.348
[epoch: 3, i:  1299] avg mini-batch loss: 0.368
[epoch: 3, i:  1399] avg mini-batch loss: 0.404
```

```
[epoch: 3, i:  1499] avg mini-batch loss: 0.356
[epoch: 3, i:  1599] avg mini-batch loss: 0.426
[epoch: 3, i:  1699] avg mini-batch loss: 0.451
[epoch: 3, i:  1799] avg mini-batch loss: 0.370
[epoch: 3, i:  1899] avg mini-batch loss: 0.400
[epoch: 3, i:  1999] avg mini-batch loss: 0.344
[epoch: 3, i:  2099] avg mini-batch loss: 0.416
[epoch: 3, i:  2199] avg mini-batch loss: 0.374
[epoch: 3, i:  2299] avg mini-batch loss: 0.397
[epoch: 3, i:  2399] avg mini-batch loss: 0.416
[epoch: 3, i:  2499] avg mini-batch loss: 0.400
[epoch: 3, i:  2599] avg mini-batch loss: 0.421
[epoch: 3, i:  2699] avg mini-batch loss: 0.395
[epoch: 3, i:  2799] avg mini-batch loss: 0.380
[epoch: 3, i:  2899] avg mini-batch loss: 0.407
[epoch: 3, i:  2999] avg mini-batch loss: 0.364
[epoch: 3, i:  3099] avg mini-batch loss: 0.357
[epoch: 3, i:  3199] avg mini-batch loss: 0.486
[epoch: 3, i:  3299] avg mini-batch loss: 0.428
[epoch: 3, i:  3399] avg mini-batch loss: 0.402
[epoch: 3, i:  3499] avg mini-batch loss: 0.457
[epoch: 3, i:  3599] avg mini-batch loss: 0.440
[epoch: 3, i:  3699] avg mini-batch loss: 0.384
[epoch: 3, i:  3799] avg mini-batch loss: 0.479
[epoch: 3, i:  3899] avg mini-batch loss: 0.399
[epoch: 3, i:  3999] avg mini-batch loss: 0.351
[epoch: 3, i:  4099] avg mini-batch loss: 0.409
[epoch: 3, i:  4199] avg mini-batch loss: 0.411
[epoch: 3, i:  4299] avg mini-batch loss: 0.343
[epoch: 3, i:  4399] avg mini-batch loss: 0.416
[epoch: 3, i:  4499] avg mini-batch loss: 0.438
[epoch: 3, i:  4599] avg mini-batch loss: 0.385
[epoch: 3, i:  4699] avg mini-batch loss: 0.394
[epoch: 3, i:  4799] avg mini-batch loss: 0.387
[epoch: 3, i:  4899] avg mini-batch loss: 0.352
[epoch: 3, i:  4999] avg mini-batch loss: 0.404
[epoch: 3, i:  5099] avg mini-batch loss: 0.428
[epoch: 3, i:  5199] avg mini-batch loss: 0.367
[epoch: 3, i:  5299] avg mini-batch loss: 0.405
[epoch: 3, i:  5399] avg mini-batch loss: 0.423
[epoch: 3, i:  5499] avg mini-batch loss: 0.445
[epoch: 3, i:  5599] avg mini-batch loss: 0.447
[epoch: 3, i:  5699] avg mini-batch loss: 0.482
[epoch: 3, i:  5799] avg mini-batch loss: 0.426
[epoch: 3, i:  5899] avg mini-batch loss: 0.378
[epoch: 3, i:  5999] avg mini-batch loss: 0.450
[epoch: 3, i:  6099] avg mini-batch loss: 0.464
[epoch: 3, i:  6199] avg mini-batch loss: 0.407
```

```
[epoch: 3, i:  6299] avg mini-batch loss: 0.492
[epoch: 3, i:  6399] avg mini-batch loss: 0.455
[epoch: 3, i:  6499] avg mini-batch loss: 0.396
[epoch: 3, i:  6599] avg mini-batch loss: 0.455
[epoch: 3, i:  6699] avg mini-batch loss: 0.380
[epoch: 3, i:  6799] avg mini-batch loss: 0.398
[epoch: 3, i:  6899] avg mini-batch loss: 0.435
[epoch: 3, i:  6999] avg mini-batch loss: 0.407
[epoch: 3, i:  7099] avg mini-batch loss: 0.357
[epoch: 3, i:  7199] avg mini-batch loss: 0.447
[epoch: 3, i:  7299] avg mini-batch loss: 0.407
[epoch: 3, i:  7399] avg mini-batch loss: 0.365
[epoch: 3, i:  7499] avg mini-batch loss: 0.452
[epoch: 3, i:  7599] avg mini-batch loss: 0.416
[epoch: 3, i:  7699] avg mini-batch loss: 0.371
[epoch: 3, i:  7799] avg mini-batch loss: 0.430
[epoch: 3, i:  7899] avg mini-batch loss: 0.436
[epoch: 3, i:  7999] avg mini-batch loss: 0.456
[epoch: 3, i:  8099] avg mini-batch loss: 0.406
[epoch: 3, i:  8199] avg mini-batch loss: 0.385
[epoch: 3, i:  8299] avg mini-batch loss: 0.376
[epoch: 3, i:  8399] avg mini-batch loss: 0.410
[epoch: 3, i:  8499] avg mini-batch loss: 0.421
[epoch: 3, i:  8599] avg mini-batch loss: 0.377
[epoch: 3, i:  8699] avg mini-batch loss: 0.353
[epoch: 3, i:  8799] avg mini-batch loss: 0.403
[epoch: 3, i:  8899] avg mini-batch loss: 0.393
[epoch: 3, i:  8999] avg mini-batch loss: 0.386
[epoch: 3, i:  9099] avg mini-batch loss: 0.479
[epoch: 3, i:  9199] avg mini-batch loss: 0.376
[epoch: 3, i:  9299] avg mini-batch loss: 0.494
[epoch: 3, i:  9399] avg mini-batch loss: 0.361
[epoch: 3, i:  9499] avg mini-batch loss: 0.435
[epoch: 3, i:  9599] avg mini-batch loss: 0.416
[epoch: 3, i:  9699] avg mini-batch loss: 0.401
[epoch: 3, i:  9799] avg mini-batch loss: 0.451
[epoch: 3, i:  9899] avg mini-batch loss: 0.367
[epoch: 3, i:  9999] avg mini-batch loss: 0.405
[epoch: 3, i: 10099] avg mini-batch loss: 0.418
[epoch: 3, i: 10199] avg mini-batch loss: 0.347
[epoch: 3, i: 10299] avg mini-batch loss: 0.420
[epoch: 3, i: 10399] avg mini-batch loss: 0.420
[epoch: 3, i: 10499] avg mini-batch loss: 0.359
[epoch: 3, i: 10599] avg mini-batch loss: 0.378
[epoch: 3, i: 10699] avg mini-batch loss: 0.355
[epoch: 3, i: 10799] avg mini-batch loss: 0.380
[epoch: 3, i: 10899] avg mini-batch loss: 0.356
[epoch: 3, i: 10999] avg mini-batch loss: 0.482
```

```
[epoch: 3, i: 11099] avg mini-batch loss: 0.397
[epoch: 3, i: 11199] avg mini-batch loss: 0.414
[epoch: 3, i: 11299] avg mini-batch loss: 0.402
[epoch: 3, i: 11399] avg mini-batch loss: 0.473
[epoch: 3, i: 11499] avg mini-batch loss: 0.342
[epoch: 3, i: 11599] avg mini-batch loss: 0.356
[epoch: 3, i: 11699] avg mini-batch loss: 0.409
[epoch: 3, i: 11799] avg mini-batch loss: 0.458
[epoch: 3, i: 11899] avg mini-batch loss: 0.387
[epoch: 3, i: 11999] avg mini-batch loss: 0.384
[epoch: 3, i: 12099] avg mini-batch loss: 0.361
[epoch: 3, i: 12199] avg mini-batch loss: 0.468
[epoch: 3, i: 12299] avg mini-batch loss: 0.448
[epoch: 3, i: 12399] avg mini-batch loss: 0.381
[epoch: 3, i: 12499] avg mini-batch loss: 0.415
[epoch: 4, i:    99] avg mini-batch loss: 0.416
[epoch: 4, i:   199] avg mini-batch loss: 0.341
[epoch: 4, i:   299] avg mini-batch loss: 0.321
[epoch: 4, i:   399] avg mini-batch loss: 0.367
[epoch: 4, i:   499] avg mini-batch loss: 0.368
[epoch: 4, i:   599] avg mini-batch loss: 0.401
[epoch: 4, i:   699] avg mini-batch loss: 0.399
[epoch: 4, i:   799] avg mini-batch loss: 0.326
[epoch: 4, i:   899] avg mini-batch loss: 0.415
[epoch: 4, i:   999] avg mini-batch loss: 0.399
[epoch: 4, i:  1099] avg mini-batch loss: 0.351
[epoch: 4, i:  1199] avg mini-batch loss: 0.418
[epoch: 4, i:  1299] avg mini-batch loss: 0.346
[epoch: 4, i:  1399] avg mini-batch loss: 0.353
[epoch: 4, i:  1499] avg mini-batch loss: 0.380
[epoch: 4, i:  1599] avg mini-batch loss: 0.371
[epoch: 4, i:  1699] avg mini-batch loss: 0.348
[epoch: 4, i:  1799] avg mini-batch loss: 0.368
[epoch: 4, i:  1899] avg mini-batch loss: 0.353
[epoch: 4, i:  1999] avg mini-batch loss: 0.344
[epoch: 4, i:  2099] avg mini-batch loss: 0.369
[epoch: 4, i:  2199] avg mini-batch loss: 0.363
[epoch: 4, i:  2299] avg mini-batch loss: 0.373
[epoch: 4, i:  2399] avg mini-batch loss: 0.397
[epoch: 4, i:  2499] avg mini-batch loss: 0.496
[epoch: 4, i:  2599] avg mini-batch loss: 0.422
[epoch: 4, i:  2699] avg mini-batch loss: 0.320
[epoch: 4, i:  2799] avg mini-batch loss: 0.406
[epoch: 4, i:  2899] avg mini-batch loss: 0.313
[epoch: 4, i:  2999] avg mini-batch loss: 0.368
[epoch: 4, i:  3099] avg mini-batch loss: 0.361
[epoch: 4, i:  3199] avg mini-batch loss: 0.422
[epoch: 4, i:  3299] avg mini-batch loss: 0.334
```

```
[epoch: 4, i:  3399] avg mini-batch loss: 0.352
[epoch: 4, i:  3499] avg mini-batch loss: 0.448
[epoch: 4, i:  3599] avg mini-batch loss: 0.388
[epoch: 4, i:  3699] avg mini-batch loss: 0.322
[epoch: 4, i:  3799] avg mini-batch loss: 0.434
[epoch: 4, i:  3899] avg mini-batch loss: 0.345
[epoch: 4, i:  3999] avg mini-batch loss: 0.340
[epoch: 4, i:  4099] avg mini-batch loss: 0.435
[epoch: 4, i:  4199] avg mini-batch loss: 0.345
[epoch: 4, i:  4299] avg mini-batch loss: 0.336
[epoch: 4, i:  4399] avg mini-batch loss: 0.437
[epoch: 4, i:  4499] avg mini-batch loss: 0.318
[epoch: 4, i:  4599] avg mini-batch loss: 0.357
[epoch: 4, i:  4699] avg mini-batch loss: 0.437
[epoch: 4, i:  4799] avg mini-batch loss: 0.378
[epoch: 4, i:  4899] avg mini-batch loss: 0.344
[epoch: 4, i:  4999] avg mini-batch loss: 0.385
[epoch: 4, i:  5099] avg mini-batch loss: 0.375
[epoch: 4, i:  5199] avg mini-batch loss: 0.394
[epoch: 4, i:  5299] avg mini-batch loss: 0.338
[epoch: 4, i:  5399] avg mini-batch loss: 0.363
[epoch: 4, i:  5499] avg mini-batch loss: 0.379
[epoch: 4, i:  5599] avg mini-batch loss: 0.345
[epoch: 4, i:  5699] avg mini-batch loss: 0.453
[epoch: 4, i:  5799] avg mini-batch loss: 0.466
[epoch: 4, i:  5899] avg mini-batch loss: 0.400
[epoch: 4, i:  5999] avg mini-batch loss: 0.301
[epoch: 4, i:  6099] avg mini-batch loss: 0.373
[epoch: 4, i:  6199] avg mini-batch loss: 0.351
[epoch: 4, i:  6299] avg mini-batch loss: 0.402
[epoch: 4, i:  6399] avg mini-batch loss: 0.450
[epoch: 4, i:  6499] avg mini-batch loss: 0.402
[epoch: 4, i:  6599] avg mini-batch loss: 0.361
[epoch: 4, i:  6699] avg mini-batch loss: 0.429
[epoch: 4, i:  6799] avg mini-batch loss: 0.392
[epoch: 4, i:  6899] avg mini-batch loss: 0.399
[epoch: 4, i:  6999] avg mini-batch loss: 0.419
[epoch: 4, i:  7099] avg mini-batch loss: 0.421
[epoch: 4, i:  7199] avg mini-batch loss: 0.433
[epoch: 4, i:  7299] avg mini-batch loss: 0.421
[epoch: 4, i:  7399] avg mini-batch loss: 0.366
[epoch: 4, i:  7499] avg mini-batch loss: 0.398
[epoch: 4, i:  7599] avg mini-batch loss: 0.314
[epoch: 4, i:  7699] avg mini-batch loss: 0.431
[epoch: 4, i:  7799] avg mini-batch loss: 0.396
[epoch: 4, i:  7899] avg mini-batch loss: 0.393
[epoch: 4, i:  7999] avg mini-batch loss: 0.349
[epoch: 4, i:  8099] avg mini-batch loss: 0.406
```

```
[epoch: 4, i:  8199] avg mini-batch loss: 0.365
[epoch: 4, i:  8299] avg mini-batch loss: 0.375
[epoch: 4, i:  8399] avg mini-batch loss: 0.348
[epoch: 4, i:  8499] avg mini-batch loss: 0.352
[epoch: 4, i:  8599] avg mini-batch loss: 0.344
[epoch: 4, i:  8699] avg mini-batch loss: 0.411
[epoch: 4, i:  8799] avg mini-batch loss: 0.385
[epoch: 4, i:  8899] avg mini-batch loss: 0.347
[epoch: 4, i:  8999] avg mini-batch loss: 0.375
[epoch: 4, i:  9099] avg mini-batch loss: 0.378
[epoch: 4, i:  9199] avg mini-batch loss: 0.350
[epoch: 4, i:  9299] avg mini-batch loss: 0.364
[epoch: 4, i:  9399] avg mini-batch loss: 0.376
[epoch: 4, i:  9499] avg mini-batch loss: 0.369
[epoch: 4, i:  9599] avg mini-batch loss: 0.395
[epoch: 4, i:  9699] avg mini-batch loss: 0.447
[epoch: 4, i:  9799] avg mini-batch loss: 0.412
[epoch: 4, i:  9899] avg mini-batch loss: 0.361
[epoch: 4, i:  9999] avg mini-batch loss: 0.397
[epoch: 4, i: 10099] avg mini-batch loss: 0.382
[epoch: 4, i: 10199] avg mini-batch loss: 0.374
[epoch: 4, i: 10299] avg mini-batch loss: 0.422
[epoch: 4, i: 10399] avg mini-batch loss: 0.447
[epoch: 4, i: 10499] avg mini-batch loss: 0.404
[epoch: 4, i: 10599] avg mini-batch loss: 0.406
[epoch: 4, i: 10699] avg mini-batch loss: 0.424
[epoch: 4, i: 10799] avg mini-batch loss: 0.380
[epoch: 4, i: 10899] avg mini-batch loss: 0.345
[epoch: 4, i: 10999] avg mini-batch loss: 0.398
[epoch: 4, i: 11099] avg mini-batch loss: 0.401
[epoch: 4, i: 11199] avg mini-batch loss: 0.333
[epoch: 4, i: 11299] avg mini-batch loss: 0.419
[epoch: 4, i: 11399] avg mini-batch loss: 0.380
[epoch: 4, i: 11499] avg mini-batch loss: 0.422
[epoch: 4, i: 11599] avg mini-batch loss: 0.381
[epoch: 4, i: 11699] avg mini-batch loss: 0.366
[epoch: 4, i: 11799] avg mini-batch loss: 0.400
[epoch: 4, i: 11899] avg mini-batch loss: 0.412
[epoch: 4, i: 11999] avg mini-batch loss: 0.456
[epoch: 4, i: 12099] avg mini-batch loss: 0.457
[epoch: 4, i: 12199] avg mini-batch loss: 0.376
[epoch: 4, i: 12299] avg mini-batch loss: 0.350
[epoch: 4, i: 12399] avg mini-batch loss: 0.422
[epoch: 4, i: 12499] avg mini-batch loss: 0.321
[epoch: 5, i:    99] avg mini-batch loss: 0.361
[epoch: 5, i:   199] avg mini-batch loss: 0.293
[epoch: 5, i:   299] avg mini-batch loss: 0.289
[epoch: 5, i:   399] avg mini-batch loss: 0.314
```

```
[epoch: 5, i:   499] avg mini-batch loss: 0.355
[epoch: 5, i:   599] avg mini-batch loss: 0.278
[epoch: 5, i:   699] avg mini-batch loss: 0.305
[epoch: 5, i:   799] avg mini-batch loss: 0.294
[epoch: 5, i:   899] avg mini-batch loss: 0.341
[epoch: 5, i:   999] avg mini-batch loss: 0.323
[epoch: 5, i:  1099] avg mini-batch loss: 0.354
[epoch: 5, i:  1199] avg mini-batch loss: 0.298
[epoch: 5, i:  1299] avg mini-batch loss: 0.318
[epoch: 5, i:  1399] avg mini-batch loss: 0.362
[epoch: 5, i:  1499] avg mini-batch loss: 0.299
[epoch: 5, i:  1599] avg mini-batch loss: 0.403
[epoch: 5, i:  1699] avg mini-batch loss: 0.393
[epoch: 5, i:  1799] avg mini-batch loss: 0.368
[epoch: 5, i:  1899] avg mini-batch loss: 0.337
[epoch: 5, i:  1999] avg mini-batch loss: 0.398
[epoch: 5, i:  2099] avg mini-batch loss: 0.440
[epoch: 5, i:  2199] avg mini-batch loss: 0.394
[epoch: 5, i:  2299] avg mini-batch loss: 0.324
[epoch: 5, i:  2399] avg mini-batch loss: 0.310
[epoch: 5, i:  2499] avg mini-batch loss: 0.336
[epoch: 5, i:  2599] avg mini-batch loss: 0.434
[epoch: 5, i:  2699] avg mini-batch loss: 0.282
[epoch: 5, i:  2799] avg mini-batch loss: 0.354
[epoch: 5, i:  2899] avg mini-batch loss: 0.400
[epoch: 5, i:  2999] avg mini-batch loss: 0.410
[epoch: 5, i:  3099] avg mini-batch loss: 0.341
[epoch: 5, i:  3199] avg mini-batch loss: 0.373
[epoch: 5, i:  3299] avg mini-batch loss: 0.366
[epoch: 5, i:  3399] avg mini-batch loss: 0.447
[epoch: 5, i:  3499] avg mini-batch loss: 0.331
[epoch: 5, i:  3599] avg mini-batch loss: 0.333
[epoch: 5, i:  3699] avg mini-batch loss: 0.411
[epoch: 5, i:  3799] avg mini-batch loss: 0.340
[epoch: 5, i:  3899] avg mini-batch loss: 0.380
[epoch: 5, i:  3999] avg mini-batch loss: 0.346
[epoch: 5, i:  4099] avg mini-batch loss: 0.339
[epoch: 5, i:  4199] avg mini-batch loss: 0.368
[epoch: 5, i:  4299] avg mini-batch loss: 0.304
[epoch: 5, i:  4399] avg mini-batch loss: 0.307
[epoch: 5, i:  4499] avg mini-batch loss: 0.353
[epoch: 5, i:  4599] avg mini-batch loss: 0.495
[epoch: 5, i:  4699] avg mini-batch loss: 0.318
[epoch: 5, i:  4799] avg mini-batch loss: 0.412
[epoch: 5, i:  4899] avg mini-batch loss: 0.375
[epoch: 5, i:  4999] avg mini-batch loss: 0.312
[epoch: 5, i:  5099] avg mini-batch loss: 0.359
[epoch: 5, i:  5199] avg mini-batch loss: 0.391
```

```
[epoch: 5, i:  5299] avg mini-batch loss: 0.348
[epoch: 5, i:  5399] avg mini-batch loss: 0.374
[epoch: 5, i:  5499] avg mini-batch loss: 0.403
[epoch: 5, i:  5599] avg mini-batch loss: 0.342
[epoch: 5, i:  5699] avg mini-batch loss: 0.297
[epoch: 5, i:  5799] avg mini-batch loss: 0.362
[epoch: 5, i:  5899] avg mini-batch loss: 0.381
[epoch: 5, i:  5999] avg mini-batch loss: 0.305
[epoch: 5, i:  6099] avg mini-batch loss: 0.337
[epoch: 5, i:  6199] avg mini-batch loss: 0.387
[epoch: 5, i:  6299] avg mini-batch loss: 0.332
[epoch: 5, i:  6399] avg mini-batch loss: 0.396
[epoch: 5, i:  6499] avg mini-batch loss: 0.345
[epoch: 5, i:  6599] avg mini-batch loss: 0.381
[epoch: 5, i:  6699] avg mini-batch loss: 0.346
[epoch: 5, i:  6799] avg mini-batch loss: 0.361
[epoch: 5, i:  6899] avg mini-batch loss: 0.339
[epoch: 5, i:  6999] avg mini-batch loss: 0.404
[epoch: 5, i:  7099] avg mini-batch loss: 0.410
[epoch: 5, i:  7199] avg mini-batch loss: 0.351
[epoch: 5, i:  7299] avg mini-batch loss: 0.395
[epoch: 5, i:  7399] avg mini-batch loss: 0.397
[epoch: 5, i:  7499] avg mini-batch loss: 0.376
[epoch: 5, i:  7599] avg mini-batch loss: 0.413
[epoch: 5, i:  7699] avg mini-batch loss: 0.372
[epoch: 5, i:  7799] avg mini-batch loss: 0.331
[epoch: 5, i:  7899] avg mini-batch loss: 0.333
[epoch: 5, i:  7999] avg mini-batch loss: 0.334
[epoch: 5, i:  8099] avg mini-batch loss: 0.373
[epoch: 5, i:  8199] avg mini-batch loss: 0.376
[epoch: 5, i:  8299] avg mini-batch loss: 0.331
[epoch: 5, i:  8399] avg mini-batch loss: 0.412
[epoch: 5, i:  8499] avg mini-batch loss: 0.414
[epoch: 5, i:  8599] avg mini-batch loss: 0.372
[epoch: 5, i:  8699] avg mini-batch loss: 0.348
[epoch: 5, i:  8799] avg mini-batch loss: 0.306
[epoch: 5, i:  8899] avg mini-batch loss: 0.376
[epoch: 5, i:  8999] avg mini-batch loss: 0.382
[epoch: 5, i:  9099] avg mini-batch loss: 0.435
[epoch: 5, i:  9199] avg mini-batch loss: 0.429
[epoch: 5, i:  9299] avg mini-batch loss: 0.408
[epoch: 5, i:  9399] avg mini-batch loss: 0.357
[epoch: 5, i:  9499] avg mini-batch loss: 0.391
[epoch: 5, i:  9599] avg mini-batch loss: 0.399
[epoch: 5, i:  9699] avg mini-batch loss: 0.394
[epoch: 5, i:  9799] avg mini-batch loss: 0.391
[epoch: 5, i:  9899] avg mini-batch loss: 0.338
[epoch: 5, i:  9999] avg mini-batch loss: 0.403
```

```
[epoch: 5, i: 10099] avg mini-batch loss: 0.400
[epoch: 5, i: 10199] avg mini-batch loss: 0.357
[epoch: 5, i: 10299] avg mini-batch loss: 0.433
[epoch: 5, i: 10399] avg mini-batch loss: 0.341
[epoch: 5, i: 10499] avg mini-batch loss: 0.382
[epoch: 5, i: 10599] avg mini-batch loss: 0.309
[epoch: 5, i: 10699] avg mini-batch loss: 0.401
[epoch: 5, i: 10799] avg mini-batch loss: 0.321
[epoch: 5, i: 10899] avg mini-batch loss: 0.396
[epoch: 5, i: 10999] avg mini-batch loss: 0.391
[epoch: 5, i: 11099] avg mini-batch loss: 0.330
[epoch: 5, i: 11199] avg mini-batch loss: 0.432
[epoch: 5, i: 11299] avg mini-batch loss: 0.405
[epoch: 5, i: 11399] avg mini-batch loss: 0.295
[epoch: 5, i: 11499] avg mini-batch loss: 0.413
[epoch: 5, i: 11599] avg mini-batch loss: 0.369
[epoch: 5, i: 11699] avg mini-batch loss: 0.411
[epoch: 5, i: 11799] avg mini-batch loss: 0.312
[epoch: 5, i: 11899] avg mini-batch loss: 0.374
[epoch: 5, i: 11999] avg mini-batch loss: 0.325
[epoch: 5, i: 12099] avg mini-batch loss: 0.350
[epoch: 5, i: 12199] avg mini-batch loss: 0.356
[epoch: 5, i: 12299] avg mini-batch loss: 0.373
[epoch: 5, i: 12399] avg mini-batch loss: 0.370
[epoch: 5, i: 12499] avg mini-batch loss: 0.412
[epoch: 6, i:    99] avg mini-batch loss: 0.344
[epoch: 6, i:   199] avg mini-batch loss: 0.317
[epoch: 6, i:   299] avg mini-batch loss: 0.311
[epoch: 6, i:   399] avg mini-batch loss: 0.286
[epoch: 6, i:   499] avg mini-batch loss: 0.276
[epoch: 6, i:   599] avg mini-batch loss: 0.342
[epoch: 6, i:   699] avg mini-batch loss: 0.352
[epoch: 6, i:   799] avg mini-batch loss: 0.329
[epoch: 6, i:   899] avg mini-batch loss: 0.358
[epoch: 6, i:   999] avg mini-batch loss: 0.311
[epoch: 6, i:  1099] avg mini-batch loss: 0.371
[epoch: 6, i:  1199] avg mini-batch loss: 0.274
[epoch: 6, i:  1299] avg mini-batch loss: 0.284
[epoch: 6, i:  1399] avg mini-batch loss: 0.227
[epoch: 6, i:  1499] avg mini-batch loss: 0.333
[epoch: 6, i:  1599] avg mini-batch loss: 0.354
[epoch: 6, i:  1699] avg mini-batch loss: 0.386
[epoch: 6, i:  1799] avg mini-batch loss: 0.370
[epoch: 6, i:  1899] avg mini-batch loss: 0.372
[epoch: 6, i:  1999] avg mini-batch loss: 0.264
[epoch: 6, i:  2099] avg mini-batch loss: 0.288
[epoch: 6, i:  2199] avg mini-batch loss: 0.325
[epoch: 6, i:  2299] avg mini-batch loss: 0.331
```

```
[epoch: 6, i:  2399] avg mini-batch loss: 0.290
[epoch: 6, i:  2499] avg mini-batch loss: 0.334
[epoch: 6, i:  2599] avg mini-batch loss: 0.343
[epoch: 6, i:  2699] avg mini-batch loss: 0.352
[epoch: 6, i:  2799] avg mini-batch loss: 0.407
[epoch: 6, i:  2899] avg mini-batch loss: 0.279
[epoch: 6, i:  2999] avg mini-batch loss: 0.326
[epoch: 6, i:  3099] avg mini-batch loss: 0.300
[epoch: 6, i:  3199] avg mini-batch loss: 0.381
[epoch: 6, i:  3299] avg mini-batch loss: 0.352
[epoch: 6, i:  3399] avg mini-batch loss: 0.291
[epoch: 6, i:  3499] avg mini-batch loss: 0.379
[epoch: 6, i:  3599] avg mini-batch loss: 0.424
[epoch: 6, i:  3699] avg mini-batch loss: 0.363
[epoch: 6, i:  3799] avg mini-batch loss: 0.307
[epoch: 6, i:  3899] avg mini-batch loss: 0.327
[epoch: 6, i:  3999] avg mini-batch loss: 0.358
[epoch: 6, i:  4099] avg mini-batch loss: 0.324
[epoch: 6, i:  4199] avg mini-batch loss: 0.305
[epoch: 6, i:  4299] avg mini-batch loss: 0.336
[epoch: 6, i:  4399] avg mini-batch loss: 0.350
[epoch: 6, i:  4499] avg mini-batch loss: 0.370
[epoch: 6, i:  4599] avg mini-batch loss: 0.331
[epoch: 6, i:  4699] avg mini-batch loss: 0.307
[epoch: 6, i:  4799] avg mini-batch loss: 0.318
[epoch: 6, i:  4899] avg mini-batch loss: 0.368
[epoch: 6, i:  4999] avg mini-batch loss: 0.363
[epoch: 6, i:  5099] avg mini-batch loss: 0.364
[epoch: 6, i:  5199] avg mini-batch loss: 0.329
[epoch: 6, i:  5299] avg mini-batch loss: 0.307
[epoch: 6, i:  5399] avg mini-batch loss: 0.395
[epoch: 6, i:  5499] avg mini-batch loss: 0.407
[epoch: 6, i:  5599] avg mini-batch loss: 0.340
[epoch: 6, i:  5699] avg mini-batch loss: 0.358
[epoch: 6, i:  5799] avg mini-batch loss: 0.334
[epoch: 6, i:  5899] avg mini-batch loss: 0.359
[epoch: 6, i:  5999] avg mini-batch loss: 0.317
[epoch: 6, i:  6099] avg mini-batch loss: 0.351
[epoch: 6, i:  6199] avg mini-batch loss: 0.307
[epoch: 6, i:  6299] avg mini-batch loss: 0.321
[epoch: 6, i:  6399] avg mini-batch loss: 0.344
[epoch: 6, i:  6499] avg mini-batch loss: 0.340
[epoch: 6, i:  6599] avg mini-batch loss: 0.330
[epoch: 6, i:  6699] avg mini-batch loss: 0.326
[epoch: 6, i:  6799] avg mini-batch loss: 0.322
[epoch: 6, i:  6899] avg mini-batch loss: 0.377
[epoch: 6, i:  6999] avg mini-batch loss: 0.376
[epoch: 6, i:  7099] avg mini-batch loss: 0.330
```

```
[epoch: 6, i:  7199] avg mini-batch loss: 0.317
[epoch: 6, i:  7299] avg mini-batch loss: 0.372
[epoch: 6, i:  7399] avg mini-batch loss: 0.403
[epoch: 6, i:  7499] avg mini-batch loss: 0.306
[epoch: 6, i:  7599] avg mini-batch loss: 0.300
[epoch: 6, i:  7699] avg mini-batch loss: 0.357
[epoch: 6, i:  7799] avg mini-batch loss: 0.352
[epoch: 6, i:  7899] avg mini-batch loss: 0.334
[epoch: 6, i:  7999] avg mini-batch loss: 0.407
[epoch: 6, i:  8099] avg mini-batch loss: 0.348
[epoch: 6, i:  8199] avg mini-batch loss: 0.398
[epoch: 6, i:  8299] avg mini-batch loss: 0.356
[epoch: 6, i:  8399] avg mini-batch loss: 0.320
[epoch: 6, i:  8499] avg mini-batch loss: 0.365
[epoch: 6, i:  8599] avg mini-batch loss: 0.386
[epoch: 6, i:  8699] avg mini-batch loss: 0.322
[epoch: 6, i:  8799] avg mini-batch loss: 0.324
[epoch: 6, i:  8899] avg mini-batch loss: 0.353
[epoch: 6, i:  8999] avg mini-batch loss: 0.321
[epoch: 6, i:  9099] avg mini-batch loss: 0.335
[epoch: 6, i:  9199] avg mini-batch loss: 0.367
[epoch: 6, i:  9299] avg mini-batch loss: 0.339
[epoch: 6, i:  9399] avg mini-batch loss: 0.373
[epoch: 6, i:  9499] avg mini-batch loss: 0.371
[epoch: 6, i:  9599] avg mini-batch loss: 0.389
[epoch: 6, i:  9699] avg mini-batch loss: 0.401
[epoch: 6, i:  9799] avg mini-batch loss: 0.353
[epoch: 6, i:  9899] avg mini-batch loss: 0.370
[epoch: 6, i:  9999] avg mini-batch loss: 0.313
[epoch: 6, i: 10099] avg mini-batch loss: 0.353
[epoch: 6, i: 10199] avg mini-batch loss: 0.393
[epoch: 6, i: 10299] avg mini-batch loss: 0.358
[epoch: 6, i: 10399] avg mini-batch loss: 0.361
[epoch: 6, i: 10499] avg mini-batch loss: 0.401
[epoch: 6, i: 10599] avg mini-batch loss: 0.424
[epoch: 6, i: 10699] avg mini-batch loss: 0.392
[epoch: 6, i: 10799] avg mini-batch loss: 0.307
[epoch: 6, i: 10899] avg mini-batch loss: 0.378
[epoch: 6, i: 10999] avg mini-batch loss: 0.290
[epoch: 6, i: 11099] avg mini-batch loss: 0.283
[epoch: 6, i: 11199] avg mini-batch loss: 0.357
[epoch: 6, i: 11299] avg mini-batch loss: 0.344
[epoch: 6, i: 11399] avg mini-batch loss: 0.339
[epoch: 6, i: 11499] avg mini-batch loss: 0.310
[epoch: 6, i: 11599] avg mini-batch loss: 0.347
[epoch: 6, i: 11699] avg mini-batch loss: 0.472
[epoch: 6, i: 11799] avg mini-batch loss: 0.351
[epoch: 6, i: 11899] avg mini-batch loss: 0.403
```

```
[epoch: 6, i: 11999] avg mini-batch loss: 0.256
[epoch: 6, i: 12099] avg mini-batch loss: 0.413
[epoch: 6, i: 12199] avg mini-batch loss: 0.408
[epoch: 6, i: 12299] avg mini-batch loss: 0.338
[epoch: 6, i: 12399] avg mini-batch loss: 0.362
[epoch: 6, i: 12499] avg mini-batch loss: 0.328
[epoch: 7, i:    99] avg mini-batch loss: 0.317
[epoch: 7, i:   199] avg mini-batch loss: 0.310
[epoch: 7, i:   299] avg mini-batch loss: 0.271
[epoch: 7, i:   399] avg mini-batch loss: 0.270
[epoch: 7, i:   499] avg mini-batch loss: 0.298
[epoch: 7, i:   599] avg mini-batch loss: 0.287
[epoch: 7, i:   699] avg mini-batch loss: 0.286
[epoch: 7, i:   799] avg mini-batch loss: 0.369
[epoch: 7, i:   899] avg mini-batch loss: 0.328
[epoch: 7, i:   999] avg mini-batch loss: 0.369
[epoch: 7, i:  1099] avg mini-batch loss: 0.263
[epoch: 7, i:  1199] avg mini-batch loss: 0.288
[epoch: 7, i:  1299] avg mini-batch loss: 0.332
[epoch: 7, i:  1399] avg mini-batch loss: 0.290
[epoch: 7, i:  1499] avg mini-batch loss: 0.298
[epoch: 7, i:  1599] avg mini-batch loss: 0.318
[epoch: 7, i:  1699] avg mini-batch loss: 0.342
[epoch: 7, i:  1799] avg mini-batch loss: 0.335
[epoch: 7, i:  1899] avg mini-batch loss: 0.379
[epoch: 7, i:  1999] avg mini-batch loss: 0.340
[epoch: 7, i:  2099] avg mini-batch loss: 0.254
[epoch: 7, i:  2199] avg mini-batch loss: 0.342
[epoch: 7, i:  2299] avg mini-batch loss: 0.293
[epoch: 7, i:  2399] avg mini-batch loss: 0.311
[epoch: 7, i:  2499] avg mini-batch loss: 0.285
[epoch: 7, i:  2599] avg mini-batch loss: 0.369
[epoch: 7, i:  2699] avg mini-batch loss: 0.254
[epoch: 7, i:  2799] avg mini-batch loss: 0.356
[epoch: 7, i:  2899] avg mini-batch loss: 0.248
[epoch: 7, i:  2999] avg mini-batch loss: 0.328
[epoch: 7, i:  3099] avg mini-batch loss: 0.342
[epoch: 7, i:  3199] avg mini-batch loss: 0.387
[epoch: 7, i:  3299] avg mini-batch loss: 0.350
[epoch: 7, i:  3399] avg mini-batch loss: 0.331
[epoch: 7, i:  3499] avg mini-batch loss: 0.356
[epoch: 7, i:  3599] avg mini-batch loss: 0.336
[epoch: 7, i:  3699] avg mini-batch loss: 0.317
[epoch: 7, i:  3799] avg mini-batch loss: 0.327
[epoch: 7, i:  3899] avg mini-batch loss: 0.355
[epoch: 7, i:  3999] avg mini-batch loss: 0.354
[epoch: 7, i:  4099] avg mini-batch loss: 0.366
[epoch: 7, i:  4199] avg mini-batch loss: 0.401
```

```
[epoch: 7, i:  4299] avg mini-batch loss: 0.327
[epoch: 7, i:  4399] avg mini-batch loss: 0.287
[epoch: 7, i:  4499] avg mini-batch loss: 0.340
[epoch: 7, i:  4599] avg mini-batch loss: 0.320
[epoch: 7, i:  4699] avg mini-batch loss: 0.363
[epoch: 7, i:  4799] avg mini-batch loss: 0.327
[epoch: 7, i:  4899] avg mini-batch loss: 0.338
[epoch: 7, i:  4999] avg mini-batch loss: 0.343
[epoch: 7, i:  5099] avg mini-batch loss: 0.356
[epoch: 7, i:  5199] avg mini-batch loss: 0.324
[epoch: 7, i:  5299] avg mini-batch loss: 0.339
[epoch: 7, i:  5399] avg mini-batch loss: 0.336
[epoch: 7, i:  5499] avg mini-batch loss: 0.320
[epoch: 7, i:  5599] avg mini-batch loss: 0.324
[epoch: 7, i:  5699] avg mini-batch loss: 0.344
[epoch: 7, i:  5799] avg mini-batch loss: 0.344
[epoch: 7, i:  5899] avg mini-batch loss: 0.382
[epoch: 7, i:  5999] avg mini-batch loss: 0.302
[epoch: 7, i:  6099] avg mini-batch loss: 0.362
[epoch: 7, i:  6199] avg mini-batch loss: 0.315
[epoch: 7, i:  6299] avg mini-batch loss: 0.245
[epoch: 7, i:  6399] avg mini-batch loss: 0.289
[epoch: 7, i:  6499] avg mini-batch loss: 0.299
[epoch: 7, i:  6599] avg mini-batch loss: 0.288
[epoch: 7, i:  6699] avg mini-batch loss: 0.305
[epoch: 7, i:  6799] avg mini-batch loss: 0.328
[epoch: 7, i:  6899] avg mini-batch loss: 0.343
[epoch: 7, i:  6999] avg mini-batch loss: 0.408
[epoch: 7, i:  7099] avg mini-batch loss: 0.349
[epoch: 7, i:  7199] avg mini-batch loss: 0.332
[epoch: 7, i:  7299] avg mini-batch loss: 0.339
[epoch: 7, i:  7399] avg mini-batch loss: 0.353
[epoch: 7, i:  7499] avg mini-batch loss: 0.347
[epoch: 7, i:  7599] avg mini-batch loss: 0.347
[epoch: 7, i:  7699] avg mini-batch loss: 0.329
[epoch: 7, i:  7799] avg mini-batch loss: 0.289
[epoch: 7, i:  7899] avg mini-batch loss: 0.359
[epoch: 7, i:  7999] avg mini-batch loss: 0.395
[epoch: 7, i:  8099] avg mini-batch loss: 0.312
[epoch: 7, i:  8199] avg mini-batch loss: 0.319
[epoch: 7, i:  8299] avg mini-batch loss: 0.319
[epoch: 7, i:  8399] avg mini-batch loss: 0.317
[epoch: 7, i:  8499] avg mini-batch loss: 0.328
[epoch: 7, i:  8599] avg mini-batch loss: 0.351
[epoch: 7, i:  8699] avg mini-batch loss: 0.357
[epoch: 7, i:  8799] avg mini-batch loss: 0.405
[epoch: 7, i:  8899] avg mini-batch loss: 0.309
[epoch: 7, i:  8999] avg mini-batch loss: 0.353
```

```
[epoch: 7, i:  9099] avg mini-batch loss: 0.322
[epoch: 7, i:  9199] avg mini-batch loss: 0.302
[epoch: 7, i:  9299] avg mini-batch loss: 0.387
[epoch: 7, i:  9399] avg mini-batch loss: 0.275
[epoch: 7, i:  9499] avg mini-batch loss: 0.298
[epoch: 7, i:  9599] avg mini-batch loss: 0.317
[epoch: 7, i:  9699] avg mini-batch loss: 0.280
[epoch: 7, i:  9799] avg mini-batch loss: 0.381
[epoch: 7, i:  9899] avg mini-batch loss: 0.316
[epoch: 7, i:  9999] avg mini-batch loss: 0.338
[epoch: 7, i: 10099] avg mini-batch loss: 0.347
[epoch: 7, i: 10199] avg mini-batch loss: 0.291
[epoch: 7, i: 10299] avg mini-batch loss: 0.330
[epoch: 7, i: 10399] avg mini-batch loss: 0.276
[epoch: 7, i: 10499] avg mini-batch loss: 0.311
[epoch: 7, i: 10599] avg mini-batch loss: 0.351
[epoch: 7, i: 10699] avg mini-batch loss: 0.305
[epoch: 7, i: 10799] avg mini-batch loss: 0.368
[epoch: 7, i: 10899] avg mini-batch loss: 0.344
[epoch: 7, i: 10999] avg mini-batch loss: 0.304
[epoch: 7, i: 11099] avg mini-batch loss: 0.314
[epoch: 7, i: 11199] avg mini-batch loss: 0.323
[epoch: 7, i: 11299] avg mini-batch loss: 0.392
[epoch: 7, i: 11399] avg mini-batch loss: 0.278
[epoch: 7, i: 11499] avg mini-batch loss: 0.286
[epoch: 7, i: 11599] avg mini-batch loss: 0.291
[epoch: 7, i: 11699] avg mini-batch loss: 0.403
[epoch: 7, i: 11799] avg mini-batch loss: 0.282
[epoch: 7, i: 11899] avg mini-batch loss: 0.315
[epoch: 7, i: 11999] avg mini-batch loss: 0.361
[epoch: 7, i: 12099] avg mini-batch loss: 0.383
[epoch: 7, i: 12199] avg mini-batch loss: 0.280
[epoch: 7, i: 12299] avg mini-batch loss: 0.351
[epoch: 7, i: 12399] avg mini-batch loss: 0.374
[epoch: 7, i: 12499] avg mini-batch loss: 0.349
[epoch: 8, i:    99] avg mini-batch loss: 0.247
[epoch: 8, i:   199] avg mini-batch loss: 0.308
[epoch: 8, i:   299] avg mini-batch loss: 0.339
[epoch: 8, i:   399] avg mini-batch loss: 0.276
[epoch: 8, i:   499] avg mini-batch loss: 0.316
[epoch: 8, i:   599] avg mini-batch loss: 0.304
[epoch: 8, i:   699] avg mini-batch loss: 0.260
[epoch: 8, i:   799] avg mini-batch loss: 0.379
[epoch: 8, i:   899] avg mini-batch loss: 0.337
[epoch: 8, i:   999] avg mini-batch loss: 0.258
[epoch: 8, i:  1099] avg mini-batch loss: 0.332
[epoch: 8, i:  1199] avg mini-batch loss: 0.246
[epoch: 8, i:  1299] avg mini-batch loss: 0.281
```

```
[epoch: 8, i:  1399] avg mini-batch loss: 0.248
[epoch: 8, i:  1499] avg mini-batch loss: 0.314
[epoch: 8, i:  1599] avg mini-batch loss: 0.301
[epoch: 8, i:  1699] avg mini-batch loss: 0.255
[epoch: 8, i:  1799] avg mini-batch loss: 0.290
[epoch: 8, i:  1899] avg mini-batch loss: 0.279
[epoch: 8, i:  1999] avg mini-batch loss: 0.246
[epoch: 8, i:  2099] avg mini-batch loss: 0.306
[epoch: 8, i:  2199] avg mini-batch loss: 0.288
[epoch: 8, i:  2299] avg mini-batch loss: 0.222
[epoch: 8, i:  2399] avg mini-batch loss: 0.329
[epoch: 8, i:  2499] avg mini-batch loss: 0.287
[epoch: 8, i:  2599] avg mini-batch loss: 0.297
[epoch: 8, i:  2699] avg mini-batch loss: 0.285
[epoch: 8, i:  2799] avg mini-batch loss: 0.276
[epoch: 8, i:  2899] avg mini-batch loss: 0.290
[epoch: 8, i:  2999] avg mini-batch loss: 0.258
[epoch: 8, i:  3099] avg mini-batch loss: 0.336
[epoch: 8, i:  3199] avg mini-batch loss: 0.290
[epoch: 8, i:  3299] avg mini-batch loss: 0.377
[epoch: 8, i:  3399] avg mini-batch loss: 0.313
[epoch: 8, i:  3499] avg mini-batch loss: 0.279
[epoch: 8, i:  3599] avg mini-batch loss: 0.355
[epoch: 8, i:  3699] avg mini-batch loss: 0.277
[epoch: 8, i:  3799] avg mini-batch loss: 0.294
[epoch: 8, i:  3899] avg mini-batch loss: 0.273
[epoch: 8, i:  3999] avg mini-batch loss: 0.354
[epoch: 8, i:  4099] avg mini-batch loss: 0.309
[epoch: 8, i:  4199] avg mini-batch loss: 0.258
[epoch: 8, i:  4299] avg mini-batch loss: 0.304
[epoch: 8, i:  4399] avg mini-batch loss: 0.266
[epoch: 8, i:  4499] avg mini-batch loss: 0.299
[epoch: 8, i:  4599] avg mini-batch loss: 0.287
[epoch: 8, i:  4699] avg mini-batch loss: 0.331
[epoch: 8, i:  4799] avg mini-batch loss: 0.345
[epoch: 8, i:  4899] avg mini-batch loss: 0.242
[epoch: 8, i:  4999] avg mini-batch loss: 0.271
[epoch: 8, i:  5099] avg mini-batch loss: 0.293
[epoch: 8, i:  5199] avg mini-batch loss: 0.384
[epoch: 8, i:  5299] avg mini-batch loss: 0.304
[epoch: 8, i:  5399] avg mini-batch loss: 0.283
[epoch: 8, i:  5499] avg mini-batch loss: 0.352
[epoch: 8, i:  5599] avg mini-batch loss: 0.266
[epoch: 8, i:  5699] avg mini-batch loss: 0.294
[epoch: 8, i:  5799] avg mini-batch loss: 0.280
[epoch: 8, i:  5899] avg mini-batch loss: 0.337
[epoch: 8, i:  5999] avg mini-batch loss: 0.321
[epoch: 8, i:  6099] avg mini-batch loss: 0.357
```

```
[epoch: 8, i:  6199] avg mini-batch loss: 0.320
[epoch: 8, i:  6299] avg mini-batch loss: 0.363
[epoch: 8, i:  6399] avg mini-batch loss: 0.318
[epoch: 8, i:  6499] avg mini-batch loss: 0.304
[epoch: 8, i:  6599] avg mini-batch loss: 0.302
[epoch: 8, i:  6699] avg mini-batch loss: 0.362
[epoch: 8, i:  6799] avg mini-batch loss: 0.322
[epoch: 8, i:  6899] avg mini-batch loss: 0.335
[epoch: 8, i:  6999] avg mini-batch loss: 0.333
[epoch: 8, i:  7099] avg mini-batch loss: 0.324
[epoch: 8, i:  7199] avg mini-batch loss: 0.325
[epoch: 8, i:  7299] avg mini-batch loss: 0.305
[epoch: 8, i:  7399] avg mini-batch loss: 0.387
[epoch: 8, i:  7499] avg mini-batch loss: 0.302
[epoch: 8, i:  7599] avg mini-batch loss: 0.303
[epoch: 8, i:  7699] avg mini-batch loss: 0.304
[epoch: 8, i:  7799] avg mini-batch loss: 0.390
[epoch: 8, i:  7899] avg mini-batch loss: 0.284
[epoch: 8, i:  7999] avg mini-batch loss: 0.367
[epoch: 8, i:  8099] avg mini-batch loss: 0.361
[epoch: 8, i:  8199] avg mini-batch loss: 0.299
[epoch: 8, i:  8299] avg mini-batch loss: 0.297
[epoch: 8, i:  8399] avg mini-batch loss: 0.265
[epoch: 8, i:  8499] avg mini-batch loss: 0.318
[epoch: 8, i:  8599] avg mini-batch loss: 0.325
[epoch: 8, i:  8699] avg mini-batch loss: 0.320
[epoch: 8, i:  8799] avg mini-batch loss: 0.255
[epoch: 8, i:  8899] avg mini-batch loss: 0.272
[epoch: 8, i:  8999] avg mini-batch loss: 0.328
[epoch: 8, i:  9099] avg mini-batch loss: 0.305
[epoch: 8, i:  9199] avg mini-batch loss: 0.349
[epoch: 8, i:  9299] avg mini-batch loss: 0.313
[epoch: 8, i:  9399] avg mini-batch loss: 0.270
[epoch: 8, i:  9499] avg mini-batch loss: 0.312
[epoch: 8, i:  9599] avg mini-batch loss: 0.293
[epoch: 8, i:  9699] avg mini-batch loss: 0.308
[epoch: 8, i:  9799] avg mini-batch loss: 0.299
[epoch: 8, i:  9899] avg mini-batch loss: 0.354
[epoch: 8, i:  9999] avg mini-batch loss: 0.317
[epoch: 8, i: 10099] avg mini-batch loss: 0.337
[epoch: 8, i: 10199] avg mini-batch loss: 0.373
[epoch: 8, i: 10299] avg mini-batch loss: 0.316
[epoch: 8, i: 10399] avg mini-batch loss: 0.373
[epoch: 8, i: 10499] avg mini-batch loss: 0.260
[epoch: 8, i: 10599] avg mini-batch loss: 0.349
[epoch: 8, i: 10699] avg mini-batch loss: 0.331
[epoch: 8, i: 10799] avg mini-batch loss: 0.366
[epoch: 8, i: 10899] avg mini-batch loss: 0.264
```

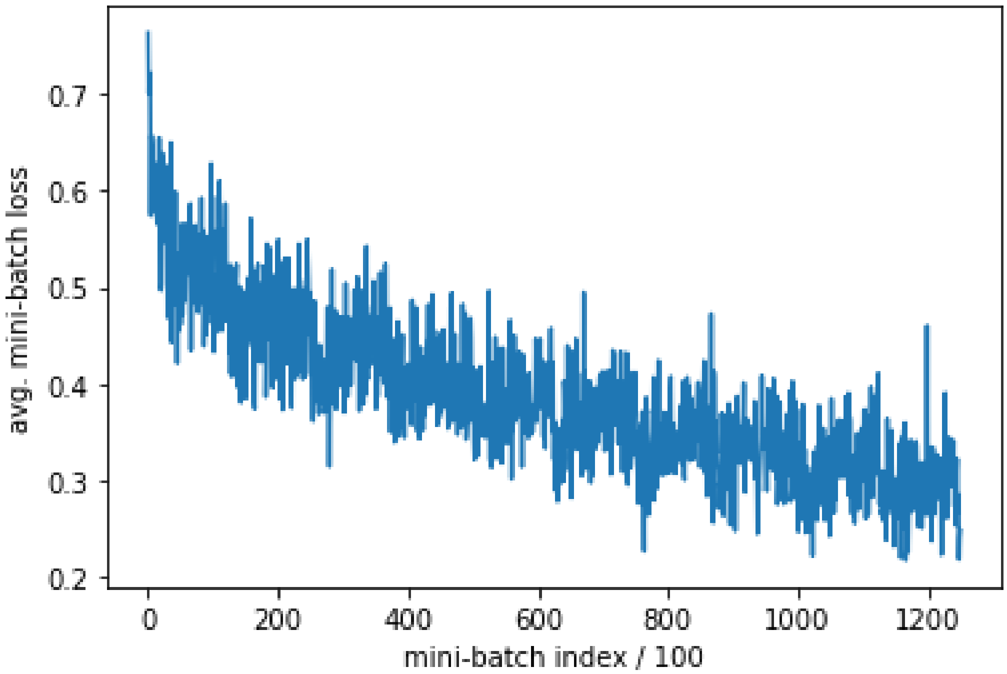
```
[epoch: 8, i: 10999] avg mini-batch loss: 0.309
[epoch: 8, i: 11099] avg mini-batch loss: 0.294
[epoch: 8, i: 11199] avg mini-batch loss: 0.307
[epoch: 8, i: 11299] avg mini-batch loss: 0.397
[epoch: 8, i: 11399] avg mini-batch loss: 0.304
[epoch: 8, i: 11499] avg mini-batch loss: 0.282
[epoch: 8, i: 11599] avg mini-batch loss: 0.351
[epoch: 8, i: 11699] avg mini-batch loss: 0.309
[epoch: 8, i: 11799] avg mini-batch loss: 0.349
[epoch: 8, i: 11899] avg mini-batch loss: 0.388
[epoch: 8, i: 11999] avg mini-batch loss: 0.311
[epoch: 8, i: 12099] avg mini-batch loss: 0.306
[epoch: 8, i: 12199] avg mini-batch loss: 0.327
[epoch: 8, i: 12299] avg mini-batch loss: 0.330
[epoch: 8, i: 12399] avg mini-batch loss: 0.411
[epoch: 8, i: 12499] avg mini-batch loss: 0.316
[epoch: 9, i:    99] avg mini-batch loss: 0.275
[epoch: 9, i:   199] avg mini-batch loss: 0.310
[epoch: 9, i:   299] avg mini-batch loss: 0.303
[epoch: 9, i:   399] avg mini-batch loss: 0.264
[epoch: 9, i:   499] avg mini-batch loss: 0.270
[epoch: 9, i:   599] avg mini-batch loss: 0.259
[epoch: 9, i:   699] avg mini-batch loss: 0.285
[epoch: 9, i:   799] avg mini-batch loss: 0.301
[epoch: 9, i:   899] avg mini-batch loss: 0.265
[epoch: 9, i:   999] avg mini-batch loss: 0.309
[epoch: 9, i:  1099] avg mini-batch loss: 0.238
[epoch: 9, i:  1199] avg mini-batch loss: 0.241
[epoch: 9, i:  1299] avg mini-batch loss: 0.364
[epoch: 9, i:  1399] avg mini-batch loss: 0.294
[epoch: 9, i:  1499] avg mini-batch loss: 0.287
[epoch: 9, i:  1599] avg mini-batch loss: 0.321
[epoch: 9, i:  1699] avg mini-batch loss: 0.271
[epoch: 9, i:  1799] avg mini-batch loss: 0.326
[epoch: 9, i:  1899] avg mini-batch loss: 0.307
[epoch: 9, i:  1999] avg mini-batch loss: 0.352
[epoch: 9, i:  2099] avg mini-batch loss: 0.264
[epoch: 9, i:  2199] avg mini-batch loss: 0.279
[epoch: 9, i:  2299] avg mini-batch loss: 0.232
[epoch: 9, i:  2399] avg mini-batch loss: 0.269
[epoch: 9, i:  2499] avg mini-batch loss: 0.283
[epoch: 9, i:  2599] avg mini-batch loss: 0.302
[epoch: 9, i:  2699] avg mini-batch loss: 0.298
[epoch: 9, i:  2799] avg mini-batch loss: 0.284
[epoch: 9, i:  2899] avg mini-batch loss: 0.261
[epoch: 9, i:  2999] avg mini-batch loss: 0.309
[epoch: 9, i:  3099] avg mini-batch loss: 0.338
[epoch: 9, i:  3199] avg mini-batch loss: 0.293
```

```
[epoch: 9, i:  3299] avg mini-batch loss: 0.220
[epoch: 9, i:  3399] avg mini-batch loss: 0.268
[epoch: 9, i:  3499] avg mini-batch loss: 0.264
[epoch: 9, i:  3599] avg mini-batch loss: 0.345
[epoch: 9, i:  3699] avg mini-batch loss: 0.249
[epoch: 9, i:  3799] avg mini-batch loss: 0.351
[epoch: 9, i:  3899] avg mini-batch loss: 0.348
[epoch: 9, i:  3999] avg mini-batch loss: 0.359
[epoch: 9, i:  4099] avg mini-batch loss: 0.217
[epoch: 9, i:  4199] avg mini-batch loss: 0.230
[epoch: 9, i:  4299] avg mini-batch loss: 0.306
[epoch: 9, i:  4399] avg mini-batch loss: 0.274
[epoch: 9, i:  4499] avg mini-batch loss: 0.226
[epoch: 9, i:  4599] avg mini-batch loss: 0.266
[epoch: 9, i:  4699] avg mini-batch loss: 0.296
[epoch: 9, i:  4799] avg mini-batch loss: 0.342
[epoch: 9, i:  4899] avg mini-batch loss: 0.341
[epoch: 9, i:  4999] avg mini-batch loss: 0.267
[epoch: 9, i:  5099] avg mini-batch loss: 0.296
[epoch: 9, i:  5199] avg mini-batch loss: 0.295
[epoch: 9, i:  5299] avg mini-batch loss: 0.270
[epoch: 9, i:  5399] avg mini-batch loss: 0.292
[epoch: 9, i:  5499] avg mini-batch loss: 0.334
[epoch: 9, i:  5599] avg mini-batch loss: 0.326
[epoch: 9, i:  5699] avg mini-batch loss: 0.340
[epoch: 9, i:  5799] avg mini-batch loss: 0.336
[epoch: 9, i:  5899] avg mini-batch loss: 0.306
[epoch: 9, i:  5999] avg mini-batch loss: 0.318
[epoch: 9, i:  6099] avg mini-batch loss: 0.252
[epoch: 9, i:  6199] avg mini-batch loss: 0.311
[epoch: 9, i:  6299] avg mini-batch loss: 0.277
[epoch: 9, i:  6399] avg mini-batch loss: 0.296
[epoch: 9, i:  6499] avg mini-batch loss: 0.305
[epoch: 9, i:  6599] avg mini-batch loss: 0.251
[epoch: 9, i:  6699] avg mini-batch loss: 0.307
[epoch: 9, i:  6799] avg mini-batch loss: 0.318
[epoch: 9, i:  6899] avg mini-batch loss: 0.300
[epoch: 9, i:  6999] avg mini-batch loss: 0.290
[epoch: 9, i:  7099] avg mini-batch loss: 0.307
[epoch: 9, i:  7199] avg mini-batch loss: 0.276
[epoch: 9, i:  7299] avg mini-batch loss: 0.264
[epoch: 9, i:  7399] avg mini-batch loss: 0.460
[epoch: 9, i:  7499] avg mini-batch loss: 0.275
[epoch: 9, i:  7599] avg mini-batch loss: 0.275
[epoch: 9, i:  7699] avg mini-batch loss: 0.334
[epoch: 9, i:  7799] avg mini-batch loss: 0.270
[epoch: 9, i:  7899] avg mini-batch loss: 0.311
[epoch: 9, i:  7999] avg mini-batch loss: 0.261
```

```
[epoch: 9, i:  8099] avg mini-batch loss: 0.237
[epoch: 9, i:  8199] avg mini-batch loss: 0.301
[epoch: 9, i:  8299] avg mini-batch loss: 0.334
[epoch: 9, i:  8399] avg mini-batch loss: 0.309
[epoch: 9, i:  8499] avg mini-batch loss: 0.287
[epoch: 9, i:  8599] avg mini-batch loss: 0.265
[epoch: 9, i:  8699] avg mini-batch loss: 0.313
[epoch: 9, i:  8799] avg mini-batch loss: 0.317
[epoch: 9, i:  8899] avg mini-batch loss: 0.326
[epoch: 9, i:  8999] avg mini-batch loss: 0.320
[epoch: 9, i:  9099] avg mini-batch loss: 0.281
[epoch: 9, i:  9199] avg mini-batch loss: 0.308
[epoch: 9, i:  9299] avg mini-batch loss: 0.323
[epoch: 9, i:  9399] avg mini-batch loss: 0.310
[epoch: 9, i:  9499] avg mini-batch loss: 0.299
[epoch: 9, i:  9599] avg mini-batch loss: 0.223
[epoch: 9, i:  9699] avg mini-batch loss: 0.302
[epoch: 9, i:  9799] avg mini-batch loss: 0.267
[epoch: 9, i:  9899] avg mini-batch loss: 0.308
[epoch: 9, i:  9999] avg mini-batch loss: 0.262
[epoch: 9, i: 10099] avg mini-batch loss: 0.293
[epoch: 9, i: 10199] avg mini-batch loss: 0.391
[epoch: 9, i: 10299] avg mini-batch loss: 0.312
[epoch: 9, i: 10399] avg mini-batch loss: 0.274
[epoch: 9, i: 10499] avg mini-batch loss: 0.261
[epoch: 9, i: 10599] avg mini-batch loss: 0.295
[epoch: 9, i: 10699] avg mini-batch loss: 0.309
[epoch: 9, i: 10799] avg mini-batch loss: 0.344
[epoch: 9, i: 10899] avg mini-batch loss: 0.319
[epoch: 9, i: 10999] avg mini-batch loss: 0.294
[epoch: 9, i: 11099] avg mini-batch loss: 0.341
[epoch: 9, i: 11199] avg mini-batch loss: 0.328
[epoch: 9, i: 11299] avg mini-batch loss: 0.314
[epoch: 9, i: 11399] avg mini-batch loss: 0.312
[epoch: 9, i: 11499] avg mini-batch loss: 0.341
[epoch: 9, i: 11599] avg mini-batch loss: 0.273
[epoch: 9, i: 11699] avg mini-batch loss: 0.323
[epoch: 9, i: 11799] avg mini-batch loss: 0.254
[epoch: 9, i: 11899] avg mini-batch loss: 0.309
[epoch: 9, i: 11999] avg mini-batch loss: 0.322
[epoch: 9, i: 12099] avg mini-batch loss: 0.266
[epoch: 9, i: 12199] avg mini-batch loss: 0.285
[epoch: 9, i: 12299] avg mini-batch loss: 0.218
[epoch: 9, i: 12399] avg mini-batch loss: 0.241
[epoch: 9, i: 12499] avg mini-batch loss: 0.248
Finished Training.
```

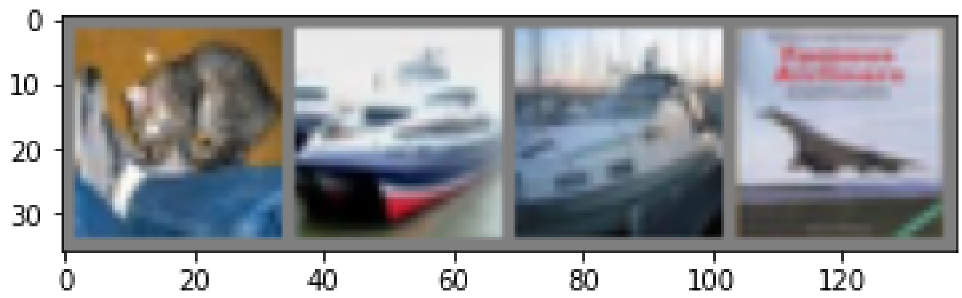
Training Loss Curve

[17]:
```
plt.plot(avg_losses)
plt.xlabel('mini-batch index / {}'.format(print_freq))
plt.ylabel('avg. mini-batch loss')
plt.show()
```



Evaluate on Test Dataset

[18]:
```
# Check several images.
dataiter = iter(testloader)
images, labels = next(dataiter)
imshow(torchvision.utils.make_grid(images))
print('GroundTruth: ', ' '.join('%5s' % classes[labels[j]] for j in range(4)))
outputs = net(images.to(device))
_, predicted = torch.max(outputs, 1)

print('Predicted: ', ' '.join('%5s' % classes[predicted[j]]
                              for j in range(4)))
```



```
GroundTruth:    cat  ship  ship plane
Predicted:    dog  ship  ship plane
```

[19]:

```python
# Get test accuracy.
correct = 0
total = 0
with torch.no_grad():
    for data in testloader:
        images, labels = data
        images, labels = images.to(device), labels.to(device)
        outputs = net(images)
        _, predicted = torch.max(outputs.data, 1)
        total += labels.size(0)
        correct += (predicted == labels).sum().item()

print('Accuracy of the network on the 10000 test images: %d %%' % (
    100 * correct / total))
```

```
Accuracy of the network on the 10000 test images: 70 %
```

[20]:

```python
# Get test accuracy for each class.
class_correct = list(0. for i in range(10))
class_total = list(0. for i in range(10))
with torch.no_grad():
    for data in testloader:
        images, labels = data
        images, labels = images.to(device), labels.to(device)
        outputs = net(images)
        _, predicted = torch.max(outputs, 1)
        c = (predicted == labels).squeeze()
        for i in range(4):
            label = labels[i]
            class_correct[label] += c[i].item()
            class_total[label] += 1
```

```
for i in range(10):
    print('Accuracy of %5s : %2d %%' % (
        classes[i], 100 * class_correct[i] / class_total[i]))
```

```
Accuracy of plane : 77 %
Accuracy of   car : 82 %
Accuracy of  bird : 57 %
Accuracy of   cat : 49 %
Accuracy of  deer : 64 %
Accuracy of   dog : 62 %
Accuracy of  frog : 79 %
Accuracy of horse : 75 %
Accuracy of  ship : 80 %
Accuracy of truck : 80 %
```

[ ]:

```
# One of the changes I made was that I added another layer in the network
# that takes the output from the second convolutional layer, applies ReLU␣
↪activation,
# and passes it through a 2x2 AvgPool layer to capture more relationships.
# I also reduced the learning rate of the optimizer to half (0.0005) which
# which allowed the optimizer to take smaller steps towards the minimum of
# the loss function which might allow for a better chance of finding the global␣
↪min of loss.
```

hw4

March 3, 2023

[1]:
```
%matplotlib inline
import matplotlib.pyplot as plt
import numpy as np
import torch
import torchvision
import torchvision.transforms as transforms
import torch.nn as nn
import torch.nn.functional as F
import torch.optim as optim
```

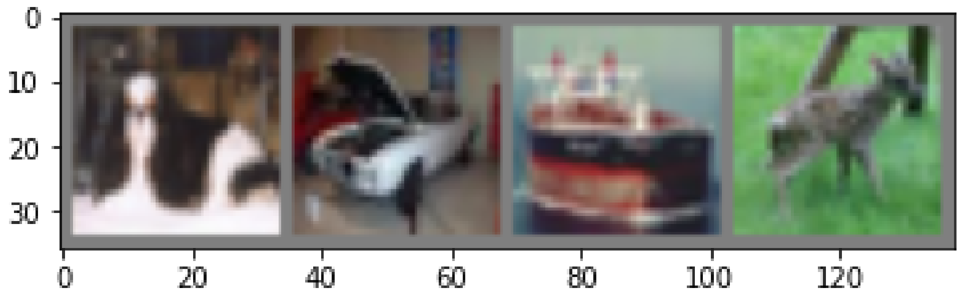
Prepare for Dataset

[2]:
```
transform = transforms.Compose(
    [transforms.ToTensor(),
     transforms.Normalize((0.5, 0.5, 0.5), (0.5, 0.5, 0.5))])

trainset = torchvision.datasets.CIFAR10(root='./data', train=True,
                                        download=True, transform=transform)
trainloader = torch.utils.data.DataLoader(trainset, batch_size=4,
                                          shuffle=True, num_workers=2)

testset = torchvision.datasets.CIFAR10(root='./data', train=False,
                                       download=True, transform=transform)
testloader = torch.utils.data.DataLoader(testset, batch_size=4,
                                         shuffle=False, num_workers=2)

classes = ('plane', 'car', 'bird', 'cat',
           'deer', 'dog', 'frog', 'horse', 'ship', 'truck')
```

```
Downloading https://www.cs.toronto.edu/~kriz/cifar-10-python.tar.gz to
./data/cifar-10-python.tar.gz

  0%|          | 0/170498071 [00:00<?, ?it/s]

Extracting ./data/cifar-10-python.tar.gz to ./data
Files already downloaded and verified
```

[3]:
```
# The function to show an image.
def imshow(img):
```

```python
    img = img / 2 + 0.5     # Unnormalize.
    npimg = img.numpy()
    plt.imshow(np.transpose(npimg, (1, 2, 0)))
    plt.show()

# Get some random training images.
dataiter = iter(trainloader)
images, labels = next(dataiter)
# Show images.
imshow(torchvision.utils.make_grid(images))
# Print labels.
print(' '.join('%5s' % classes[labels[j]] for j in range(4)))
```




```
  dog   car  ship  deer
```

Choose a Device

[4]:
```python
# If there are GPUs, choose the first one for computing. Otherwise use CPU.
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
print(device)
# If 'cuda:0' is printed, it means GPU is available.
```

```
cpu
```

Network Definition

[5]:
```python
class Net(nn.Module):
    def __init__(self):
        super(Net, self).__init__()
        ###### Fill the blank here ######
        self.conv1 = nn.Conv2d(3, 10, kernel_size=3, padding=1)

        self.relu1 = nn.ReLU()

        self.pool1 = nn.AvgPool2d(kernel_size=2, stride=2)

        self.conv2 = nn.Conv2d(10, 20, kernel_size=3, padding=1)
```

```python
        self.relu2 = nn.ReLU()

        self.pool2 = nn.AvgPool2d(kernel_size=2, stride=2)

        self.fc1 = nn.Linear(20 * 8 * 8, 100)

        self.relu3 = nn.ReLU()

        self.fc2 = nn.Linear(100, 10)


    def forward(self, x):
        ###### Fill the blank here ######
        x = self.pool1(self.relu1(self.conv1(x)))

        x = self.pool2(self.relu2(self.conv2(x)))

        x = x.view(-1, 20 * 8 * 8)

        x = self.relu3(self.fc1(x))

        x = self.fc2(x)

        return x



net = Net()      # Create the network instance.
net.to(device)  # Move the network parameters to the specified device.
```

[5]:
```
Net(
  (conv1): Conv2d(3, 10, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
  (relu1): ReLU()
  (pool1): AvgPool2d(kernel_size=2, stride=2, padding=0)
  (conv2): Conv2d(10, 20, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
  (relu2): ReLU()
  (pool2): AvgPool2d(kernel_size=2, stride=2, padding=0)
  (fc1): Linear(in_features=1280, out_features=100, bias=True)
  (relu3): ReLU()
  (fc2): Linear(in_features=100, out_features=10, bias=True)
)
```

Optimizer and Loss Function

[6]:
```python
# We use cross-entropy as loss function.
loss_func = nn.CrossEntropyLoss()
```

```python
# We use stochastic gradient descent (SGD) as optimizer.
opt = optim.SGD(net.parameters(), lr=0.001, momentum=0.9)
```

Training Procedure

[7]:
```python
avg_losses = []   # Avg. losses.
epochs = 10       # Total epochs.
print_freq = 100  # Print frequency.

for epoch in range(epochs):  # Loop over the dataset multiple times.
    running_loss = 0.0       # Initialize running loss.
    for i, data in enumerate(trainloader, 0):
        # Get the inputs.
        inputs, labels = data

        # Move the inputs to the specified device.
        inputs, labels = inputs.to(device), labels.to(device)

        # Zero the parameter gradients.
        opt.zero_grad()

        # Forward step.
        outputs = net(inputs)
        loss = loss_func(outputs, labels)

        # Backward step.
        loss.backward()

        # Optimization step (update the parameters).
        opt.step()

        # Print statistics.
        running_loss += loss.item()
        if i % print_freq == print_freq - 1: # Print every several mini-batches.
            avg_loss = running_loss / print_freq
            print('[epoch: {}, i: {:5d}] avg mini-batch loss: {:.3f}'.format(
                epoch, i, avg_loss))
            avg_losses.append(avg_loss)
            running_loss = 0.0

print('Finished Training.')
```

```
[epoch: 0, i:    99] avg mini-batch loss: 2.298
[epoch: 0, i:   199] avg mini-batch loss: 2.304
[epoch: 0, i:   299] avg mini-batch loss: 2.301
[epoch: 0, i:   399] avg mini-batch loss: 2.295
[epoch: 0, i:   499] avg mini-batch loss: 2.289
[epoch: 0, i:   599] avg mini-batch loss: 2.278
```

```
[epoch: 0, i:   699] avg mini-batch loss: 2.263
[epoch: 0, i:   799] avg mini-batch loss: 2.228
[epoch: 0, i:   899] avg mini-batch loss: 2.161
[epoch: 0, i:   999] avg mini-batch loss: 2.119
[epoch: 0, i:  1099] avg mini-batch loss: 2.070
[epoch: 0, i:  1199] avg mini-batch loss: 2.027
[epoch: 0, i:  1299] avg mini-batch loss: 2.004
[epoch: 0, i:  1399] avg mini-batch loss: 2.072
[epoch: 0, i:  1499] avg mini-batch loss: 2.025
[epoch: 0, i:  1599] avg mini-batch loss: 1.985
[epoch: 0, i:  1699] avg mini-batch loss: 1.937
[epoch: 0, i:  1799] avg mini-batch loss: 1.979
[epoch: 0, i:  1899] avg mini-batch loss: 1.878
[epoch: 0, i:  1999] avg mini-batch loss: 1.843
[epoch: 0, i:  2099] avg mini-batch loss: 1.884
[epoch: 0, i:  2199] avg mini-batch loss: 1.903
[epoch: 0, i:  2299] avg mini-batch loss: 1.871
[epoch: 0, i:  2399] avg mini-batch loss: 1.790
[epoch: 0, i:  2499] avg mini-batch loss: 1.906
[epoch: 0, i:  2599] avg mini-batch loss: 1.847
[epoch: 0, i:  2699] avg mini-batch loss: 1.795
[epoch: 0, i:  2799] avg mini-batch loss: 1.755
[epoch: 0, i:  2899] avg mini-batch loss: 1.823
[epoch: 0, i:  2999] avg mini-batch loss: 1.799
[epoch: 0, i:  3099] avg mini-batch loss: 1.786
[epoch: 0, i:  3199] avg mini-batch loss: 1.812
[epoch: 0, i:  3299] avg mini-batch loss: 1.775
[epoch: 0, i:  3399] avg mini-batch loss: 1.742
[epoch: 0, i:  3499] avg mini-batch loss: 1.831
[epoch: 0, i:  3599] avg mini-batch loss: 1.781
[epoch: 0, i:  3699] avg mini-batch loss: 1.752
[epoch: 0, i:  3799] avg mini-batch loss: 1.743
[epoch: 0, i:  3899] avg mini-batch loss: 1.663
[epoch: 0, i:  3999] avg mini-batch loss: 1.764
[epoch: 0, i:  4099] avg mini-batch loss: 1.669
[epoch: 0, i:  4199] avg mini-batch loss: 1.682
[epoch: 0, i:  4299] avg mini-batch loss: 1.604
[epoch: 0, i:  4399] avg mini-batch loss: 1.641
[epoch: 0, i:  4499] avg mini-batch loss: 1.621
[epoch: 0, i:  4599] avg mini-batch loss: 1.620
[epoch: 0, i:  4699] avg mini-batch loss: 1.671
[epoch: 0, i:  4799] avg mini-batch loss: 1.639
[epoch: 0, i:  4899] avg mini-batch loss: 1.648
[epoch: 0, i:  4999] avg mini-batch loss: 1.630
[epoch: 0, i:  5099] avg mini-batch loss: 1.693
[epoch: 0, i:  5199] avg mini-batch loss: 1.582
[epoch: 0, i:  5299] avg mini-batch loss: 1.614
[epoch: 0, i:  5399] avg mini-batch loss: 1.594
```

```
[epoch: 0, i:  5499] avg mini-batch loss: 1.545
[epoch: 0, i:  5599] avg mini-batch loss: 1.573
[epoch: 0, i:  5699] avg mini-batch loss: 1.575
[epoch: 0, i:  5799] avg mini-batch loss: 1.598
[epoch: 0, i:  5899] avg mini-batch loss: 1.567
[epoch: 0, i:  5999] avg mini-batch loss: 1.603
[epoch: 0, i:  6099] avg mini-batch loss: 1.563
[epoch: 0, i:  6199] avg mini-batch loss: 1.535
[epoch: 0, i:  6299] avg mini-batch loss: 1.552
[epoch: 0, i:  6399] avg mini-batch loss: 1.560
[epoch: 0, i:  6499] avg mini-batch loss: 1.596
[epoch: 0, i:  6599] avg mini-batch loss: 1.589
[epoch: 0, i:  6699] avg mini-batch loss: 1.486
[epoch: 0, i:  6799] avg mini-batch loss: 1.575
[epoch: 0, i:  6899] avg mini-batch loss: 1.591
[epoch: 0, i:  6999] avg mini-batch loss: 1.580
[epoch: 0, i:  7099] avg mini-batch loss: 1.531
[epoch: 0, i:  7199] avg mini-batch loss: 1.488
[epoch: 0, i:  7299] avg mini-batch loss: 1.582
[epoch: 0, i:  7399] avg mini-batch loss: 1.482
[epoch: 0, i:  7499] avg mini-batch loss: 1.482
[epoch: 0, i:  7599] avg mini-batch loss: 1.468
[epoch: 0, i:  7699] avg mini-batch loss: 1.541
[epoch: 0, i:  7799] avg mini-batch loss: 1.566
[epoch: 0, i:  7899] avg mini-batch loss: 1.532
[epoch: 0, i:  7999] avg mini-batch loss: 1.427
[epoch: 0, i:  8099] avg mini-batch loss: 1.538
[epoch: 0, i:  8199] avg mini-batch loss: 1.465
[epoch: 0, i:  8299] avg mini-batch loss: 1.486
[epoch: 0, i:  8399] avg mini-batch loss: 1.530
[epoch: 0, i:  8499] avg mini-batch loss: 1.476
[epoch: 0, i:  8599] avg mini-batch loss: 1.495
[epoch: 0, i:  8699] avg mini-batch loss: 1.453
[epoch: 0, i:  8799] avg mini-batch loss: 1.506
[epoch: 0, i:  8899] avg mini-batch loss: 1.441
[epoch: 0, i:  8999] avg mini-batch loss: 1.350
[epoch: 0, i:  9099] avg mini-batch loss: 1.513
[epoch: 0, i:  9199] avg mini-batch loss: 1.458
[epoch: 0, i:  9299] avg mini-batch loss: 1.451
[epoch: 0, i:  9399] avg mini-batch loss: 1.451
[epoch: 0, i:  9499] avg mini-batch loss: 1.463
[epoch: 0, i:  9599] avg mini-batch loss: 1.437
[epoch: 0, i:  9699] avg mini-batch loss: 1.434
[epoch: 0, i:  9799] avg mini-batch loss: 1.443
[epoch: 0, i:  9899] avg mini-batch loss: 1.466
[epoch: 0, i:  9999] avg mini-batch loss: 1.411
[epoch: 0, i: 10099] avg mini-batch loss: 1.454
[epoch: 0, i: 10199] avg mini-batch loss: 1.460
```

```
[epoch: 0, i: 10299] avg mini-batch loss: 1.435
[epoch: 0, i: 10399] avg mini-batch loss: 1.398
[epoch: 0, i: 10499] avg mini-batch loss: 1.415
[epoch: 0, i: 10599] avg mini-batch loss: 1.395
[epoch: 0, i: 10699] avg mini-batch loss: 1.371
[epoch: 0, i: 10799] avg mini-batch loss: 1.346
[epoch: 0, i: 10899] avg mini-batch loss: 1.438
[epoch: 0, i: 10999] avg mini-batch loss: 1.341
[epoch: 0, i: 11099] avg mini-batch loss: 1.387
[epoch: 0, i: 11199] avg mini-batch loss: 1.409
[epoch: 0, i: 11299] avg mini-batch loss: 1.406
[epoch: 0, i: 11399] avg mini-batch loss: 1.378
[epoch: 0, i: 11499] avg mini-batch loss: 1.402
[epoch: 0, i: 11599] avg mini-batch loss: 1.452
[epoch: 0, i: 11699] avg mini-batch loss: 1.362
[epoch: 0, i: 11799] avg mini-batch loss: 1.352
[epoch: 0, i: 11899] avg mini-batch loss: 1.349
[epoch: 0, i: 11999] avg mini-batch loss: 1.317
[epoch: 0, i: 12099] avg mini-batch loss: 1.336
[epoch: 0, i: 12199] avg mini-batch loss: 1.329
[epoch: 0, i: 12299] avg mini-batch loss: 1.383
[epoch: 0, i: 12399] avg mini-batch loss: 1.351
[epoch: 0, i: 12499] avg mini-batch loss: 1.374
[epoch: 1, i:    99] avg mini-batch loss: 1.297
[epoch: 1, i:   199] avg mini-batch loss: 1.386
[epoch: 1, i:   299] avg mini-batch loss: 1.300
[epoch: 1, i:   399] avg mini-batch loss: 1.334
[epoch: 1, i:   499] avg mini-batch loss: 1.347
[epoch: 1, i:   599] avg mini-batch loss: 1.278
[epoch: 1, i:   699] avg mini-batch loss: 1.275
[epoch: 1, i:   799] avg mini-batch loss: 1.313
[epoch: 1, i:   899] avg mini-batch loss: 1.382
[epoch: 1, i:   999] avg mini-batch loss: 1.324
[epoch: 1, i:  1099] avg mini-batch loss: 1.267
[epoch: 1, i:  1199] avg mini-batch loss: 1.384
[epoch: 1, i:  1299] avg mini-batch loss: 1.386
[epoch: 1, i:  1399] avg mini-batch loss: 1.259
[epoch: 1, i:  1499] avg mini-batch loss: 1.340
[epoch: 1, i:  1599] avg mini-batch loss: 1.292
[epoch: 1, i:  1699] avg mini-batch loss: 1.254
[epoch: 1, i:  1799] avg mini-batch loss: 1.376
[epoch: 1, i:  1899] avg mini-batch loss: 1.261
[epoch: 1, i:  1999] avg mini-batch loss: 1.293
[epoch: 1, i:  2099] avg mini-batch loss: 1.369
[epoch: 1, i:  2199] avg mini-batch loss: 1.308
[epoch: 1, i:  2299] avg mini-batch loss: 1.295
[epoch: 1, i:  2399] avg mini-batch loss: 1.282
[epoch: 1, i:  2499] avg mini-batch loss: 1.286
```

```
[epoch: 1, i:  2599] avg mini-batch loss: 1.389
[epoch: 1, i:  2699] avg mini-batch loss: 1.235
[epoch: 1, i:  2799] avg mini-batch loss: 1.294
[epoch: 1, i:  2899] avg mini-batch loss: 1.305
[epoch: 1, i:  2999] avg mini-batch loss: 1.352
[epoch: 1, i:  3099] avg mini-batch loss: 1.336
[epoch: 1, i:  3199] avg mini-batch loss: 1.243
[epoch: 1, i:  3299] avg mini-batch loss: 1.369
[epoch: 1, i:  3399] avg mini-batch loss: 1.207
[epoch: 1, i:  3499] avg mini-batch loss: 1.394
[epoch: 1, i:  3599] avg mini-batch loss: 1.218
[epoch: 1, i:  3699] avg mini-batch loss: 1.391
[epoch: 1, i:  3799] avg mini-batch loss: 1.261
[epoch: 1, i:  3899] avg mini-batch loss: 1.254
[epoch: 1, i:  3999] avg mini-batch loss: 1.249
[epoch: 1, i:  4099] avg mini-batch loss: 1.288
[epoch: 1, i:  4199] avg mini-batch loss: 1.236
[epoch: 1, i:  4299] avg mini-batch loss: 1.343
[epoch: 1, i:  4399] avg mini-batch loss: 1.243
[epoch: 1, i:  4499] avg mini-batch loss: 1.255
[epoch: 1, i:  4599] avg mini-batch loss: 1.235
[epoch: 1, i:  4699] avg mini-batch loss: 1.236
[epoch: 1, i:  4799] avg mini-batch loss: 1.416
[epoch: 1, i:  4899] avg mini-batch loss: 1.247
[epoch: 1, i:  4999] avg mini-batch loss: 1.213
[epoch: 1, i:  5099] avg mini-batch loss: 1.263
[epoch: 1, i:  5199] avg mini-batch loss: 1.310
[epoch: 1, i:  5299] avg mini-batch loss: 1.236
[epoch: 1, i:  5399] avg mini-batch loss: 1.390
[epoch: 1, i:  5499] avg mini-batch loss: 1.232
[epoch: 1, i:  5599] avg mini-batch loss: 1.216
[epoch: 1, i:  5699] avg mini-batch loss: 1.213
[epoch: 1, i:  5799] avg mini-batch loss: 1.186
[epoch: 1, i:  5899] avg mini-batch loss: 1.180
[epoch: 1, i:  5999] avg mini-batch loss: 1.290
[epoch: 1, i:  6099] avg mini-batch loss: 1.330
[epoch: 1, i:  6199] avg mini-batch loss: 1.199
[epoch: 1, i:  6299] avg mini-batch loss: 1.284
[epoch: 1, i:  6399] avg mini-batch loss: 1.225
[epoch: 1, i:  6499] avg mini-batch loss: 1.259
[epoch: 1, i:  6599] avg mini-batch loss: 1.227
[epoch: 1, i:  6699] avg mini-batch loss: 1.301
[epoch: 1, i:  6799] avg mini-batch loss: 1.213
[epoch: 1, i:  6899] avg mini-batch loss: 1.368
[epoch: 1, i:  6999] avg mini-batch loss: 1.217
[epoch: 1, i:  7099] avg mini-batch loss: 1.269
[epoch: 1, i:  7199] avg mini-batch loss: 1.221
[epoch: 1, i:  7299] avg mini-batch loss: 1.298
```

```
[epoch: 1, i:  7399] avg mini-batch loss: 1.209
[epoch: 1, i:  7499] avg mini-batch loss: 1.192
[epoch: 1, i:  7599] avg mini-batch loss: 1.259
[epoch: 1, i:  7699] avg mini-batch loss: 1.205
[epoch: 1, i:  7799] avg mini-batch loss: 1.245
[epoch: 1, i:  7899] avg mini-batch loss: 1.215
[epoch: 1, i:  7999] avg mini-batch loss: 1.193
[epoch: 1, i:  8099] avg mini-batch loss: 1.166
[epoch: 1, i:  8199] avg mini-batch loss: 1.123
[epoch: 1, i:  8299] avg mini-batch loss: 1.159
[epoch: 1, i:  8399] avg mini-batch loss: 1.173
[epoch: 1, i:  8499] avg mini-batch loss: 1.321
[epoch: 1, i:  8599] avg mini-batch loss: 1.256
[epoch: 1, i:  8699] avg mini-batch loss: 1.201
[epoch: 1, i:  8799] avg mini-batch loss: 1.285
[epoch: 1, i:  8899] avg mini-batch loss: 1.230
[epoch: 1, i:  8999] avg mini-batch loss: 1.147
[epoch: 1, i:  9099] avg mini-batch loss: 1.175
[epoch: 1, i:  9199] avg mini-batch loss: 1.121
[epoch: 1, i:  9299] avg mini-batch loss: 1.205
[epoch: 1, i:  9399] avg mini-batch loss: 1.226
[epoch: 1, i:  9499] avg mini-batch loss: 1.156
[epoch: 1, i:  9599] avg mini-batch loss: 1.228
[epoch: 1, i:  9699] avg mini-batch loss: 1.373
[epoch: 1, i:  9799] avg mini-batch loss: 1.255
[epoch: 1, i:  9899] avg mini-batch loss: 1.300
[epoch: 1, i:  9999] avg mini-batch loss: 1.281
[epoch: 1, i: 10099] avg mini-batch loss: 1.149
[epoch: 1, i: 10199] avg mini-batch loss: 1.242
[epoch: 1, i: 10299] avg mini-batch loss: 1.188
[epoch: 1, i: 10399] avg mini-batch loss: 1.281
[epoch: 1, i: 10499] avg mini-batch loss: 1.158
[epoch: 1, i: 10599] avg mini-batch loss: 1.226
[epoch: 1, i: 10699] avg mini-batch loss: 1.199
[epoch: 1, i: 10799] avg mini-batch loss: 1.124
[epoch: 1, i: 10899] avg mini-batch loss: 1.230
[epoch: 1, i: 10999] avg mini-batch loss: 1.150
[epoch: 1, i: 11099] avg mini-batch loss: 1.089
[epoch: 1, i: 11199] avg mini-batch loss: 1.176
[epoch: 1, i: 11299] avg mini-batch loss: 1.231
[epoch: 1, i: 11399] avg mini-batch loss: 1.355
[epoch: 1, i: 11499] avg mini-batch loss: 1.211
[epoch: 1, i: 11599] avg mini-batch loss: 1.242
[epoch: 1, i: 11699] avg mini-batch loss: 1.247
[epoch: 1, i: 11799] avg mini-batch loss: 1.088
[epoch: 1, i: 11899] avg mini-batch loss: 1.285
[epoch: 1, i: 11999] avg mini-batch loss: 1.218
[epoch: 1, i: 12099] avg mini-batch loss: 1.193
```

```
[epoch: 1, i: 12199] avg mini-batch loss: 1.079
[epoch: 1, i: 12299] avg mini-batch loss: 1.202
[epoch: 1, i: 12399] avg mini-batch loss: 1.109
[epoch: 1, i: 12499] avg mini-batch loss: 1.205
[epoch: 2, i:    99] avg mini-batch loss: 1.185
[epoch: 2, i:   199] avg mini-batch loss: 1.098
[epoch: 2, i:   299] avg mini-batch loss: 1.100
[epoch: 2, i:   399] avg mini-batch loss: 1.134
[epoch: 2, i:   499] avg mini-batch loss: 1.115
[epoch: 2, i:   599] avg mini-batch loss: 1.116
[epoch: 2, i:   699] avg mini-batch loss: 1.150
[epoch: 2, i:   799] avg mini-batch loss: 1.186
[epoch: 2, i:   899] avg mini-batch loss: 1.153
[epoch: 2, i:   999] avg mini-batch loss: 1.134
[epoch: 2, i:  1099] avg mini-batch loss: 1.150
[epoch: 2, i:  1199] avg mini-batch loss: 1.047
[epoch: 2, i:  1299] avg mini-batch loss: 1.069
[epoch: 2, i:  1399] avg mini-batch loss: 1.042
[epoch: 2, i:  1499] avg mini-batch loss: 1.103
[epoch: 2, i:  1599] avg mini-batch loss: 1.118
[epoch: 2, i:  1699] avg mini-batch loss: 1.129
[epoch: 2, i:  1799] avg mini-batch loss: 1.071
[epoch: 2, i:  1899] avg mini-batch loss: 1.170
[epoch: 2, i:  1999] avg mini-batch loss: 1.095
[epoch: 2, i:  2099] avg mini-batch loss: 1.121
[epoch: 2, i:  2199] avg mini-batch loss: 1.080
[epoch: 2, i:  2299] avg mini-batch loss: 1.072
[epoch: 2, i:  2399] avg mini-batch loss: 1.060
[epoch: 2, i:  2499] avg mini-batch loss: 1.150
[epoch: 2, i:  2599] avg mini-batch loss: 1.086
[epoch: 2, i:  2699] avg mini-batch loss: 1.143
[epoch: 2, i:  2799] avg mini-batch loss: 1.080
[epoch: 2, i:  2899] avg mini-batch loss: 1.084
[epoch: 2, i:  2999] avg mini-batch loss: 1.115
[epoch: 2, i:  3099] avg mini-batch loss: 1.093
[epoch: 2, i:  3199] avg mini-batch loss: 1.134
[epoch: 2, i:  3299] avg mini-batch loss: 1.112
[epoch: 2, i:  3399] avg mini-batch loss: 1.179
[epoch: 2, i:  3499] avg mini-batch loss: 1.104
[epoch: 2, i:  3599] avg mini-batch loss: 1.012
[epoch: 2, i:  3699] avg mini-batch loss: 1.125
[epoch: 2, i:  3799] avg mini-batch loss: 1.098
[epoch: 2, i:  3899] avg mini-batch loss: 1.133
[epoch: 2, i:  3999] avg mini-batch loss: 1.153
[epoch: 2, i:  4099] avg mini-batch loss: 1.132
[epoch: 2, i:  4199] avg mini-batch loss: 1.091
[epoch: 2, i:  4299] avg mini-batch loss: 1.038
[epoch: 2, i:  4399] avg mini-batch loss: 1.112
```

```
[epoch: 2, i:  4499] avg mini-batch loss: 1.100
[epoch: 2, i:  4599] avg mini-batch loss: 1.122
[epoch: 2, i:  4699] avg mini-batch loss: 0.987
[epoch: 2, i:  4799] avg mini-batch loss: 1.152
[epoch: 2, i:  4899] avg mini-batch loss: 1.176
[epoch: 2, i:  4999] avg mini-batch loss: 1.017
[epoch: 2, i:  5099] avg mini-batch loss: 1.067
[epoch: 2, i:  5199] avg mini-batch loss: 1.115
[epoch: 2, i:  5299] avg mini-batch loss: 1.134
[epoch: 2, i:  5399] avg mini-batch loss: 1.139
[epoch: 2, i:  5499] avg mini-batch loss: 1.060
[epoch: 2, i:  5599] avg mini-batch loss: 1.134
[epoch: 2, i:  5699] avg mini-batch loss: 1.220
[epoch: 2, i:  5799] avg mini-batch loss: 1.059
[epoch: 2, i:  5899] avg mini-batch loss: 1.094
[epoch: 2, i:  5999] avg mini-batch loss: 1.109
[epoch: 2, i:  6099] avg mini-batch loss: 1.018
[epoch: 2, i:  6199] avg mini-batch loss: 1.096
[epoch: 2, i:  6299] avg mini-batch loss: 1.135
[epoch: 2, i:  6399] avg mini-batch loss: 1.071
[epoch: 2, i:  6499] avg mini-batch loss: 1.048
[epoch: 2, i:  6599] avg mini-batch loss: 1.111
[epoch: 2, i:  6699] avg mini-batch loss: 1.113
[epoch: 2, i:  6799] avg mini-batch loss: 1.049
[epoch: 2, i:  6899] avg mini-batch loss: 1.101
[epoch: 2, i:  6999] avg mini-batch loss: 1.188
[epoch: 2, i:  7099] avg mini-batch loss: 1.098
[epoch: 2, i:  7199] avg mini-batch loss: 1.124
[epoch: 2, i:  7299] avg mini-batch loss: 1.149
[epoch: 2, i:  7399] avg mini-batch loss: 1.191
[epoch: 2, i:  7499] avg mini-batch loss: 1.063
[epoch: 2, i:  7599] avg mini-batch loss: 1.025
[epoch: 2, i:  7699] avg mini-batch loss: 1.116
[epoch: 2, i:  7799] avg mini-batch loss: 1.072
[epoch: 2, i:  7899] avg mini-batch loss: 1.116
[epoch: 2, i:  7999] avg mini-batch loss: 1.143
[epoch: 2, i:  8099] avg mini-batch loss: 1.020
[epoch: 2, i:  8199] avg mini-batch loss: 1.083
[epoch: 2, i:  8299] avg mini-batch loss: 1.063
[epoch: 2, i:  8399] avg mini-batch loss: 1.119
[epoch: 2, i:  8499] avg mini-batch loss: 1.197
[epoch: 2, i:  8599] avg mini-batch loss: 1.081
[epoch: 2, i:  8699] avg mini-batch loss: 1.111
[epoch: 2, i:  8799] avg mini-batch loss: 1.158
[epoch: 2, i:  8899] avg mini-batch loss: 1.066
[epoch: 2, i:  8999] avg mini-batch loss: 1.156
[epoch: 2, i:  9099] avg mini-batch loss: 1.120
[epoch: 2, i:  9199] avg mini-batch loss: 1.085
```

```
[epoch: 2, i:  9299] avg mini-batch loss: 1.036
[epoch: 2, i:  9399] avg mini-batch loss: 1.054
[epoch: 2, i:  9499] avg mini-batch loss: 1.111
[epoch: 2, i:  9599] avg mini-batch loss: 0.997
[epoch: 2, i:  9699] avg mini-batch loss: 1.018
[epoch: 2, i:  9799] avg mini-batch loss: 1.048
[epoch: 2, i:  9899] avg mini-batch loss: 1.010
[epoch: 2, i:  9999] avg mini-batch loss: 1.182
[epoch: 2, i: 10099] avg mini-batch loss: 1.151
[epoch: 2, i: 10199] avg mini-batch loss: 1.159
[epoch: 2, i: 10299] avg mini-batch loss: 1.111
[epoch: 2, i: 10399] avg mini-batch loss: 1.037
[epoch: 2, i: 10499] avg mini-batch loss: 1.044
[epoch: 2, i: 10599] avg mini-batch loss: 1.055
[epoch: 2, i: 10699] avg mini-batch loss: 1.031
[epoch: 2, i: 10799] avg mini-batch loss: 1.038
[epoch: 2, i: 10899] avg mini-batch loss: 0.981
[epoch: 2, i: 10999] avg mini-batch loss: 1.056
[epoch: 2, i: 11099] avg mini-batch loss: 1.059
[epoch: 2, i: 11199] avg mini-batch loss: 1.041
[epoch: 2, i: 11299] avg mini-batch loss: 1.032
[epoch: 2, i: 11399] avg mini-batch loss: 1.079
[epoch: 2, i: 11499] avg mini-batch loss: 0.991
[epoch: 2, i: 11599] avg mini-batch loss: 1.134
[epoch: 2, i: 11699] avg mini-batch loss: 1.062
[epoch: 2, i: 11799] avg mini-batch loss: 1.000
[epoch: 2, i: 11899] avg mini-batch loss: 0.979
[epoch: 2, i: 11999] avg mini-batch loss: 1.007
[epoch: 2, i: 12099] avg mini-batch loss: 0.919
[epoch: 2, i: 12199] avg mini-batch loss: 1.092
[epoch: 2, i: 12299] avg mini-batch loss: 1.132
[epoch: 2, i: 12399] avg mini-batch loss: 1.044
[epoch: 2, i: 12499] avg mini-batch loss: 1.041
[epoch: 3, i:    99] avg mini-batch loss: 0.920
[epoch: 3, i:   199] avg mini-batch loss: 1.029
[epoch: 3, i:   299] avg mini-batch loss: 1.009
[epoch: 3, i:   399] avg mini-batch loss: 1.037
[epoch: 3, i:   499] avg mini-batch loss: 0.964
[epoch: 3, i:   599] avg mini-batch loss: 1.035
[epoch: 3, i:   699] avg mini-batch loss: 0.967
[epoch: 3, i:   799] avg mini-batch loss: 1.048
[epoch: 3, i:   899] avg mini-batch loss: 1.027
[epoch: 3, i:   999] avg mini-batch loss: 0.949
[epoch: 3, i:  1099] avg mini-batch loss: 0.972
[epoch: 3, i:  1199] avg mini-batch loss: 0.972
[epoch: 3, i:  1299] avg mini-batch loss: 1.034
[epoch: 3, i:  1399] avg mini-batch loss: 0.985
[epoch: 3, i:  1499] avg mini-batch loss: 1.089
```

```
[epoch: 3, i:  1599] avg mini-batch loss: 1.002
[epoch: 3, i:  1699] avg mini-batch loss: 1.055
[epoch: 3, i:  1799] avg mini-batch loss: 0.966
[epoch: 3, i:  1899] avg mini-batch loss: 0.891
[epoch: 3, i:  1999] avg mini-batch loss: 0.875
[epoch: 3, i:  2099] avg mini-batch loss: 0.980
[epoch: 3, i:  2199] avg mini-batch loss: 0.920
[epoch: 3, i:  2299] avg mini-batch loss: 0.995
[epoch: 3, i:  2399] avg mini-batch loss: 0.973
[epoch: 3, i:  2499] avg mini-batch loss: 0.944
[epoch: 3, i:  2599] avg mini-batch loss: 0.955
[epoch: 3, i:  2699] avg mini-batch loss: 0.994
[epoch: 3, i:  2799] avg mini-batch loss: 0.856
[epoch: 3, i:  2899] avg mini-batch loss: 1.064
[epoch: 3, i:  2999] avg mini-batch loss: 0.979
[epoch: 3, i:  3099] avg mini-batch loss: 1.020
[epoch: 3, i:  3199] avg mini-batch loss: 0.919
[epoch: 3, i:  3299] avg mini-batch loss: 0.970
[epoch: 3, i:  3399] avg mini-batch loss: 0.935
[epoch: 3, i:  3499] avg mini-batch loss: 1.019
[epoch: 3, i:  3599] avg mini-batch loss: 0.923
[epoch: 3, i:  3699] avg mini-batch loss: 0.891
[epoch: 3, i:  3799] avg mini-batch loss: 1.004
[epoch: 3, i:  3899] avg mini-batch loss: 1.040
[epoch: 3, i:  3999] avg mini-batch loss: 1.022
[epoch: 3, i:  4099] avg mini-batch loss: 0.993
[epoch: 3, i:  4199] avg mini-batch loss: 1.041
[epoch: 3, i:  4299] avg mini-batch loss: 1.065
[epoch: 3, i:  4399] avg mini-batch loss: 0.914
[epoch: 3, i:  4499] avg mini-batch loss: 0.944
[epoch: 3, i:  4599] avg mini-batch loss: 0.990
[epoch: 3, i:  4699] avg mini-batch loss: 0.975
[epoch: 3, i:  4799] avg mini-batch loss: 0.978
[epoch: 3, i:  4899] avg mini-batch loss: 0.985
[epoch: 3, i:  4999] avg mini-batch loss: 1.055
[epoch: 3, i:  5099] avg mini-batch loss: 0.954
[epoch: 3, i:  5199] avg mini-batch loss: 0.954
[epoch: 3, i:  5299] avg mini-batch loss: 0.970
[epoch: 3, i:  5399] avg mini-batch loss: 1.089
[epoch: 3, i:  5499] avg mini-batch loss: 0.954
[epoch: 3, i:  5599] avg mini-batch loss: 0.983
[epoch: 3, i:  5699] avg mini-batch loss: 0.934
[epoch: 3, i:  5799] avg mini-batch loss: 0.991
[epoch: 3, i:  5899] avg mini-batch loss: 0.980
[epoch: 3, i:  5999] avg mini-batch loss: 0.981
[epoch: 3, i:  6099] avg mini-batch loss: 0.871
[epoch: 3, i:  6199] avg mini-batch loss: 1.024
[epoch: 3, i:  6299] avg mini-batch loss: 0.940
```

```
[epoch: 3, i:  6399] avg mini-batch loss: 1.009
[epoch: 3, i:  6499] avg mini-batch loss: 0.916
[epoch: 3, i:  6599] avg mini-batch loss: 0.987
[epoch: 3, i:  6699] avg mini-batch loss: 1.020
[epoch: 3, i:  6799] avg mini-batch loss: 0.970
[epoch: 3, i:  6899] avg mini-batch loss: 0.951
[epoch: 3, i:  6999] avg mini-batch loss: 1.027
[epoch: 3, i:  7099] avg mini-batch loss: 0.959
[epoch: 3, i:  7199] avg mini-batch loss: 1.079
[epoch: 3, i:  7299] avg mini-batch loss: 0.874
[epoch: 3, i:  7399] avg mini-batch loss: 0.939
[epoch: 3, i:  7499] avg mini-batch loss: 0.877
[epoch: 3, i:  7599] avg mini-batch loss: 0.969
[epoch: 3, i:  7699] avg mini-batch loss: 0.998
[epoch: 3, i:  7799] avg mini-batch loss: 0.937
[epoch: 3, i:  7899] avg mini-batch loss: 0.922
[epoch: 3, i:  7999] avg mini-batch loss: 1.011
[epoch: 3, i:  8099] avg mini-batch loss: 1.014
[epoch: 3, i:  8199] avg mini-batch loss: 1.021
[epoch: 3, i:  8299] avg mini-batch loss: 1.053
[epoch: 3, i:  8399] avg mini-batch loss: 1.024
[epoch: 3, i:  8499] avg mini-batch loss: 0.886
[epoch: 3, i:  8599] avg mini-batch loss: 0.963
[epoch: 3, i:  8699] avg mini-batch loss: 0.987
[epoch: 3, i:  8799] avg mini-batch loss: 0.999
[epoch: 3, i:  8899] avg mini-batch loss: 0.907
[epoch: 3, i:  8999] avg mini-batch loss: 0.982
[epoch: 3, i:  9099] avg mini-batch loss: 0.897
[epoch: 3, i:  9199] avg mini-batch loss: 1.019
[epoch: 3, i:  9299] avg mini-batch loss: 0.932
[epoch: 3, i:  9399] avg mini-batch loss: 0.975
[epoch: 3, i:  9499] avg mini-batch loss: 0.953
[epoch: 3, i:  9599] avg mini-batch loss: 1.023
[epoch: 3, i:  9699] avg mini-batch loss: 1.026
[epoch: 3, i:  9799] avg mini-batch loss: 1.020
[epoch: 3, i:  9899] avg mini-batch loss: 0.942
[epoch: 3, i:  9999] avg mini-batch loss: 1.068
[epoch: 3, i: 10099] avg mini-batch loss: 0.957
[epoch: 3, i: 10199] avg mini-batch loss: 1.010
[epoch: 3, i: 10299] avg mini-batch loss: 0.929
[epoch: 3, i: 10399] avg mini-batch loss: 0.918
[epoch: 3, i: 10499] avg mini-batch loss: 0.915
[epoch: 3, i: 10599] avg mini-batch loss: 0.886
[epoch: 3, i: 10699] avg mini-batch loss: 0.974
[epoch: 3, i: 10799] avg mini-batch loss: 0.940
[epoch: 3, i: 10899] avg mini-batch loss: 0.977
[epoch: 3, i: 10999] avg mini-batch loss: 0.967
[epoch: 3, i: 11099] avg mini-batch loss: 0.943
```

```
[epoch: 3, i: 11199] avg mini-batch loss: 0.881
[epoch: 3, i: 11299] avg mini-batch loss: 1.034
[epoch: 3, i: 11399] avg mini-batch loss: 0.992
[epoch: 3, i: 11499] avg mini-batch loss: 0.993
[epoch: 3, i: 11599] avg mini-batch loss: 0.955
[epoch: 3, i: 11699] avg mini-batch loss: 1.012
[epoch: 3, i: 11799] avg mini-batch loss: 0.952
[epoch: 3, i: 11899] avg mini-batch loss: 1.016
[epoch: 3, i: 11999] avg mini-batch loss: 1.038
[epoch: 3, i: 12099] avg mini-batch loss: 1.024
[epoch: 3, i: 12199] avg mini-batch loss: 1.020
[epoch: 3, i: 12299] avg mini-batch loss: 0.925
[epoch: 3, i: 12399] avg mini-batch loss: 1.013
[epoch: 3, i: 12499] avg mini-batch loss: 0.952
[epoch: 4, i:    99] avg mini-batch loss: 0.838
[epoch: 4, i:   199] avg mini-batch loss: 0.880
[epoch: 4, i:   299] avg mini-batch loss: 0.817
[epoch: 4, i:   399] avg mini-batch loss: 0.829
[epoch: 4, i:   499] avg mini-batch loss: 0.869
[epoch: 4, i:   599] avg mini-batch loss: 0.849
[epoch: 4, i:   699] avg mini-batch loss: 0.796
[epoch: 4, i:   799] avg mini-batch loss: 0.969
[epoch: 4, i:   899] avg mini-batch loss: 0.881
[epoch: 4, i:   999] avg mini-batch loss: 0.902
[epoch: 4, i:  1099] avg mini-batch loss: 0.772
[epoch: 4, i:  1199] avg mini-batch loss: 0.962
[epoch: 4, i:  1299] avg mini-batch loss: 0.864
[epoch: 4, i:  1399] avg mini-batch loss: 0.847
[epoch: 4, i:  1499] avg mini-batch loss: 0.878
[epoch: 4, i:  1599] avg mini-batch loss: 0.914
[epoch: 4, i:  1699] avg mini-batch loss: 0.797
[epoch: 4, i:  1799] avg mini-batch loss: 0.999
[epoch: 4, i:  1899] avg mini-batch loss: 0.808
[epoch: 4, i:  1999] avg mini-batch loss: 0.940
[epoch: 4, i:  2099] avg mini-batch loss: 0.778
[epoch: 4, i:  2199] avg mini-batch loss: 0.995
[epoch: 4, i:  2299] avg mini-batch loss: 0.899
[epoch: 4, i:  2399] avg mini-batch loss: 0.879
[epoch: 4, i:  2499] avg mini-batch loss: 0.852
[epoch: 4, i:  2599] avg mini-batch loss: 0.869
[epoch: 4, i:  2699] avg mini-batch loss: 0.926
[epoch: 4, i:  2799] avg mini-batch loss: 0.827
[epoch: 4, i:  2899] avg mini-batch loss: 0.887
[epoch: 4, i:  2999] avg mini-batch loss: 0.887
[epoch: 4, i:  3099] avg mini-batch loss: 0.828
[epoch: 4, i:  3199] avg mini-batch loss: 0.867
[epoch: 4, i:  3299] avg mini-batch loss: 0.824
[epoch: 4, i:  3399] avg mini-batch loss: 0.844
```

```
[epoch: 4, i:  3499] avg mini-batch loss: 0.875
[epoch: 4, i:  3599] avg mini-batch loss: 0.780
[epoch: 4, i:  3699] avg mini-batch loss: 0.963
[epoch: 4, i:  3799] avg mini-batch loss: 0.935
[epoch: 4, i:  3899] avg mini-batch loss: 0.839
[epoch: 4, i:  3999] avg mini-batch loss: 0.810
[epoch: 4, i:  4099] avg mini-batch loss: 0.859
[epoch: 4, i:  4199] avg mini-batch loss: 0.907
[epoch: 4, i:  4299] avg mini-batch loss: 0.885
[epoch: 4, i:  4399] avg mini-batch loss: 0.924
[epoch: 4, i:  4499] avg mini-batch loss: 1.004
[epoch: 4, i:  4599] avg mini-batch loss: 0.864
[epoch: 4, i:  4699] avg mini-batch loss: 0.906
[epoch: 4, i:  4799] avg mini-batch loss: 0.925
[epoch: 4, i:  4899] avg mini-batch loss: 0.820
[epoch: 4, i:  4999] avg mini-batch loss: 0.862
[epoch: 4, i:  5099] avg mini-batch loss: 0.853
[epoch: 4, i:  5199] avg mini-batch loss: 0.860
[epoch: 4, i:  5299] avg mini-batch loss: 0.847
[epoch: 4, i:  5399] avg mini-batch loss: 0.885
[epoch: 4, i:  5499] avg mini-batch loss: 0.924
[epoch: 4, i:  5599] avg mini-batch loss: 0.919
[epoch: 4, i:  5699] avg mini-batch loss: 0.858
[epoch: 4, i:  5799] avg mini-batch loss: 0.912
[epoch: 4, i:  5899] avg mini-batch loss: 0.950
[epoch: 4, i:  5999] avg mini-batch loss: 0.831
[epoch: 4, i:  6099] avg mini-batch loss: 0.912
[epoch: 4, i:  6199] avg mini-batch loss: 0.876
[epoch: 4, i:  6299] avg mini-batch loss: 0.875
[epoch: 4, i:  6399] avg mini-batch loss: 0.912
[epoch: 4, i:  6499] avg mini-batch loss: 0.844
[epoch: 4, i:  6599] avg mini-batch loss: 0.831
[epoch: 4, i:  6699] avg mini-batch loss: 0.910
[epoch: 4, i:  6799] avg mini-batch loss: 0.888
[epoch: 4, i:  6899] avg mini-batch loss: 0.906
[epoch: 4, i:  6999] avg mini-batch loss: 0.962
[epoch: 4, i:  7099] avg mini-batch loss: 0.970
[epoch: 4, i:  7199] avg mini-batch loss: 0.909
[epoch: 4, i:  7299] avg mini-batch loss: 0.886
[epoch: 4, i:  7399] avg mini-batch loss: 0.890
[epoch: 4, i:  7499] avg mini-batch loss: 0.914
[epoch: 4, i:  7599] avg mini-batch loss: 0.896
[epoch: 4, i:  7699] avg mini-batch loss: 0.981
[epoch: 4, i:  7799] avg mini-batch loss: 0.977
[epoch: 4, i:  7899] avg mini-batch loss: 0.890
[epoch: 4, i:  7999] avg mini-batch loss: 0.873
[epoch: 4, i:  8099] avg mini-batch loss: 0.964
[epoch: 4, i:  8199] avg mini-batch loss: 0.918
```

```
[epoch: 4, i:  8299] avg mini-batch loss: 0.933
[epoch: 4, i:  8399] avg mini-batch loss: 0.991
[epoch: 4, i:  8499] avg mini-batch loss: 0.807
[epoch: 4, i:  8599] avg mini-batch loss: 0.809
[epoch: 4, i:  8699] avg mini-batch loss: 0.803
[epoch: 4, i:  8799] avg mini-batch loss: 0.794
[epoch: 4, i:  8899] avg mini-batch loss: 0.908
[epoch: 4, i:  8999] avg mini-batch loss: 0.946
[epoch: 4, i:  9099] avg mini-batch loss: 0.888
[epoch: 4, i:  9199] avg mini-batch loss: 0.966
[epoch: 4, i:  9299] avg mini-batch loss: 0.853
[epoch: 4, i:  9399] avg mini-batch loss: 0.785
[epoch: 4, i:  9499] avg mini-batch loss: 0.862
[epoch: 4, i:  9599] avg mini-batch loss: 0.953
[epoch: 4, i:  9699] avg mini-batch loss: 0.942
[epoch: 4, i:  9799] avg mini-batch loss: 0.932
[epoch: 4, i:  9899] avg mini-batch loss: 0.915
[epoch: 4, i:  9999] avg mini-batch loss: 0.973
[epoch: 4, i: 10099] avg mini-batch loss: 0.897
[epoch: 4, i: 10199] avg mini-batch loss: 0.816
[epoch: 4, i: 10299] avg mini-batch loss: 0.846
[epoch: 4, i: 10399] avg mini-batch loss: 0.938
[epoch: 4, i: 10499] avg mini-batch loss: 0.790
[epoch: 4, i: 10599] avg mini-batch loss: 0.949
[epoch: 4, i: 10699] avg mini-batch loss: 0.869
[epoch: 4, i: 10799] avg mini-batch loss: 0.862
[epoch: 4, i: 10899] avg mini-batch loss: 0.965
[epoch: 4, i: 10999] avg mini-batch loss: 0.742
[epoch: 4, i: 11099] avg mini-batch loss: 0.898
[epoch: 4, i: 11199] avg mini-batch loss: 0.883
[epoch: 4, i: 11299] avg mini-batch loss: 0.804
[epoch: 4, i: 11399] avg mini-batch loss: 0.889
[epoch: 4, i: 11499] avg mini-batch loss: 0.853
[epoch: 4, i: 11599] avg mini-batch loss: 0.895
[epoch: 4, i: 11699] avg mini-batch loss: 0.878
[epoch: 4, i: 11799] avg mini-batch loss: 0.883
[epoch: 4, i: 11899] avg mini-batch loss: 0.821
[epoch: 4, i: 11999] avg mini-batch loss: 0.832
[epoch: 4, i: 12099] avg mini-batch loss: 1.045
[epoch: 4, i: 12199] avg mini-batch loss: 0.898
[epoch: 4, i: 12299] avg mini-batch loss: 0.888
[epoch: 4, i: 12399] avg mini-batch loss: 0.989
[epoch: 4, i: 12499] avg mini-batch loss: 0.970
[epoch: 5, i:    99] avg mini-batch loss: 0.791
[epoch: 5, i:   199] avg mini-batch loss: 0.746
[epoch: 5, i:   299] avg mini-batch loss: 0.808
[epoch: 5, i:   399] avg mini-batch loss: 0.695
[epoch: 5, i:   499] avg mini-batch loss: 0.754
```

```
[epoch: 5, i:   599] avg mini-batch loss: 0.803
[epoch: 5, i:   699] avg mini-batch loss: 0.835
[epoch: 5, i:   799] avg mini-batch loss: 0.727
[epoch: 5, i:   899] avg mini-batch loss: 0.769
[epoch: 5, i:   999] avg mini-batch loss: 0.748
[epoch: 5, i:  1099] avg mini-batch loss: 0.733
[epoch: 5, i:  1199] avg mini-batch loss: 0.919
[epoch: 5, i:  1299] avg mini-batch loss: 0.903
[epoch: 5, i:  1399] avg mini-batch loss: 0.868
[epoch: 5, i:  1499] avg mini-batch loss: 0.795
[epoch: 5, i:  1599] avg mini-batch loss: 0.783
[epoch: 5, i:  1699] avg mini-batch loss: 0.649
[epoch: 5, i:  1799] avg mini-batch loss: 0.727
[epoch: 5, i:  1899] avg mini-batch loss: 0.797
[epoch: 5, i:  1999] avg mini-batch loss: 0.793
[epoch: 5, i:  2099] avg mini-batch loss: 0.916
[epoch: 5, i:  2199] avg mini-batch loss: 0.819
[epoch: 5, i:  2299] avg mini-batch loss: 0.856
[epoch: 5, i:  2399] avg mini-batch loss: 0.777
[epoch: 5, i:  2499] avg mini-batch loss: 0.784
[epoch: 5, i:  2599] avg mini-batch loss: 0.693
[epoch: 5, i:  2699] avg mini-batch loss: 0.874
[epoch: 5, i:  2799] avg mini-batch loss: 0.680
[epoch: 5, i:  2899] avg mini-batch loss: 0.810
[epoch: 5, i:  2999] avg mini-batch loss: 0.792
[epoch: 5, i:  3099] avg mini-batch loss: 0.737
[epoch: 5, i:  3199] avg mini-batch loss: 0.732
[epoch: 5, i:  3299] avg mini-batch loss: 0.927
[epoch: 5, i:  3399] avg mini-batch loss: 0.892
[epoch: 5, i:  3499] avg mini-batch loss: 0.803
[epoch: 5, i:  3599] avg mini-batch loss: 0.757
[epoch: 5, i:  3699] avg mini-batch loss: 0.814
[epoch: 5, i:  3799] avg mini-batch loss: 0.788
[epoch: 5, i:  3899] avg mini-batch loss: 0.813
[epoch: 5, i:  3999] avg mini-batch loss: 0.809
[epoch: 5, i:  4099] avg mini-batch loss: 0.845
[epoch: 5, i:  4199] avg mini-batch loss: 0.777
[epoch: 5, i:  4299] avg mini-batch loss: 0.761
[epoch: 5, i:  4399] avg mini-batch loss: 0.777
[epoch: 5, i:  4499] avg mini-batch loss: 0.934
[epoch: 5, i:  4599] avg mini-batch loss: 0.891
[epoch: 5, i:  4699] avg mini-batch loss: 0.726
[epoch: 5, i:  4799] avg mini-batch loss: 0.800
[epoch: 5, i:  4899] avg mini-batch loss: 0.842
[epoch: 5, i:  4999] avg mini-batch loss: 0.841
[epoch: 5, i:  5099] avg mini-batch loss: 0.715
[epoch: 5, i:  5199] avg mini-batch loss: 0.863
[epoch: 5, i:  5299] avg mini-batch loss: 0.760
```

```
[epoch: 5, i:  5399] avg mini-batch loss: 0.727
[epoch: 5, i:  5499] avg mini-batch loss: 0.789
[epoch: 5, i:  5599] avg mini-batch loss: 0.737
[epoch: 5, i:  5699] avg mini-batch loss: 0.794
[epoch: 5, i:  5799] avg mini-batch loss: 0.788
[epoch: 5, i:  5899] avg mini-batch loss: 0.821
[epoch: 5, i:  5999] avg mini-batch loss: 0.854
[epoch: 5, i:  6099] avg mini-batch loss: 0.785
[epoch: 5, i:  6199] avg mini-batch loss: 0.824
[epoch: 5, i:  6299] avg mini-batch loss: 0.856
[epoch: 5, i:  6399] avg mini-batch loss: 0.814
[epoch: 5, i:  6499] avg mini-batch loss: 0.818
[epoch: 5, i:  6599] avg mini-batch loss: 0.822
[epoch: 5, i:  6699] avg mini-batch loss: 0.823
[epoch: 5, i:  6799] avg mini-batch loss: 0.809
[epoch: 5, i:  6899] avg mini-batch loss: 0.773
[epoch: 5, i:  6999] avg mini-batch loss: 0.808
[epoch: 5, i:  7099] avg mini-batch loss: 0.834
[epoch: 5, i:  7199] avg mini-batch loss: 0.793
[epoch: 5, i:  7299] avg mini-batch loss: 0.762
[epoch: 5, i:  7399] avg mini-batch loss: 0.755
[epoch: 5, i:  7499] avg mini-batch loss: 0.783
[epoch: 5, i:  7599] avg mini-batch loss: 0.750
[epoch: 5, i:  7699] avg mini-batch loss: 0.806
[epoch: 5, i:  7799] avg mini-batch loss: 0.751
[epoch: 5, i:  7899] avg mini-batch loss: 0.839
[epoch: 5, i:  7999] avg mini-batch loss: 0.839
[epoch: 5, i:  8099] avg mini-batch loss: 0.856
[epoch: 5, i:  8199] avg mini-batch loss: 0.823
[epoch: 5, i:  8299] avg mini-batch loss: 0.785
[epoch: 5, i:  8399] avg mini-batch loss: 0.987
[epoch: 5, i:  8499] avg mini-batch loss: 0.773
[epoch: 5, i:  8599] avg mini-batch loss: 0.912
[epoch: 5, i:  8699] avg mini-batch loss: 0.733
[epoch: 5, i:  8799] avg mini-batch loss: 0.781
[epoch: 5, i:  8899] avg mini-batch loss: 0.703
[epoch: 5, i:  8999] avg mini-batch loss: 0.854
[epoch: 5, i:  9099] avg mini-batch loss: 0.873
[epoch: 5, i:  9199] avg mini-batch loss: 0.893
[epoch: 5, i:  9299] avg mini-batch loss: 0.800
[epoch: 5, i:  9399] avg mini-batch loss: 0.754
[epoch: 5, i:  9499] avg mini-batch loss: 0.829
[epoch: 5, i:  9599] avg mini-batch loss: 0.862
[epoch: 5, i:  9699] avg mini-batch loss: 0.813
[epoch: 5, i:  9799] avg mini-batch loss: 0.738
[epoch: 5, i:  9899] avg mini-batch loss: 0.815
[epoch: 5, i:  9999] avg mini-batch loss: 0.818
[epoch: 5, i: 10099] avg mini-batch loss: 0.822
```

```
[epoch: 5, i: 10199] avg mini-batch loss: 0.811
[epoch: 5, i: 10299] avg mini-batch loss: 0.856
[epoch: 5, i: 10399] avg mini-batch loss: 0.830
[epoch: 5, i: 10499] avg mini-batch loss: 0.791
[epoch: 5, i: 10599] avg mini-batch loss: 0.807
[epoch: 5, i: 10699] avg mini-batch loss: 0.784
[epoch: 5, i: 10799] avg mini-batch loss: 0.804
[epoch: 5, i: 10899] avg mini-batch loss: 0.931
[epoch: 5, i: 10999] avg mini-batch loss: 0.847
[epoch: 5, i: 11099] avg mini-batch loss: 0.799
[epoch: 5, i: 11199] avg mini-batch loss: 0.877
[epoch: 5, i: 11299] avg mini-batch loss: 0.794
[epoch: 5, i: 11399] avg mini-batch loss: 0.898
[epoch: 5, i: 11499] avg mini-batch loss: 0.762
[epoch: 5, i: 11599] avg mini-batch loss: 0.737
[epoch: 5, i: 11699] avg mini-batch loss: 0.867
[epoch: 5, i: 11799] avg mini-batch loss: 0.844
[epoch: 5, i: 11899] avg mini-batch loss: 0.838
[epoch: 5, i: 11999] avg mini-batch loss: 0.832
[epoch: 5, i: 12099] avg mini-batch loss: 0.878
[epoch: 5, i: 12199] avg mini-batch loss: 0.844
[epoch: 5, i: 12299] avg mini-batch loss: 0.781
[epoch: 5, i: 12399] avg mini-batch loss: 0.912
[epoch: 5, i: 12499] avg mini-batch loss: 0.754
[epoch: 6, i:    99] avg mini-batch loss: 0.717
[epoch: 6, i:   199] avg mini-batch loss: 0.689
[epoch: 6, i:   299] avg mini-batch loss: 0.706
[epoch: 6, i:   399] avg mini-batch loss: 0.690
[epoch: 6, i:   499] avg mini-batch loss: 0.730
[epoch: 6, i:   599] avg mini-batch loss: 0.694
[epoch: 6, i:   699] avg mini-batch loss: 0.747
[epoch: 6, i:   799] avg mini-batch loss: 0.636
[epoch: 6, i:   899] avg mini-batch loss: 0.649
[epoch: 6, i:   999] avg mini-batch loss: 0.742
[epoch: 6, i:  1099] avg mini-batch loss: 0.585
[epoch: 6, i:  1199] avg mini-batch loss: 0.666
[epoch: 6, i:  1299] avg mini-batch loss: 0.728
[epoch: 6, i:  1399] avg mini-batch loss: 0.685
[epoch: 6, i:  1499] avg mini-batch loss: 0.763
[epoch: 6, i:  1599] avg mini-batch loss: 0.712
[epoch: 6, i:  1699] avg mini-batch loss: 0.665
[epoch: 6, i:  1799] avg mini-batch loss: 0.825
[epoch: 6, i:  1899] avg mini-batch loss: 0.652
[epoch: 6, i:  1999] avg mini-batch loss: 0.638
[epoch: 6, i:  2099] avg mini-batch loss: 0.782
[epoch: 6, i:  2199] avg mini-batch loss: 0.661
[epoch: 6, i:  2299] avg mini-batch loss: 0.703
[epoch: 6, i:  2399] avg mini-batch loss: 0.727
```

```
[epoch: 6, i:  2499] avg mini-batch loss: 0.735
[epoch: 6, i:  2599] avg mini-batch loss: 0.758
[epoch: 6, i:  2699] avg mini-batch loss: 0.741
[epoch: 6, i:  2799] avg mini-batch loss: 0.691
[epoch: 6, i:  2899] avg mini-batch loss: 0.697
[epoch: 6, i:  2999] avg mini-batch loss: 0.687
[epoch: 6, i:  3099] avg mini-batch loss: 0.682
[epoch: 6, i:  3199] avg mini-batch loss: 0.800
[epoch: 6, i:  3399] avg mini-batch loss: 0.749
[epoch: 6, i:  3499] avg mini-batch loss: 0.785
[epoch: 6, i:  3599] avg mini-batch loss: 0.664
[epoch: 6, i:  3699] avg mini-batch loss: 0.705
[epoch: 6, i:  3799] avg mini-batch loss: 0.791
[epoch: 6, i:  3899] avg mini-batch loss: 0.749
[epoch: 6, i:  3999] avg mini-batch loss: 0.731
[epoch: 6, i:  4099] avg mini-batch loss: 0.832
[epoch: 6, i:  4199] avg mini-batch loss: 0.733
[epoch: 6, i:  4299] avg mini-batch loss: 0.657
[epoch: 6, i:  4399] avg mini-batch loss: 0.712
[epoch: 6, i:  4499] avg mini-batch loss: 0.669
[epoch: 6, i:  4599] avg mini-batch loss: 0.789
[epoch: 6, i:  4699] avg mini-batch loss: 0.742
[epoch: 6, i:  4799] avg mini-batch loss: 0.783
[epoch: 6, i:  4899] avg mini-batch loss: 0.696
[epoch: 6, i:  4999] avg mini-batch loss: 0.781
[epoch: 6, i:  5099] avg mini-batch loss: 0.707
[epoch: 6, i:  5199] avg mini-batch loss: 0.793
[epoch: 6, i:  5299] avg mini-batch loss: 0.728
[epoch: 6, i:  5399] avg mini-batch loss: 0.713
[epoch: 6, i:  5499] avg mini-batch loss: 0.691
[epoch: 6, i:  5599] avg mini-batch loss: 0.720
[epoch: 6, i:  5699] avg mini-batch loss: 0.758
[epoch: 6, i:  5799] avg mini-batch loss: 0.765
[epoch: 6, i:  5899] avg mini-batch loss: 0.822
[epoch: 6, i:  5999] avg mini-batch loss: 0.696
[epoch: 6, i:  6099] avg mini-batch loss: 0.780
[epoch: 6, i:  6199] avg mini-batch loss: 0.818
[epoch: 6, i:  6299] avg mini-batch loss: 0.805
[epoch: 6, i:  6399] avg mini-batch loss: 0.671
[epoch: 6, i:  6499] avg mini-batch loss: 0.779
[epoch: 6, i:  6599] avg mini-batch loss: 0.814
[epoch: 6, i:  6699] avg mini-batch loss: 0.778
[epoch: 6, i:  6799] avg mini-batch loss: 0.644
[epoch: 6, i:  6899] avg mini-batch loss: 0.788
[epoch: 6, i:  6999] avg mini-batch loss: 0.706
[epoch: 6, i:  7099] avg mini-batch loss: 0.854
[epoch: 6, i:  7199] avg mini-batch loss: 0.808
[epoch: 6, i:  7299] avg mini-batch loss: 0.816
```

```
[epoch: 6, i:  7399] avg mini-batch loss: 0.749
[epoch: 6, i:  7499] avg mini-batch loss: 0.738
[epoch: 6, i:  7599] avg mini-batch loss: 0.759
[epoch: 6, i:  7699] avg mini-batch loss: 0.747
[epoch: 6, i:  7799] avg mini-batch loss: 0.729
[epoch: 6, i:  7899] avg mini-batch loss: 0.776
[epoch: 6, i:  7999] avg mini-batch loss: 0.772
[epoch: 6, i:  8099] avg mini-batch loss: 0.657
[epoch: 6, i:  8199] avg mini-batch loss: 0.740
[epoch: 6, i:  8299] avg mini-batch loss: 0.845
[epoch: 6, i:  8399] avg mini-batch loss: 0.721
[epoch: 6, i:  8499] avg mini-batch loss: 0.766
[epoch: 6, i:  8599] avg mini-batch loss: 0.726
[epoch: 6, i:  8699] avg mini-batch loss: 0.839
[epoch: 6, i:  8799] avg mini-batch loss: 0.837
[epoch: 6, i:  8899] avg mini-batch loss: 0.795
[epoch: 6, i:  8999] avg mini-batch loss: 0.777
[epoch: 6, i:  9099] avg mini-batch loss: 0.800
[epoch: 6, i:  9199] avg mini-batch loss: 0.754
[epoch: 6, i:  9299] avg mini-batch loss: 0.791
[epoch: 6, i:  9399] avg mini-batch loss: 0.681
[epoch: 6, i:  9499] avg mini-batch loss: 0.771
[epoch: 6, i:  9599] avg mini-batch loss: 0.699
[epoch: 6, i:  9699] avg mini-batch loss: 0.810
[epoch: 6, i:  9799] avg mini-batch loss: 0.717
[epoch: 6, i:  9899] avg mini-batch loss: 0.748
[epoch: 6, i:  9999] avg mini-batch loss: 0.775
[epoch: 6, i: 10099] avg mini-batch loss: 0.862
[epoch: 6, i: 10199] avg mini-batch loss: 0.751
[epoch: 6, i: 10299] avg mini-batch loss: 0.686
[epoch: 6, i: 10399] avg mini-batch loss: 0.771
[epoch: 6, i: 10499] avg mini-batch loss: 0.753
[epoch: 6, i: 10599] avg mini-batch loss: 0.763
[epoch: 6, i: 10699] avg mini-batch loss: 0.726
[epoch: 6, i: 10799] avg mini-batch loss: 0.694
[epoch: 6, i: 10899] avg mini-batch loss: 0.685
[epoch: 6, i: 10999] avg mini-batch loss: 0.770
[epoch: 6, i: 11099] avg mini-batch loss: 0.694
[epoch: 6, i: 11199] avg mini-batch loss: 0.773
[epoch: 6, i: 11299] avg mini-batch loss: 0.699
[epoch: 6, i: 11399] avg mini-batch loss: 0.690
[epoch: 6, i: 11499] avg mini-batch loss: 0.712
[epoch: 6, i: 11599] avg mini-batch loss: 0.801
[epoch: 6, i: 11699] avg mini-batch loss: 0.711
[epoch: 6, i: 11799] avg mini-batch loss: 0.723
[epoch: 6, i: 11899] avg mini-batch loss: 0.798
[epoch: 6, i: 11999] avg mini-batch loss: 0.807
[epoch: 6, i: 12099] avg mini-batch loss: 0.813
```

```
[epoch: 6, i: 12199] avg mini-batch loss: 0.757
[epoch: 6, i: 12299] avg mini-batch loss: 0.832
[epoch: 6, i: 12399] avg mini-batch loss: 0.747
[epoch: 6, i: 12499] avg mini-batch loss: 0.849
[epoch: 7, i:    99] avg mini-batch loss: 0.697
[epoch: 7, i:   199] avg mini-batch loss: 0.621
[epoch: 7, i:   299] avg mini-batch loss: 0.639
[epoch: 7, i:   399] avg mini-batch loss: 0.656
[epoch: 7, i:   499] avg mini-batch loss: 0.754
[epoch: 7, i:   599] avg mini-batch loss: 0.558
[epoch: 7, i:   699] avg mini-batch loss: 0.627
[epoch: 7, i:   799] avg mini-batch loss: 0.695
[epoch: 7, i:   899] avg mini-batch loss: 0.640
[epoch: 7, i:   999] avg mini-batch loss: 0.666
[epoch: 7, i:  1099] avg mini-batch loss: 0.564
[epoch: 7, i:  1199] avg mini-batch loss: 0.663
[epoch: 7, i:  1299] avg mini-batch loss: 0.677
[epoch: 7, i:  1399] avg mini-batch loss: 0.646
[epoch: 7, i:  1499] avg mini-batch loss: 0.644
[epoch: 7, i:  1599] avg mini-batch loss: 0.679
[epoch: 7, i:  1699] avg mini-batch loss: 0.656
[epoch: 7, i:  1799] avg mini-batch loss: 0.680
[epoch: 7, i:  1899] avg mini-batch loss: 0.658
[epoch: 7, i:  1999] avg mini-batch loss: 0.646
[epoch: 7, i:  2099] avg mini-batch loss: 0.696
[epoch: 7, i:  2199] avg mini-batch loss: 0.620
[epoch: 7, i:  2299] avg mini-batch loss: 0.609
[epoch: 7, i:  2399] avg mini-batch loss: 0.669
[epoch: 7, i:  2499] avg mini-batch loss: 0.646
[epoch: 7, i:  2599] avg mini-batch loss: 0.720
[epoch: 7, i:  2699] avg mini-batch loss: 0.660
[epoch: 7, i:  2799] avg mini-batch loss: 0.646
[epoch: 7, i:  2899] avg mini-batch loss: 0.760
[epoch: 7, i:  2999] avg mini-batch loss: 0.658
[epoch: 7, i:  3099] avg mini-batch loss: 0.690
[epoch: 7, i:  3199] avg mini-batch loss: 0.681
[epoch: 7, i:  3299] avg mini-batch loss: 0.622
[epoch: 7, i:  3399] avg mini-batch loss: 0.632
[epoch: 7, i:  3499] avg mini-batch loss: 0.659
[epoch: 7, i:  3599] avg mini-batch loss: 0.727
[epoch: 7, i:  3699] avg mini-batch loss: 0.740
[epoch: 7, i:  3799] avg mini-batch loss: 0.632
[epoch: 7, i:  3899] avg mini-batch loss: 0.722
[epoch: 7, i:  3999] avg mini-batch loss: 0.666
[epoch: 7, i:  4099] avg mini-batch loss: 0.671
[epoch: 7, i:  4199] avg mini-batch loss: 0.735
[epoch: 7, i:  4299] avg mini-batch loss: 0.577
[epoch: 7, i:  4399] avg mini-batch loss: 0.747
```

```
[epoch: 7, i:  4499] avg mini-batch loss: 0.686
[epoch: 7, i:  4599] avg mini-batch loss: 0.710
[epoch: 7, i:  4699] avg mini-batch loss: 0.767
[epoch: 7, i:  4799] avg mini-batch loss: 0.771
[epoch: 7, i:  4899] avg mini-batch loss: 0.646
[epoch: 7, i:  4999] avg mini-batch loss: 0.639
[epoch: 7, i:  5099] avg mini-batch loss: 0.755
[epoch: 7, i:  5199] avg mini-batch loss: 0.661
[epoch: 7, i:  5299] avg mini-batch loss: 0.665
[epoch: 7, i:  5399] avg mini-batch loss: 0.719
[epoch: 7, i:  5499] avg mini-batch loss: 0.702
[epoch: 7, i:  5599] avg mini-batch loss: 0.627
[epoch: 7, i:  5699] avg mini-batch loss: 0.654
[epoch: 7, i:  5799] avg mini-batch loss: 0.709
[epoch: 7, i:  5899] avg mini-batch loss: 0.690
[epoch: 7, i:  5999] avg mini-batch loss: 0.574
[epoch: 7, i:  6099] avg mini-batch loss: 0.651
[epoch: 7, i:  6199] avg mini-batch loss: 0.601
[epoch: 7, i:  6299] avg mini-batch loss: 0.663
[epoch: 7, i:  6399] avg mini-batch loss: 0.693
[epoch: 7, i:  6499] avg mini-batch loss: 0.672
[epoch: 7, i:  6599] avg mini-batch loss: 0.770
[epoch: 7, i:  6699] avg mini-batch loss: 0.671
[epoch: 7, i:  6799] avg mini-batch loss: 0.776
[epoch: 7, i:  6899] avg mini-batch loss: 0.690
[epoch: 7, i:  6999] avg mini-batch loss: 0.616
[epoch: 7, i:  7099] avg mini-batch loss: 0.704
[epoch: 7, i:  7199] avg mini-batch loss: 0.626
[epoch: 7, i:  7299] avg mini-batch loss: 0.712
[epoch: 7, i:  7399] avg mini-batch loss: 0.799
[epoch: 7, i:  7499] avg mini-batch loss: 0.682
[epoch: 7, i:  7599] avg mini-batch loss: 0.709
[epoch: 7, i:  7699] avg mini-batch loss: 0.690
[epoch: 7, i:  7799] avg mini-batch loss: 0.722
[epoch: 7, i:  7899] avg mini-batch loss: 0.619
[epoch: 7, i:  7999] avg mini-batch loss: 0.686
[epoch: 7, i:  8099] avg mini-batch loss: 0.761
[epoch: 7, i:  8199] avg mini-batch loss: 0.675
[epoch: 7, i:  8299] avg mini-batch loss: 0.772
[epoch: 7, i:  8399] avg mini-batch loss: 0.633
[epoch: 7, i:  8499] avg mini-batch loss: 0.682
[epoch: 7, i:  8599] avg mini-batch loss: 0.655
[epoch: 7, i:  8699] avg mini-batch loss: 0.743
[epoch: 7, i:  8799] avg mini-batch loss: 0.651
[epoch: 7, i:  8899] avg mini-batch loss: 0.686
[epoch: 7, i:  8999] avg mini-batch loss: 0.813
[epoch: 7, i:  9099] avg mini-batch loss: 0.747
[epoch: 7, i:  9199] avg mini-batch loss: 0.770
```

```
[epoch: 7, i:  9299] avg mini-batch loss: 0.682
[epoch: 7, i:  9399] avg mini-batch loss: 0.665
[epoch: 7, i:  9499] avg mini-batch loss: 0.695
[epoch: 7, i:  9599] avg mini-batch loss: 0.621
[epoch: 7, i:  9699] avg mini-batch loss: 0.649
[epoch: 7, i:  9799] avg mini-batch loss: 0.538
[epoch: 7, i:  9899] avg mini-batch loss: 0.763
[epoch: 7, i:  9999] avg mini-batch loss: 0.737
[epoch: 7, i: 10099] avg mini-batch loss: 0.679
[epoch: 7, i: 10199] avg mini-batch loss: 0.749
[epoch: 7, i: 10299] avg mini-batch loss: 0.608
[epoch: 7, i: 10399] avg mini-batch loss: 0.701
[epoch: 7, i: 10499] avg mini-batch loss: 0.725
[epoch: 7, i: 10599] avg mini-batch loss: 0.683
[epoch: 7, i: 10699] avg mini-batch loss: 0.733
[epoch: 7, i: 10799] avg mini-batch loss: 0.692
[epoch: 7, i: 10899] avg mini-batch loss: 0.650
[epoch: 7, i: 10999] avg mini-batch loss: 0.689
[epoch: 7, i: 11099] avg mini-batch loss: 0.715
[epoch: 7, i: 11199] avg mini-batch loss: 0.817
[epoch: 7, i: 11299] avg mini-batch loss: 0.662
[epoch: 7, i: 11399] avg mini-batch loss: 0.691
[epoch: 7, i: 11499] avg mini-batch loss: 0.699
[epoch: 7, i: 11599] avg mini-batch loss: 0.767
[epoch: 7, i: 11699] avg mini-batch loss: 0.755
[epoch: 7, i: 11799] avg mini-batch loss: 0.765
[epoch: 7, i: 11899] avg mini-batch loss: 0.789
[epoch: 7, i: 11999] avg mini-batch loss: 0.659
[epoch: 7, i: 12099] avg mini-batch loss: 0.652
[epoch: 7, i: 12199] avg mini-batch loss: 0.650
[epoch: 7, i: 12299] avg mini-batch loss: 0.651
[epoch: 7, i: 12399] avg mini-batch loss: 0.628
[epoch: 7, i: 12499] avg mini-batch loss: 0.759
[epoch: 8, i:    99] avg mini-batch loss: 0.678
[epoch: 8, i:   199] avg mini-batch loss: 0.578
[epoch: 8, i:   299] avg mini-batch loss: 0.500
[epoch: 8, i:   399] avg mini-batch loss: 0.594
[epoch: 8, i:   499] avg mini-batch loss: 0.523
[epoch: 8, i:   599] avg mini-batch loss: 0.561
[epoch: 8, i:   699] avg mini-batch loss: 0.532
[epoch: 8, i:   799] avg mini-batch loss: 0.593
[epoch: 8, i:   899] avg mini-batch loss: 0.618
[epoch: 8, i:   999] avg mini-batch loss: 0.575
[epoch: 8, i:  1099] avg mini-batch loss: 0.568
[epoch: 8, i:  1199] avg mini-batch loss: 0.554
[epoch: 8, i:  1299] avg mini-batch loss: 0.516
[epoch: 8, i:  1399] avg mini-batch loss: 0.687
[epoch: 8, i:  1499] avg mini-batch loss: 0.544
```

```
[epoch: 8, i:  1599] avg mini-batch loss: 0.610
[epoch: 8, i:  1699] avg mini-batch loss: 0.539
[epoch: 8, i:  1799] avg mini-batch loss: 0.590
[epoch: 8, i:  1899] avg mini-batch loss: 0.599
[epoch: 8, i:  1999] avg mini-batch loss: 0.571
[epoch: 8, i:  2099] avg mini-batch loss: 0.634
[epoch: 8, i:  2199] avg mini-batch loss: 0.553
[epoch: 8, i:  2299] avg mini-batch loss: 0.584
[epoch: 8, i:  2399] avg mini-batch loss: 0.537
[epoch: 8, i:  2499] avg mini-batch loss: 0.448
[epoch: 8, i:  2599] avg mini-batch loss: 0.667
[epoch: 8, i:  2699] avg mini-batch loss: 0.625
[epoch: 8, i:  2799] avg mini-batch loss: 0.579
[epoch: 8, i:  2899] avg mini-batch loss: 0.627
[epoch: 8, i:  2999] avg mini-batch loss: 0.595
[epoch: 8, i:  3099] avg mini-batch loss: 0.587
[epoch: 8, i:  3199] avg mini-batch loss: 0.569
[epoch: 8, i:  3299] avg mini-batch loss: 0.679
[epoch: 8, i:  3399] avg mini-batch loss: 0.649
[epoch: 8, i:  3499] avg mini-batch loss: 0.616
[epoch: 8, i:  3599] avg mini-batch loss: 0.598
[epoch: 8, i:  3699] avg mini-batch loss: 0.630
[epoch: 8, i:  3799] avg mini-batch loss: 0.524
[epoch: 8, i:  3899] avg mini-batch loss: 0.671
[epoch: 8, i:  3999] avg mini-batch loss: 0.671
[epoch: 8, i:  4099] avg mini-batch loss: 0.627
[epoch: 8, i:  4199] avg mini-batch loss: 0.675
[epoch: 8, i:  4299] avg mini-batch loss: 0.652
[epoch: 8, i:  4399] avg mini-batch loss: 0.625
[epoch: 8, i:  4499] avg mini-batch loss: 0.636
[epoch: 8, i:  4599] avg mini-batch loss: 0.628
[epoch: 8, i:  4699] avg mini-batch loss: 0.612
[epoch: 8, i:  4799] avg mini-batch loss: 0.579
[epoch: 8, i:  4899] avg mini-batch loss: 0.632
[epoch: 8, i:  4999] avg mini-batch loss: 0.623
[epoch: 8, i:  5099] avg mini-batch loss: 0.575
[epoch: 8, i:  5199] avg mini-batch loss: 0.545
[epoch: 8, i:  5299] avg mini-batch loss: 0.682
[epoch: 8, i:  5399] avg mini-batch loss: 0.650
[epoch: 8, i:  5499] avg mini-batch loss: 0.558
[epoch: 8, i:  5599] avg mini-batch loss: 0.606
[epoch: 8, i:  5699] avg mini-batch loss: 0.625
[epoch: 8, i:  5799] avg mini-batch loss: 0.657
[epoch: 8, i:  5899] avg mini-batch loss: 0.610
[epoch: 8, i:  5999] avg mini-batch loss: 0.659
[epoch: 8, i:  6099] avg mini-batch loss: 0.703
[epoch: 8, i:  6199] avg mini-batch loss: 0.656
[epoch: 8, i:  6299] avg mini-batch loss: 0.646
```

```
[epoch: 8, i:  6399] avg mini-batch loss: 0.695
[epoch: 8, i:  6499] avg mini-batch loss: 0.673
[epoch: 8, i:  6599] avg mini-batch loss: 0.698
[epoch: 8, i:  6699] avg mini-batch loss: 0.604
[epoch: 8, i:  6799] avg mini-batch loss: 0.609
[epoch: 8, i:  6899] avg mini-batch loss: 0.653
[epoch: 8, i:  6999] avg mini-batch loss: 0.638
[epoch: 8, i:  7099] avg mini-batch loss: 0.605
[epoch: 8, i:  7199] avg mini-batch loss: 0.595
[epoch: 8, i:  7299] avg mini-batch loss: 0.673
[epoch: 8, i:  7399] avg mini-batch loss: 0.668
[epoch: 8, i:  7499] avg mini-batch loss: 0.713
[epoch: 8, i:  7599] avg mini-batch loss: 0.550
[epoch: 8, i:  7699] avg mini-batch loss: 0.739
[epoch: 8, i:  7799] avg mini-batch loss: 0.696
[epoch: 8, i:  7899] avg mini-batch loss: 0.600
[epoch: 8, i:  7999] avg mini-batch loss: 0.686
[epoch: 8, i:  8099] avg mini-batch loss: 0.757
[epoch: 8, i:  8199] avg mini-batch loss: 0.598
[epoch: 8, i:  8299] avg mini-batch loss: 0.539
[epoch: 8, i:  8399] avg mini-batch loss: 0.650
[epoch: 8, i:  8499] avg mini-batch loss: 0.614
[epoch: 8, i:  8599] avg mini-batch loss: 0.605
[epoch: 8, i:  8699] avg mini-batch loss: 0.720
[epoch: 8, i:  8799] avg mini-batch loss: 0.617
[epoch: 8, i:  8899] avg mini-batch loss: 0.672
[epoch: 8, i:  8999] avg mini-batch loss: 0.673
[epoch: 8, i:  9099] avg mini-batch loss: 0.540
[epoch: 8, i:  9199] avg mini-batch loss: 0.689
[epoch: 8, i:  9299] avg mini-batch loss: 0.691
[epoch: 8, i:  9399] avg mini-batch loss: 0.687
[epoch: 8, i:  9499] avg mini-batch loss: 0.660
[epoch: 8, i:  9599] avg mini-batch loss: 0.726
[epoch: 8, i:  9699] avg mini-batch loss: 0.604
[epoch: 8, i:  9799] avg mini-batch loss: 0.668
[epoch: 8, i:  9899] avg mini-batch loss: 0.621
[epoch: 8, i:  9999] avg mini-batch loss: 0.592
[epoch: 8, i: 10099] avg mini-batch loss: 0.642
[epoch: 8, i: 10199] avg mini-batch loss: 0.715
[epoch: 8, i: 10299] avg mini-batch loss: 0.699
[epoch: 8, i: 10399] avg mini-batch loss: 0.724
[epoch: 8, i: 10499] avg mini-batch loss: 0.602
[epoch: 8, i: 10599] avg mini-batch loss: 0.696
[epoch: 8, i: 10699] avg mini-batch loss: 0.667
[epoch: 8, i: 10799] avg mini-batch loss: 0.717
[epoch: 8, i: 10899] avg mini-batch loss: 0.637
[epoch: 8, i: 10999] avg mini-batch loss: 0.743
[epoch: 8, i: 11099] avg mini-batch loss: 0.632
```

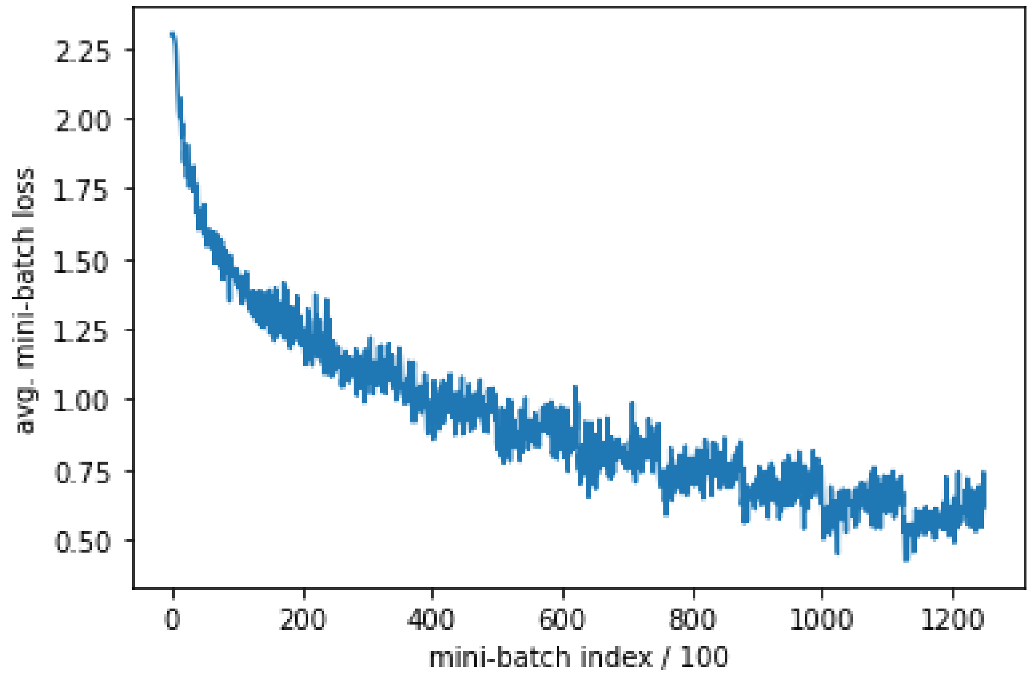
```
[epoch: 8, i: 11199] avg mini-batch loss: 0.576
[epoch: 8, i: 11299] avg mini-batch loss: 0.713
[epoch: 8, i: 11399] avg mini-batch loss: 0.658
[epoch: 8, i: 11499] avg mini-batch loss: 0.593
[epoch: 8, i: 11599] avg mini-batch loss: 0.616
[epoch: 8, i: 11699] avg mini-batch loss: 0.650
[epoch: 8, i: 11799] avg mini-batch loss: 0.685
[epoch: 8, i: 11899] avg mini-batch loss: 0.639
[epoch: 8, i: 11999] avg mini-batch loss: 0.651
[epoch: 8, i: 12099] avg mini-batch loss: 0.633
[epoch: 8, i: 12199] avg mini-batch loss: 0.682
[epoch: 8, i: 12299] avg mini-batch loss: 0.725
[epoch: 8, i: 12399] avg mini-batch loss: 0.616
[epoch: 8, i: 12499] avg mini-batch loss: 0.668
[epoch: 9, i:    99] avg mini-batch loss: 0.527
[epoch: 9, i:   199] avg mini-batch loss: 0.575
[epoch: 9, i:   299] avg mini-batch loss: 0.515
[epoch: 9, i:   399] avg mini-batch loss: 0.425
[epoch: 9, i:   499] avg mini-batch loss: 0.441
[epoch: 9, i:   599] avg mini-batch loss: 0.542
[epoch: 9, i:   699] avg mini-batch loss: 0.427
[epoch: 9, i:   799] avg mini-batch loss: 0.484
[epoch: 9, i:   899] avg mini-batch loss: 0.553
[epoch: 9, i:   999] avg mini-batch loss: 0.531
[epoch: 9, i:  1099] avg mini-batch loss: 0.552
[epoch: 9, i:  1199] avg mini-batch loss: 0.549
[epoch: 9, i:  1299] avg mini-batch loss: 0.548
[epoch: 9, i:  1399] avg mini-batch loss: 0.514
[epoch: 9, i:  1499] avg mini-batch loss: 0.533
[epoch: 9, i:  1599] avg mini-batch loss: 0.554
[epoch: 9, i:  1699] avg mini-batch loss: 0.454
[epoch: 9, i:  1799] avg mini-batch loss: 0.523
[epoch: 9, i:  1899] avg mini-batch loss: 0.521
[epoch: 9, i:  1999] avg mini-batch loss: 0.606
[epoch: 9, i:  2099] avg mini-batch loss: 0.594
[epoch: 9, i:  2199] avg mini-batch loss: 0.521
[epoch: 9, i:  2299] avg mini-batch loss: 0.513
[epoch: 9, i:  2399] avg mini-batch loss: 0.512
[epoch: 9, i:  2499] avg mini-batch loss: 0.526
[epoch: 9, i:  2599] avg mini-batch loss: 0.551
[epoch: 9, i:  2699] avg mini-batch loss: 0.540
[epoch: 9, i:  2799] avg mini-batch loss: 0.599
[epoch: 9, i:  2899] avg mini-batch loss: 0.556
[epoch: 9, i:  2999] avg mini-batch loss: 0.529
[epoch: 9, i:  3099] avg mini-batch loss: 0.536
[epoch: 9, i:  3199] avg mini-batch loss: 0.575
[epoch: 9, i:  3299] avg mini-batch loss: 0.560
[epoch: 9, i:  3399] avg mini-batch loss: 0.517
```

```
[epoch: 9, i:  3499] avg mini-batch loss: 0.616
[epoch: 9, i:  3599] avg mini-batch loss: 0.581
[epoch: 9, i:  3699] avg mini-batch loss: 0.541
[epoch: 9, i:  3799] avg mini-batch loss: 0.555
[epoch: 9, i:  3899] avg mini-batch loss: 0.536
[epoch: 9, i:  3999] avg mini-batch loss: 0.592
[epoch: 9, i:  4099] avg mini-batch loss: 0.566
[epoch: 9, i:  4199] avg mini-batch loss: 0.552
[epoch: 9, i:  4299] avg mini-batch loss: 0.599
[epoch: 9, i:  4399] avg mini-batch loss: 0.550
[epoch: 9, i:  4499] avg mini-batch loss: 0.589
[epoch: 9, i:  4599] avg mini-batch loss: 0.527
[epoch: 9, i:  4699] avg mini-batch loss: 0.579
[epoch: 9, i:  4799] avg mini-batch loss: 0.502
[epoch: 9, i:  4899] avg mini-batch loss: 0.565
[epoch: 9, i:  4999] avg mini-batch loss: 0.609
[epoch: 9, i:  5099] avg mini-batch loss: 0.505
[epoch: 9, i:  5199] avg mini-batch loss: 0.606
[epoch: 9, i:  5299] avg mini-batch loss: 0.604
[epoch: 9, i:  5399] avg mini-batch loss: 0.571
[epoch: 9, i:  5499] avg mini-batch loss: 0.559
[epoch: 9, i:  5599] avg mini-batch loss: 0.541
[epoch: 9, i:  5699] avg mini-batch loss: 0.550
[epoch: 9, i:  5799] avg mini-batch loss: 0.569
[epoch: 9, i:  5899] avg mini-batch loss: 0.552
[epoch: 9, i:  5999] avg mini-batch loss: 0.576
[epoch: 9, i:  6099] avg mini-batch loss: 0.595
[epoch: 9, i:  6199] avg mini-batch loss: 0.535
[epoch: 9, i:  6299] avg mini-batch loss: 0.631
[epoch: 9, i:  6399] avg mini-batch loss: 0.603
[epoch: 9, i:  6499] avg mini-batch loss: 0.534
[epoch: 9, i:  6599] avg mini-batch loss: 0.722
[epoch: 9, i:  6699] avg mini-batch loss: 0.620
[epoch: 9, i:  6799] avg mini-batch loss: 0.574
[epoch: 9, i:  6899] avg mini-batch loss: 0.531
[epoch: 9, i:  6999] avg mini-batch loss: 0.519
[epoch: 9, i:  7099] avg mini-batch loss: 0.512
[epoch: 9, i:  7199] avg mini-batch loss: 0.607
[epoch: 9, i:  7299] avg mini-batch loss: 0.593
[epoch: 9, i:  7399] avg mini-batch loss: 0.546
[epoch: 9, i:  7499] avg mini-batch loss: 0.616
[epoch: 9, i:  7599] avg mini-batch loss: 0.520
[epoch: 9, i:  7699] avg mini-batch loss: 0.633
[epoch: 9, i:  7799] avg mini-batch loss: 0.649
[epoch: 9, i:  7899] avg mini-batch loss: 0.488
[epoch: 9, i:  7999] avg mini-batch loss: 0.591
[epoch: 9, i:  8099] avg mini-batch loss: 0.618
[epoch: 9, i:  8199] avg mini-batch loss: 0.643
```

```
[epoch: 9, i:  8299] avg mini-batch loss: 0.540
[epoch: 9, i:  8399] avg mini-batch loss: 0.579
[epoch: 9, i:  8499] avg mini-batch loss: 0.739
[epoch: 9, i:  8599] avg mini-batch loss: 0.550
[epoch: 9, i:  8699] avg mini-batch loss: 0.580
[epoch: 9, i:  8799] avg mini-batch loss: 0.564
[epoch: 9, i:  8899] avg mini-batch loss: 0.614
[epoch: 9, i:  8999] avg mini-batch loss: 0.562
[epoch: 9, i:  9099] avg mini-batch loss: 0.602
[epoch: 9, i:  9199] avg mini-batch loss: 0.608
[epoch: 9, i:  9299] avg mini-batch loss: 0.617
[epoch: 9, i:  9399] avg mini-batch loss: 0.609
[epoch: 9, i:  9499] avg mini-batch loss: 0.674
[epoch: 9, i:  9599] avg mini-batch loss: 0.622
[epoch: 9, i:  9699] avg mini-batch loss: 0.648
[epoch: 9, i:  9799] avg mini-batch loss: 0.588
[epoch: 9, i:  9899] avg mini-batch loss: 0.716
[epoch: 9, i:  9999] avg mini-batch loss: 0.622
[epoch: 9, i: 10099] avg mini-batch loss: 0.548
[epoch: 9, i: 10199] avg mini-batch loss: 0.685
[epoch: 9, i: 10299] avg mini-batch loss: 0.573
[epoch: 9, i: 10399] avg mini-batch loss: 0.553
[epoch: 9, i: 10499] avg mini-batch loss: 0.576
[epoch: 9, i: 10599] avg mini-batch loss: 0.545
[epoch: 9, i: 10699] avg mini-batch loss: 0.676
[epoch: 9, i: 10799] avg mini-batch loss: 0.575
[epoch: 9, i: 10899] avg mini-batch loss: 0.574
[epoch: 9, i: 10999] avg mini-batch loss: 0.624
[epoch: 9, i: 11099] avg mini-batch loss: 0.605
[epoch: 9, i: 11199] avg mini-batch loss: 0.527
[epoch: 9, i: 11299] avg mini-batch loss: 0.631
[epoch: 9, i: 11399] avg mini-batch loss: 0.537
[epoch: 9, i: 11499] avg mini-batch loss: 0.573
[epoch: 9, i: 11599] avg mini-batch loss: 0.620
[epoch: 9, i: 11699] avg mini-batch loss: 0.690
[epoch: 9, i: 11799] avg mini-batch loss: 0.588
[epoch: 9, i: 11899] avg mini-batch loss: 0.632
[epoch: 9, i: 11999] avg mini-batch loss: 0.634
[epoch: 9, i: 12099] avg mini-batch loss: 0.545
[epoch: 9, i: 12199] avg mini-batch loss: 0.623
[epoch: 9, i: 12299] avg mini-batch loss: 0.668
[epoch: 9, i: 12399] avg mini-batch loss: 0.739
[epoch: 9, i: 12499] avg mini-batch loss: 0.614
Finished Training.
```

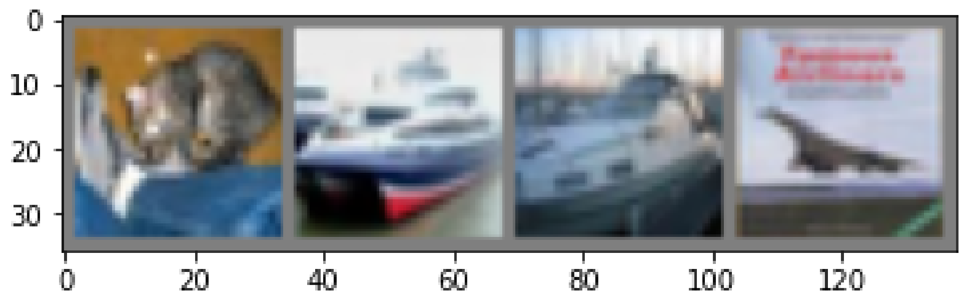
Training Loss Curve

[8]:
```
plt.plot(avg_losses)
plt.xlabel('mini-batch index / {}'.format(print_freq))
plt.ylabel('avg. mini-batch loss')
plt.show()
```



Evaluate on Test Dataset

[9]:
```
# Check several images.
dataiter = iter(testloader)
images, labels = next(dataiter)
imshow(torchvision.utils.make_grid(images))
print('GroundTruth: ', ' '.join('%5s' % classes[labels[j]] for j in range(4)))
outputs = net(images.to(device))
_, predicted = torch.max(outputs, 1)

print('Predicted: ', ' '.join('%5s' % classes[predicted[j]]
                              for j in range(4)))
```



```
GroundTruth:    cat  ship  ship plane
Predicted:    cat  ship  ship plane
```

[10]:

```python
# Get test accuracy.
correct = 0
total = 0
with torch.no_grad():
    for data in testloader:
        images, labels = data
        images, labels = images.to(device), labels.to(device)
        outputs = net(images)
        _, predicted = torch.max(outputs.data, 1)
        total += labels.size(0)
        correct += (predicted == labels).sum().item()

print('Accuracy of the network on the 10000 test images: %d %%' % (
    100 * correct / total))
```

```
Accuracy of the network on the 10000 test images: 67 %
```

[11]:

```python
# Get test accuracy for each class.
class_correct = list(0. for i in range(10))
class_total = list(0. for i in range(10))
with torch.no_grad():
    for data in testloader:
        images, labels = data
        images, labels = images.to(device), labels.to(device)
        outputs = net(images)
        _, predicted = torch.max(outputs, 1)
        c = (predicted == labels).squeeze()
        for i in range(4):
            label = labels[i]
            class_correct[label] += c[i].item()
            class_total[label] += 1
```

```python
for i in range(10):
    print('Accuracy of %5s : %2d %%' % (
        classes[i], 100 * class_correct[i] / class_total[i]))
```

```
Accuracy of plane : 71 %
Accuracy of   car : 80 %
Accuracy of  bird : 55 %
Accuracy of   cat : 49 %
Accuracy of  deer : 63 %
Accuracy of   dog : 60 %
Accuracy of  frog : 68 %
Accuracy of horse : 73 %
Accuracy of  ship : 75 %
Accuracy of truck : 72 %
```