

gan_ipynb

April 21, 2025

We would like to acknowledge Stanford University's CS231n on which we based the development of this project.

```
[12]: USE_COLAB = True

import sys
import os

if USE_COLAB:
    # This mounts your Google Drive to the Colab VM.
    from google.colab import drive
    drive.mount('/content/drive')

    # TODO: Enter the foldername in your Drive where you have saved the unzipped
    # project folder, e.g. '239AS.2/project1/gan'
    FOLDERNAME = None
    assert FOLDERNAME is not None, "[!] Enter the foldername."

    # Now that we've mounted your Drive, this ensures that
    # the Python interpreter of the Colab VM can load
    # python files from within it.

    sys.path.append('/content/drive/My Drive/{}'.format(FOLDERNAME))

    %cd /content/drive/My\ Drive/$FOLDERNAME

# You can use os.listdir() to find the folder
# Google drive is...weird.
print('cwd', os.getcwd())
```

```
-----
ModuleNotFoundError                                Traceback (most recent call last)
Cell In[12], line 8
      4 import os
      6 if USE_COLAB:
      7     # This mounts your Google Drive to the Colab VM.
----> 8     from google.colab import drive
```

```

9     drive.mount('/content/drive')
11    # TODO: Enter the foldername in your Drive where you have saved the
↪unzipped
12    # project folder, e.g. '239AS.2/project1/gan'

```

`ModuleNotFoundError: No module named 'google.colab'`

TO FIND YOUR FILES IN COLAB:

Click the folder on the left sidebar (content->)drive->MyDrive->WhereverYouPutTheseFiles

Double clicking gan.py will open an editor on the right.

0.1 Enabling GPU:

Runtime > Change runtime type > Hardware accelerator: select T4 GPU or v2-8 TPU. You may run out of compute allowance at some point. **When you will be away for a long period of time, select Runtime->Disconnect and delete runtime! Make sure to preserve your compute allowance!**

1 Generative Adversarial Networks (GANs)

In C147/C247, all the applications of neural networks that we have explored have been **discriminative models** that take an input and are trained to produce a labeled output. In this notebook, we will expand our repertoire, and build **generative models** using neural networks. Specifically, we will learn how to build models which generate novel images that resemble a set of training images.

1.0.1 What is a GAN?

In 2014, [Goodfellow et al.](#) presented a method for training generative models called Generative Adversarial Networks (GANs for short). In a GAN, we build two different neural networks. Our first network is a traditional classification network, called the **discriminator**. We will train the discriminator to take images and classify them as being real (belonging to the training set) or fake (not present in the training set). Our other network, called the **generator**, will take random noise as input and transform it using a neural network to produce images. The goal of the generator is to fool the discriminator into thinking the images it produced are real.

We can think of this back and forth process of the generator (G) trying to fool the discriminator (D) and the discriminator trying to correctly classify real vs. fake as a minimax game:

$$\underset{G}{\text{minimize}} \underset{D}{\text{maximize}} \mathbb{E}_{x \sim p_{\text{data}}} [\log D(x)] + \mathbb{E}_{z \sim p(z)} [\log (1 - D(G(z)))]$$

where $z \sim p(z)$ are the random noise samples, $G(z)$ are the generated images using the neural network generator G , and D is the output of the discriminator, specifying the probability of an input being real. In [Goodfellow et al.](#), they analyze this minimax game and show how it relates to minimizing the Jensen-Shannon divergence between the training data distribution and the generated samples from G .

To optimize this minimax game, we will alternate between taking gradient *descent* steps on the objective for G and gradient *ascent* steps on the objective for D : 1. update the **generator** (G) to minimize the probability of the **discriminator making the correct choice**. 2. update the **discriminator** (D) to maximize the probability of the **discriminator making the correct choice**.

While these updates are useful for analysis, they do not perform well in practice. Instead, we will use a different objective when we update the generator: maximize the probability of the **discriminator making the incorrect choice**. This small change helps to alleviate problems with the generator gradient vanishing when the discriminator is confident. This is the standard update used in most GAN papers and was used in the original paper from [Goodfellow et al.](#).

In this assignment, we will alternate the following updates: 1. Update the generator (G) to maximize the probability of the discriminator making the incorrect choice on generated data:

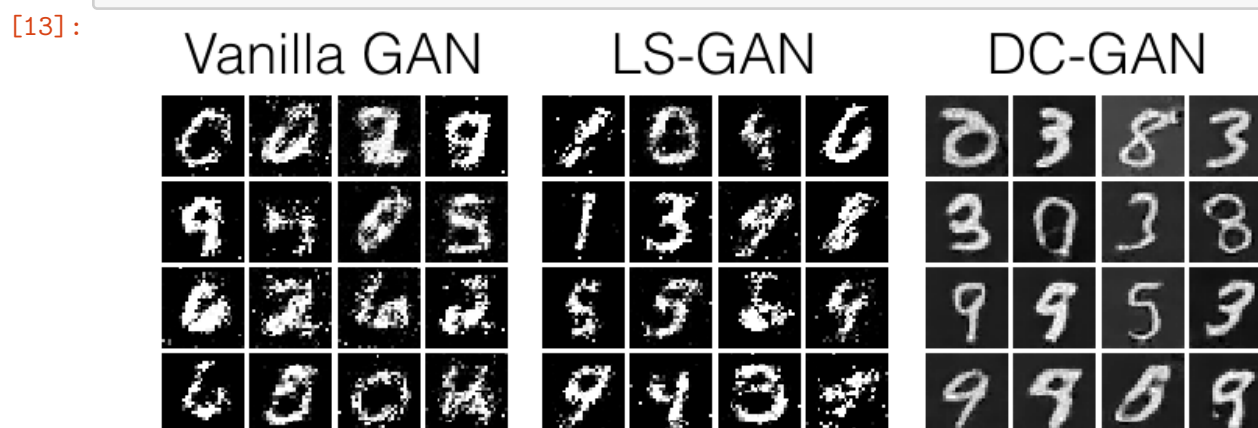
$$\underset{G}{\text{maximize}} \mathbb{E}_{z \sim p(z)} [\log D(G(z))]$$

2. Update the discriminator (D), to maximize the probability of the discriminator making the correct choice on real and generated data:

$$\underset{D}{\text{maximize}} \mathbb{E}_{x \sim p_{\text{data}}} [\log D(x)] + \mathbb{E}_{z \sim p(z)} [\log (1 - D(G(z)))]$$

Here's an example of what your outputs from the 3 different models you're going to train should look like. Note that GANs are sometimes finicky, so your outputs might not look exactly like this. This is just meant to be a *rough* guideline of the kind of quality you can expect:

```
[13]: # Run this cell to see sample outputs.
from IPython.display import Image
Image('nnd12/gan_outputs.png')
```



```
[14]: # Setup cell.
import numpy as np
import torch
import torch.nn as nn
```

```

from torch.nn import init
import torchvision
import torchvision.transforms as T
import torch.optim as optim
from torch.utils.data import DataLoader
from torch.utils.data import sampler
import torchvision.datasets as dset
import matplotlib.pyplot as plt
import matplotlib.gridspec as gridspec
from gan import preprocess_img, deprocess_img, rel_error, count_params,
    ↪ChunkSampler

%matplotlib inline
plt.rcParams['figure.figsize'] = (10.0, 8.0) # Set default size of plots.
plt.rcParams['image.interpolation'] = 'nearest'
plt.rcParams['image.cmap'] = 'gray'

%load_ext autoreload
%autoreload 2

def show_images(images):
    images = np.reshape(images, [images.shape[0], -1]) # Images reshape to
    ↪(batch_size, D).
    sqrtn = int(np.ceil(np.sqrt(images.shape[0])))
    sqrtimg = int(np.ceil(np.sqrt(images.shape[1])))

    fig = plt.figure(figsize=(sqrtn, sqrtn))
    gs = gridspec.GridSpec(sqrtn, sqrtn)
    gs.update(wspace=0.05, hspace=0.05)

    for i, img in enumerate(images):
        ax = plt.subplot(gs[i])
        plt.axis('off')
        ax.set_xticklabels([])
        ax.set_yticklabels([])
        ax.set_aspect('equal')
        plt.imshow(img.reshape([sqrtimg, sqrtimg]))
    return

answers = dict(np.load('nndl2/gan-checks.npz'))
dtype = torch.cuda.FloatTensor if torch.cuda.is_available() else torch.
    ↪FloatTensor
print(dtype)

```

The autoreload extension is already loaded. To reload it, use:

```

%reload_ext autoreload
<class 'torch.FloatTensor'>

```

1.1 Dataset

GANs are notoriously finicky with hyperparameters, and also require many training epochs. In order to make this assignment approachable, we will be working on the MNIST dataset, which is 60,000 training and 10,000 test images. Each picture contains a centered image of white digit on black background (0 through 9). This was one of the first datasets used to train convolutional neural networks and it is fairly easy – a standard CNN model can easily exceed 99% accuracy.

To simplify our code here, we will use the PyTorch MNIST wrapper, which downloads and loads the MNIST dataset. See the [documentation](#) for more information about the interface. The default parameters will take 5,000 of the training examples and place them into a validation dataset. The data will be saved into a folder called MNIST.

```
[15]: NUM_TRAIN = 50000
      NUM_VAL = 5000

      NOISE_DIM = 96
      batch_size = 128

      mnist_train = dset.MNIST(
          './nnd12',
          train=True,
          download=True,
          transform=T.ToTensor()
      )
      loader_train = DataLoader(
          mnist_train,
          batch_size=batch_size,
          sampler=ChunkSampler(NUM_TRAIN, 0)
      )

      mnist_val = dset.MNIST(
          './nnd12',
          train=True,
          download=True,
          transform=T.ToTensor()
      )
      loader_val = DataLoader(
          mnist_val,
          batch_size=batch_size,
          sampler=ChunkSampler(NUM_VAL, NUM_TRAIN)
      )

      iterator = iter(loader_train)
      imgs, labels = next(iterator)
      imgs = imgs.view(batch_size, 784).numpy().squeeze()
      show_images(imgs)
```

1.2 Random Noise (1 point)

Generate uniform noise from -1 to 1 with shape `[batch_size, dim]`.

Implement `sample_noise` in `gan.py`.

Hint: use `torch.rand`.

Make sure noise is the correct shape and type:

```
[16]: from gan import sample_noise

def test_sample_noise():
    batch_size = 3
    dim = 4
    torch.manual_seed(231)
    z = sample_noise(batch_size, dim)
    np_z = z.cpu().numpy()
    assert np_z.shape == (batch_size, dim)
    assert torch.is_tensor(z)
    assert np.all(np_z >= -1.0) and np.all(np_z <= 1.0)
    assert np.any(np_z < 0.0) and np.any(np_z > 0.0)
    print('All tests passed!')

test_sample_noise()
```

All tests passed!

1.3 Flatten

We provide an `Unflatten`, which you might want to use when implementing the convolutional generator. We also provide a weight initializer (and call it for you) that uses Xavier initialization instead of PyTorch's uniform default.

```
[17]: from gan import Flatten, Unflatten, initialize_weights
```

2 Discriminator (1 point)

Our first step is to build a discriminator. Fill in the architecture as part of the `nn.Sequential` constructor in the function below. All fully connected layers should include bias terms. The architecture is: * Fully connected layer with input size 784 and output size 256 * LeakyReLU with alpha 0.01 * Fully connected layer with input_size 256 and output size 256 * LeakyReLU with alpha 0.01 * Fully connected layer with input size 256 and output size 1

Recall that the Leaky ReLU nonlinearity computes $f(x) = \max(\alpha x, x)$ for some fixed constant α ; for the LeakyReLU nonlinearities in the architecture above we set $\alpha = 0.01$.

The output of the discriminator should have shape `[batch_size, 1]`, and contain real numbers corresponding to the scores that each of the `batch_size` inputs is a real image.

Implement `discriminator` in `gan.py`

Test to make sure the number of parameters in the discriminator is correct:

```
[18]: from gan import discriminator

def test_discriminator(true_count=267009):
    model = discriminator()
    cur_count = count_params(model)
    if cur_count != true_count:
        print('Incorrect number of parameters in discriminator. Check your_
↪architecture.')
    else:
        print('Correct number of parameters in discriminator.')

test_discriminator()
```

Correct number of parameters in discriminator.

3 Generator (1 point)

Now to build the generator network: * Fully connected layer from noise_dim to 1024 * ReLU * Fully connected layer with size 1024 * ReLU * Fully connected layer with size 784 * TanH (to clip the image to be in the range of [-1,1])

All fully connected layers should include bias terms. Implement **generator** in **gan.py**

Test to make sure the number of parameters in the generator is correct:

```
[19]: from gan import generator

def test_generator(true_count=1858320):
    model = generator(4)
    cur_count = count_params(model)
    if cur_count != true_count:
        print('Incorrect number of parameters in generator. Check your_
↪architecture.')
    else:
        print('Correct number of parameters in generator.')

test_generator()
```

Correct number of parameters in generator.

4 GAN Loss (2 points)

Compute the generator and discriminator loss. The generator loss is:

$$\ell_G = -\mathbb{E}_{z \sim p(z)} [\log D(G(z))]$$

and the discriminator loss is:

$$\ell_D = -\mathbb{E}_{x \sim p_{\text{data}}} [\log D(x)] - \mathbb{E}_{z \sim p(z)} [\log (1 - D(G(z)))]$$

Note that these are negated from the equations presented earlier as we will be *minimizing* these losses.

HINTS: You should use the `bce_loss` function defined below to compute the binary cross entropy loss which is needed to compute the log probability of the true label given the logits output from the discriminator. Given a score $s \in \mathbb{R}$ and a label $y \in \{0, 1\}$, the binary cross entropy loss is

$$bce(s, y) = -y * \log(s) - (1 - y) * \log(1 - s)$$

A naive implementation of this formula can be numerically unstable, so we have provided a numerically stable implementation that relies on PyTorch's `nn.BCEWithLogitsLoss`.

You will also need to compute labels corresponding to real or fake and use the logit arguments to determine their size. Make sure you cast these labels to the correct data type using the global `dtype` variable for compatibility with `BCEWithLogitsLoss`, which uses floating point data types.

```
true_labels = torch.ones(size).type(dtype)
```

Instead of computing the expectation of $\log D(G(z))$, $\log D(x)$ and $\log(1 - D(G(z)))$, we will be averaging over elements of the minibatch. This is taken care of in `bce_loss` which combines the loss by averaging. **Hint:** Do NOT concatenate real and fake losses. Sum the two (empirical) expectations instead.

Implement `discriminator_loss` and `generator_loss` in `gan.py`

Test your generator and discriminator loss. You should see errors $< 1e-7$.

```
[20]: from gan import bce_loss, discriminator_loss, generator_loss

def test_discriminator_loss(logits_real, logits_fake, d_loss_true):
    d_loss = discriminator_loss(torch.Tensor(logits_real).type(dtype),
                                torch.Tensor(logits_fake).type(dtype)).cpu().
    ↪numpy()
    print("Maximum error in d_loss: %g"%rel_error(d_loss_true, d_loss))

test_discriminator_loss(
    answers['logits_real'],
    answers['logits_fake'],
    answers['d_loss_true']
)
```

Maximum error in d_loss: 3.97058e-09

```
[21]: def test_generator_loss(logits_fake, g_loss_true):
    g_loss = generator_loss(torch.Tensor(logits_fake).type(dtype)).cpu().numpy()
    print("Maximum error in g_loss: %g"%rel_error(g_loss_true, g_loss))

test_generator_loss(
    answers['logits_fake'],
    answers['g_loss_true']
)
```


Maximum error in g_loss: 4.4518e-09

5 Optimizing Our Loss (1 point)

Make a function that returns an `optim.Adam` optimizer for the given model with a `1e-3` learning rate, `beta1=0.5`, `beta2=0.999`. You'll use this to construct optimizers for the generators and discriminators for the rest of the notebook.

Implement `get_optimizer` in `gan.py`

6 Training a GAN!

We provide you the main training loop. You won't need to change `run_a_gan` in `gan.py`, but we encourage you to read through it for your own understanding. If you train with the CPU, it takes about 7 minutes. If you train with the T4 GPU, it takes about 1 minute and 30 seconds.

If it doesn't work the first time, check the discriminator: Did you use `Flatten()`?

```
[24]: from gan import get_optimizer, run_a_gan

# Make the discriminator
D = discriminator().type(dtype)

# Make the generator
G = generator().type(dtype)

# Use the function you wrote earlier to get optimizers for the Discriminator
  ↪and the Generator
D_solver = get_optimizer(D)
G_solver = get_optimizer(G)

# Run it!
images = run_a_gan(
    D,
    G,
    D_solver,
    G_solver,
    discriminator_loss,
    generator_loss,
    loader_train
)
```

```
Iter: 0, D: 1.359, G:0.7201
Iter: 250, D: 1.372, G:1.395
Iter: 500, D: 1.374, G:0.9569
Iter: 750, D: 1.23, G:1.014
Iter: 1000, D: 1.246, G:0.8469
Iter: 1250, D: 1.236, G:1.022
```

```
Iter: 1500, D: 1.226, G:1.004
Iter: 1750, D: 1.346, G:0.9483
Iter: 2000, D: 1.335, G:0.8204
Iter: 2250, D: 1.38, G:0.7497
Iter: 2500, D: 1.221, G:0.8082
Iter: 2750, D: 1.303, G:0.9215
Iter: 3000, D: 1.261, G:0.9795
Iter: 3250, D: 1.319, G:0.7978
Iter: 3500, D: 1.362, G:0.7645
Iter: 3750, D: 1.339, G:0.7403
```

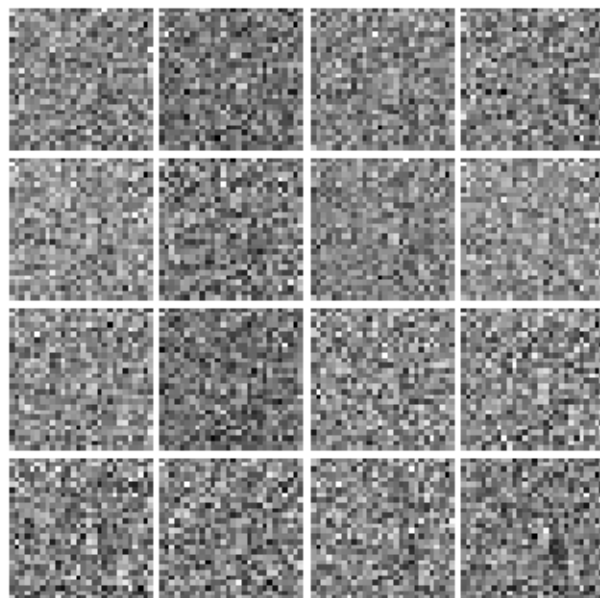
Run the cell below to show the generated images.

```
[25]: numIter = 0
      for img in images:
          print("Iter: {}".format(numIter))
          show_images(img)
          plt.show()
          numIter += 250
          print()
```

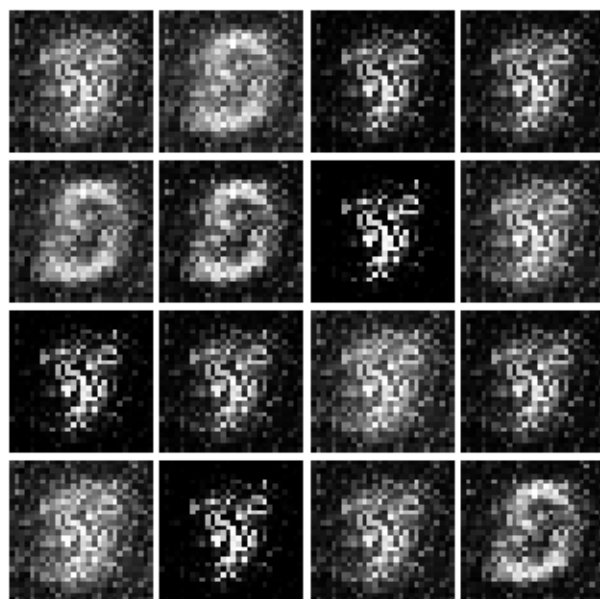
```
Iter: 0
```

5	0	4	1	9	2	1	3	1	4	3	5
3	6	1	7	2	8	6	9	4	0	9	1
1	2	4	3	2	7	3	8	6	9	0	5
6	0	7	6	1	8	7	9	3	9	8	5
9	3	3	0	7	4	9	8	0	9	4	1
4	4	6	0	4	5	6	1	0	0	1	7
1	6	3	0	2	1	1	7	9	0	2	6
7	8	3	9	0	4	6	7	4	6	8	0
7	8	3	1	5	7	1	7	1	1	6	3
0	2	9	3	1	1	0	4	9	2	0	0
2	0	2	7	1	8	6	4				

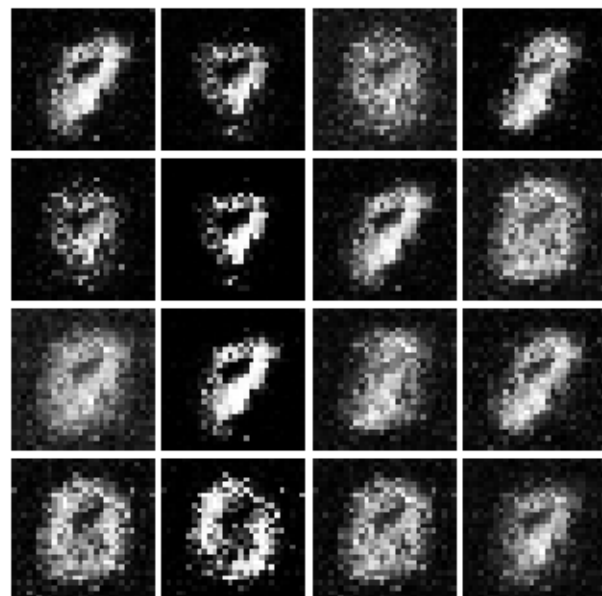
5	0	4	1	9	2	1	3	1	4	3	5
3	6	1	7	2	8	6	9	4	0	9	1
1	2	4	3	2	7	3	8	6	9	0	5
6	0	7	6	1	8	7	9	3	9	8	5
9	3	3	0	7	4	9	8	0	9	4	1
4	4	6	0	4	5	6	1	0	0	1	7
1	6	3	0	2	1	1	7	9	0	2	6
7	8	3	9	0	4	6	7	4	6	8	0
7	8	3	1	5	7	1	7	1	1	6	3
0	2	9	3	1	1	0	4	9	2	0	0
2	0	2	7	1	8	6	4				



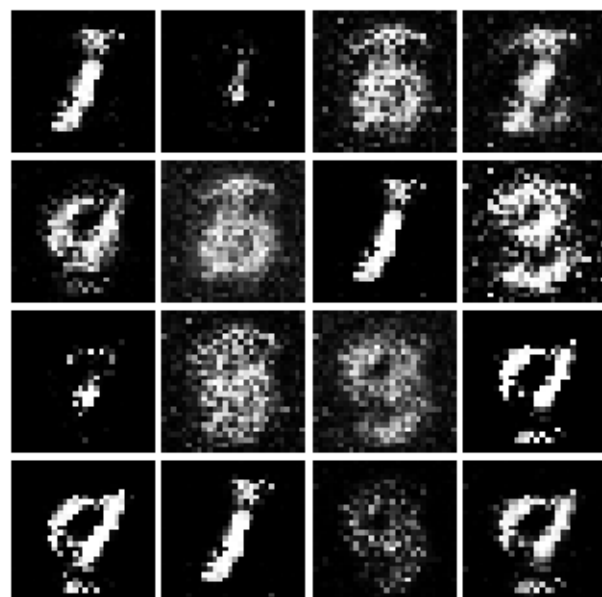
Iter: 250



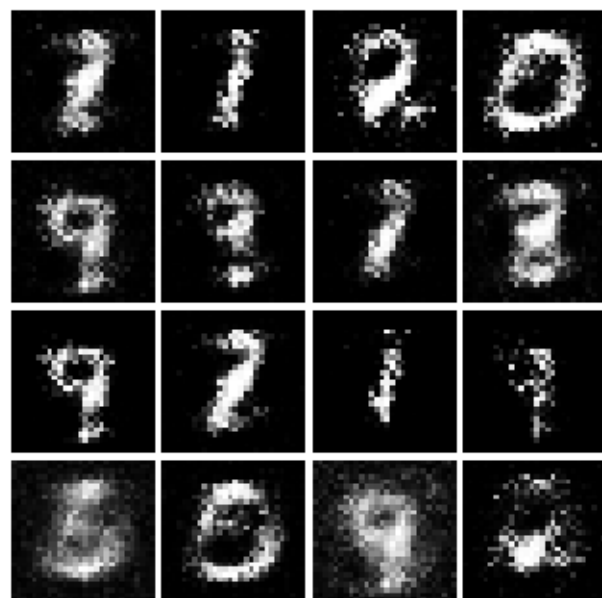
Iter: 500



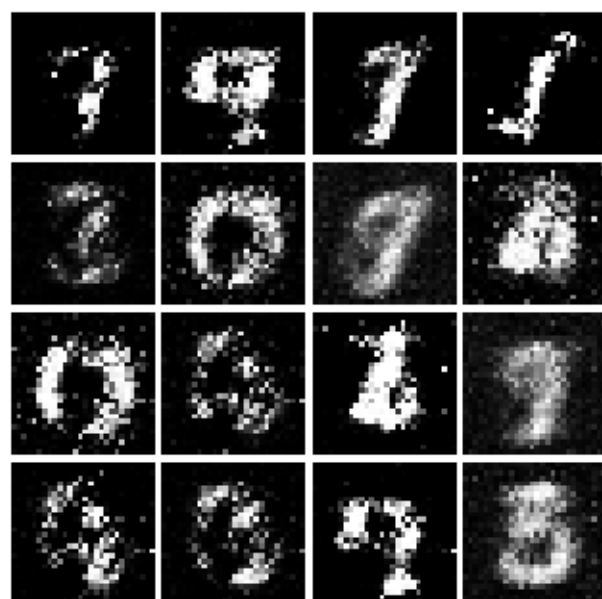
Iter: 750



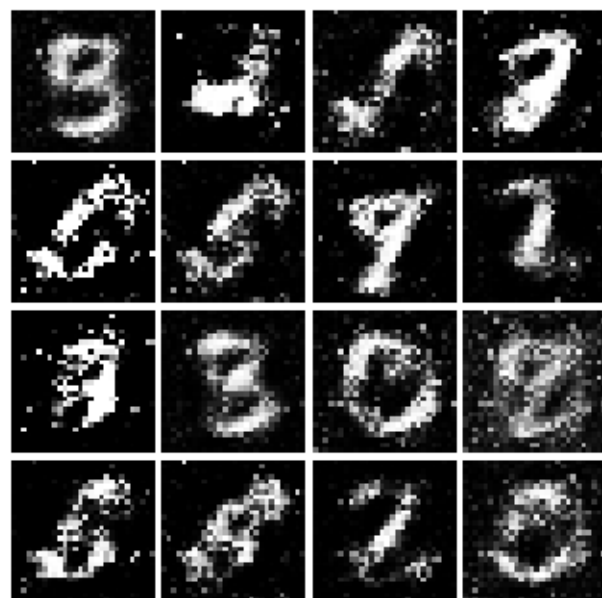
Iter: 1000



Iter: 1250



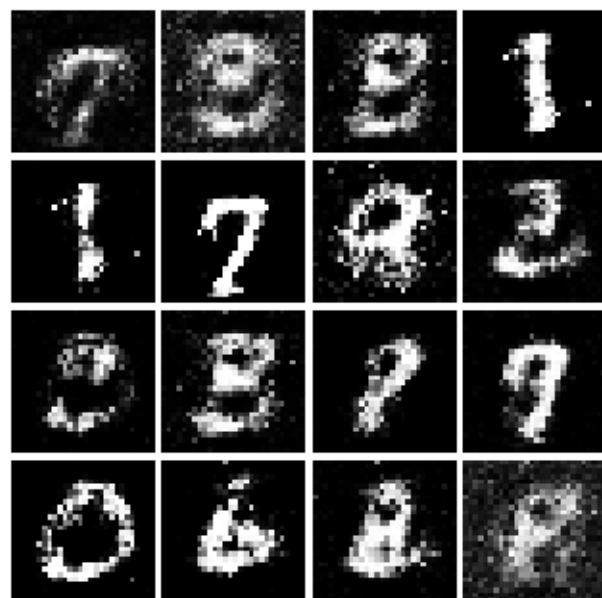
Iter: 1500



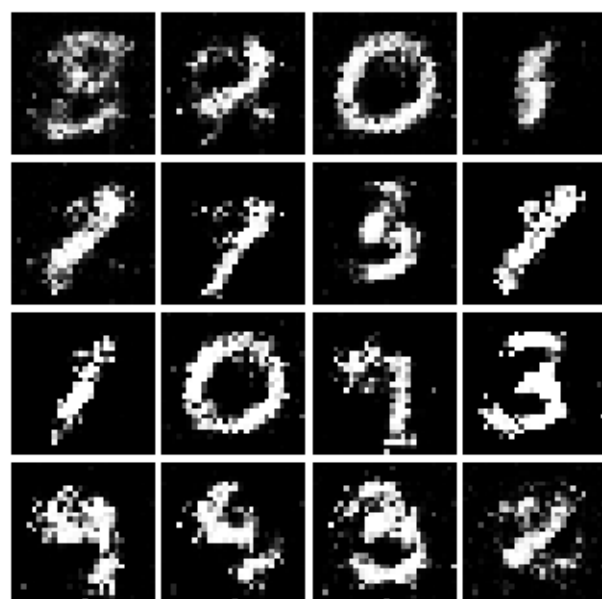
Iter: 1750



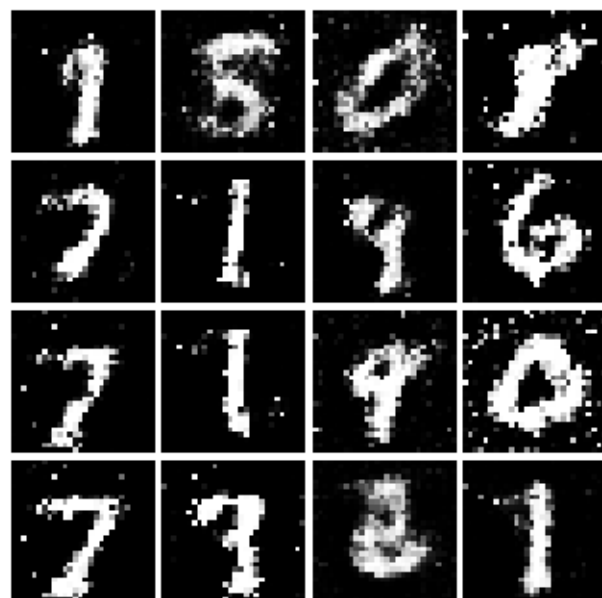
Iter: 2000



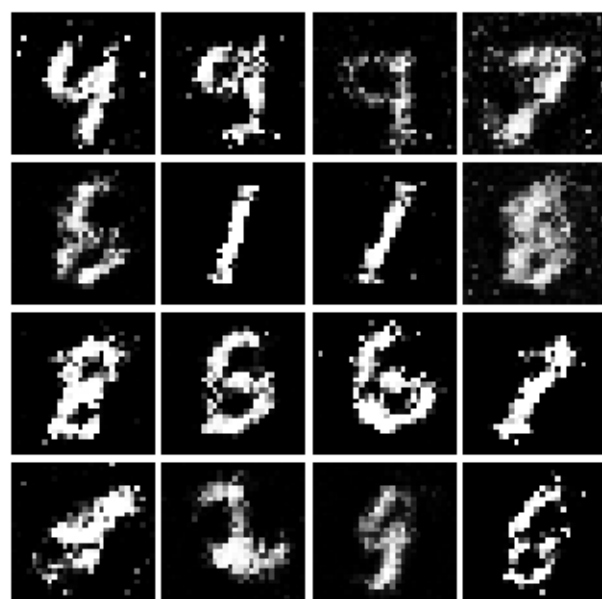
Iter: 2250



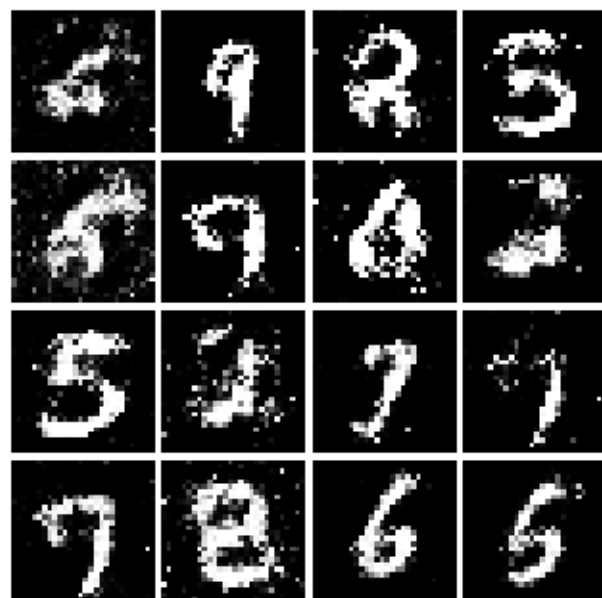
Iter: 2500



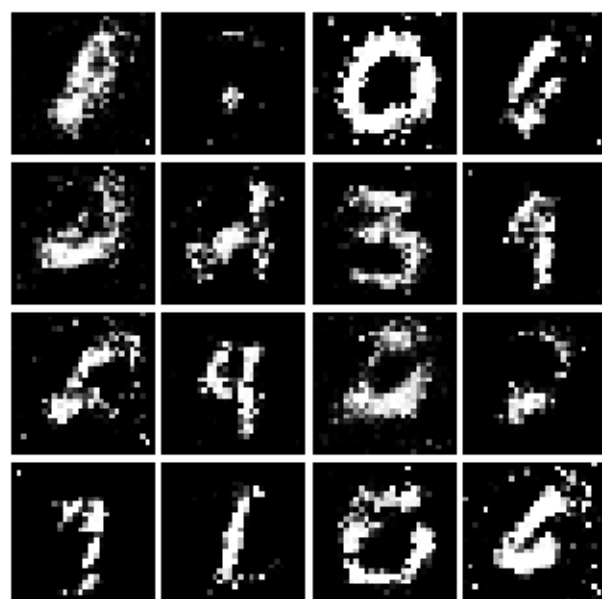
Iter: 2750



Iter: 3000



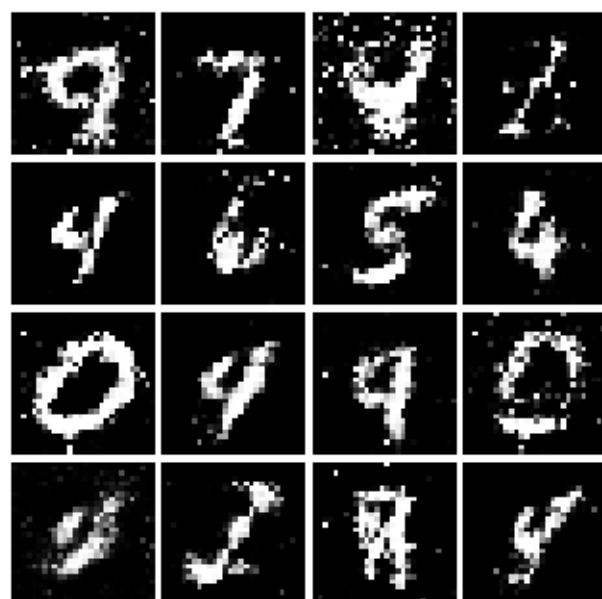
Iter: 3250



Iter: 3500



Iter: 3750

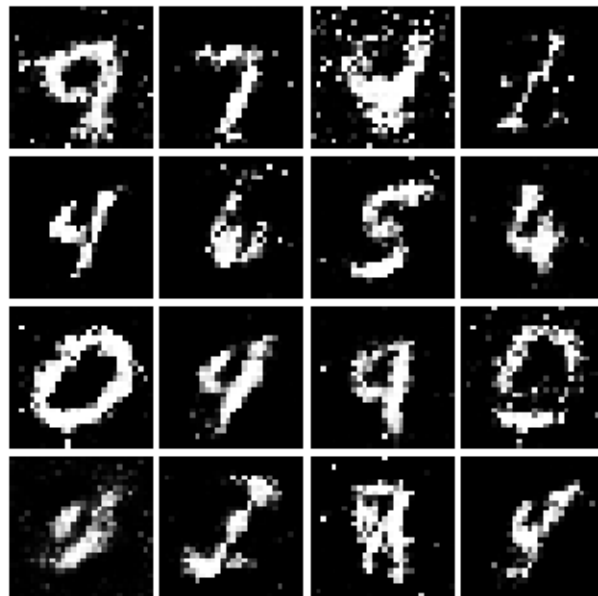


6.1 Inline Question 1

What does your final vanilla GAN image look like?

```
[26]: # This output is your answer.  
print("Vanilla GAN final image:")  
show_images(images[-1])  
plt.show()
```

Vanilla GAN final image:



Well that wasn't so hard, was it? In the iterations in the low 100s you should see black backgrounds, fuzzy shapes as you approach iteration 1000, and decent shapes, about half of which will be sharp and clearly recognizable as we pass 3000.

7 Least Squares GAN (2 points)

We'll now look at [Least Squares GAN](#), a newer, more stable alternative to the original GAN loss function. For this part, all we have to do is change the loss function and retrain the model. We'll implement equation (9) in the paper, with the generator loss:

$$\ell_G = \frac{1}{2} \mathbb{E}_{z \sim p(z)} [(D(G(z)) - 1)^2]$$

and the discriminator loss:

$$\ell_D = \frac{1}{2} \mathbb{E}_{x \sim p_{\text{data}}} [(D(x) - 1)^2] + \frac{1}{2} \mathbb{E}_{z \sim p(z)} [(D(G(z)))^2]$$

HINTS: Instead of computing the expectation, we will be averaging over elements of the minibatch, so make sure to combine the loss by averaging instead of summing. When plugging in for $D(x)$

and $D(G(z))$ use the direct output from the discriminator (`scores_real` and `scores_fake`). If you get errors of 0.333..., you probably forgot to include the $1/2$ term.

Implement `ls_discriminator_loss`, `ls_generator_loss` in `gan.py`

Before running a GAN with our new loss function, let's check it:

```
[27]: from gan import ls_discriminator_loss, ls_generator_loss

def test_lsgan_loss(score_real, score_fake, d_loss_true, g_loss_true):
    score_real = torch.Tensor(score_real).type(dtype)
    score_fake = torch.Tensor(score_fake).type(dtype)
    d_loss = ls_discriminator_loss(score_real, score_fake).cpu().numpy()
    g_loss = ls_generator_loss(score_fake).cpu().numpy()
    print("Maximum error in d_loss: %g"%rel_error(d_loss_true, d_loss))
    print("Maximum error in g_loss: %g"%rel_error(g_loss_true, g_loss))

test_lsgan_loss(
    answers['logits_real'],
    answers['logits_fake'],
    answers['d_loss_lsgan_true'],
    answers['g_loss_lsgan_true']
)
```

Maximum error in d_loss: 1.64377e-08

Maximum error in g_loss: 3.36961e-08

Run the following cell to train your model! If you train with the CPU, it takes about 7 minutes. If you train with the T4 GPU, it takes about 1 minute and 30 seconds.

```
[28]: D_LS = discriminator().type(dtype)
      G_LS = generator().type(dtype)

      D_LS_solver = get_optimizer(D_LS)
      G_LS_solver = get_optimizer(G_LS)

      images = run_a_gan(
          D_LS,
          G_LS,
          D_LS_solver,
          G_LS_solver,
          ls_discriminator_loss,
          ls_generator_loss,
          loader_train
      )
```

Iter: 0, D: 0.4695, G:0.5124

Iter: 250, D: 0.1858, G:0.173

Iter: 500, D: 0.1549, G:0.272

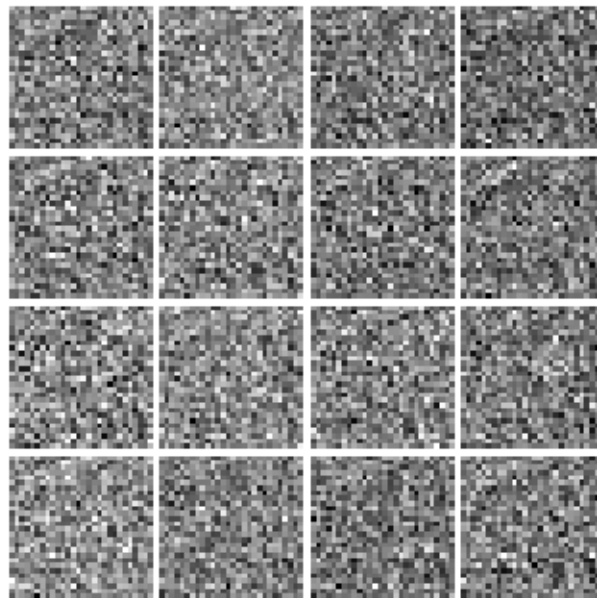
Iter: 750, D: 0.1254, G:0.291

```
Iter: 1000, D: 0.1641, G:0.3987
Iter: 1250, D: 0.1617, G:0.2608
Iter: 1500, D: 0.2038, G:0.2073
Iter: 1750, D: 0.2224, G:0.2079
Iter: 2000, D: 0.2307, G:0.1841
Iter: 2250, D: 0.2041, G:0.2092
Iter: 2500, D: 0.2216, G:0.1685
Iter: 2750, D: 0.208, G:0.1685
Iter: 3000, D: 0.24, G:0.1541
Iter: 3250, D: 0.2279, G:0.1541
Iter: 3500, D: 0.2299, G:0.1587
Iter: 3750, D: 0.2254, G:0.1516
```

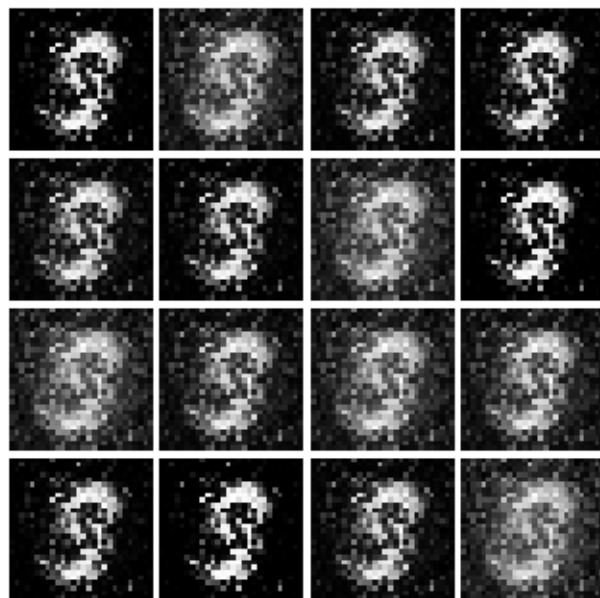
Run the cell below to show generated images.

```
[29]: numIter = 0
      for img in images:
          print("Iter: {}".format(numIter))
          show_images(img)
          plt.show()
          numIter += 250
          print()
```

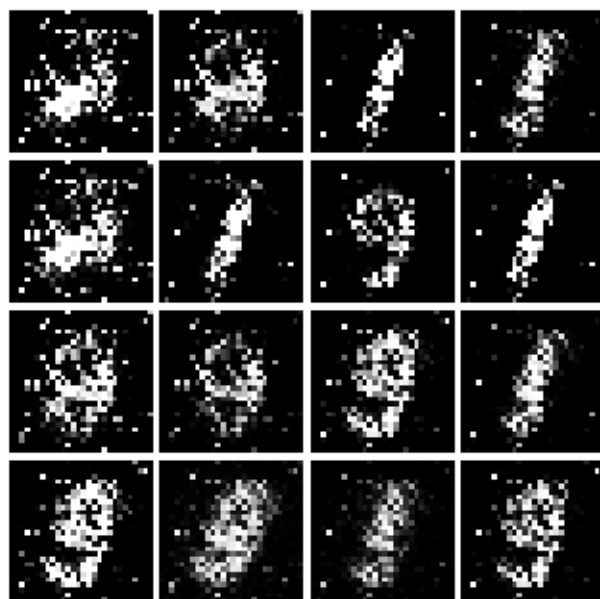
Iter: 0



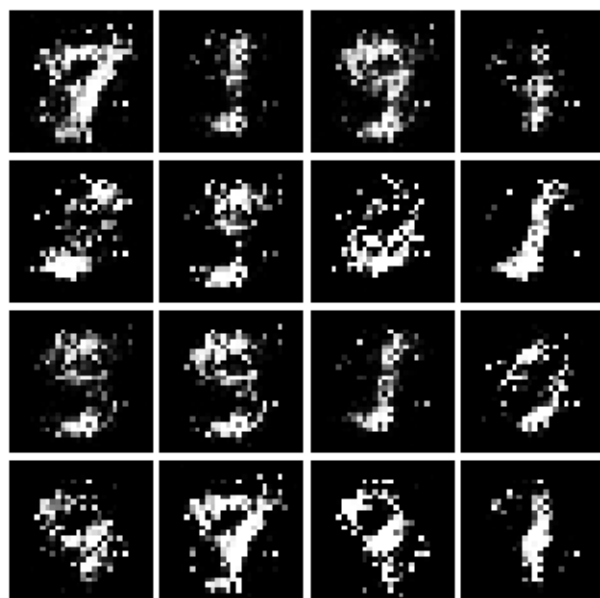
Iter: 250



Iter: 500



Iter: 750



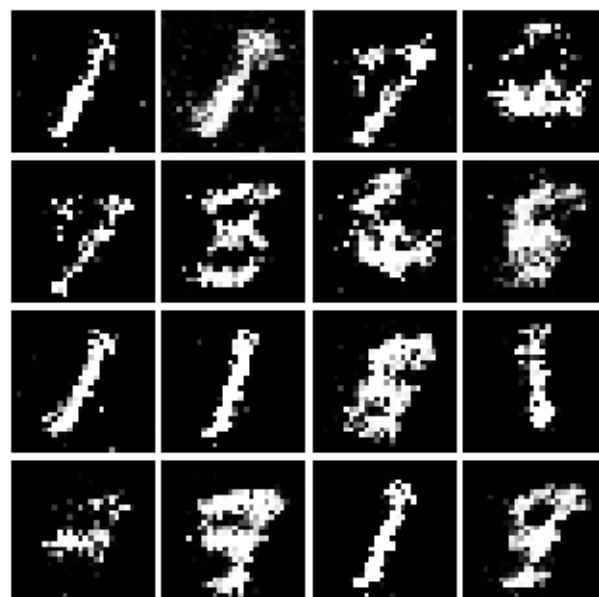
Iter: 1000



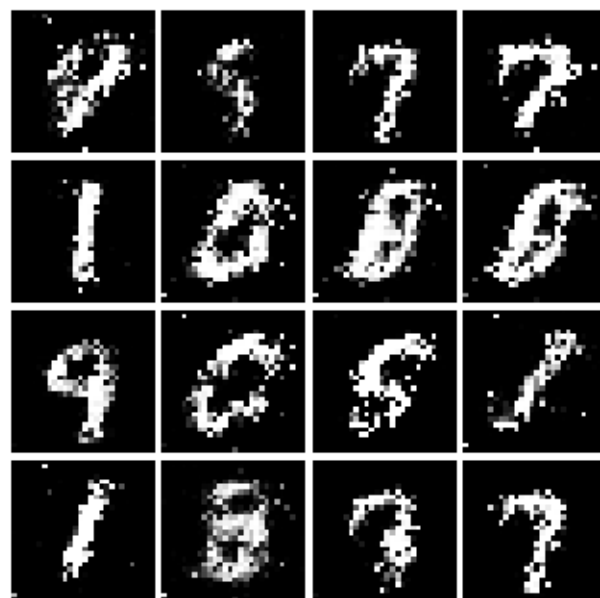
Iter: 1250



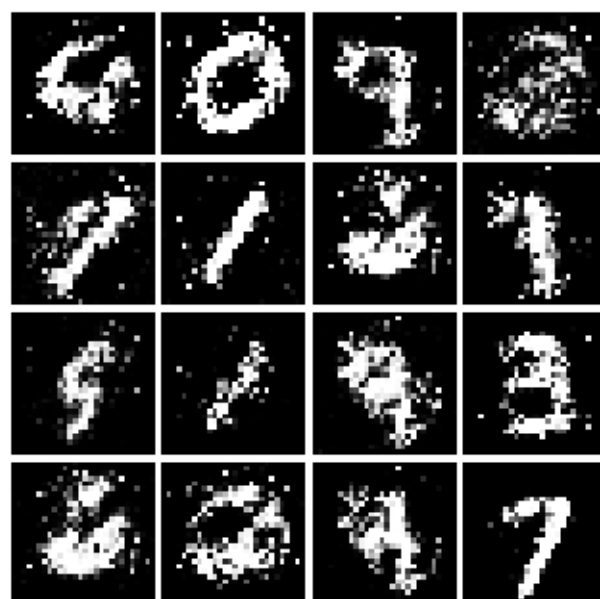
Iter: 1500



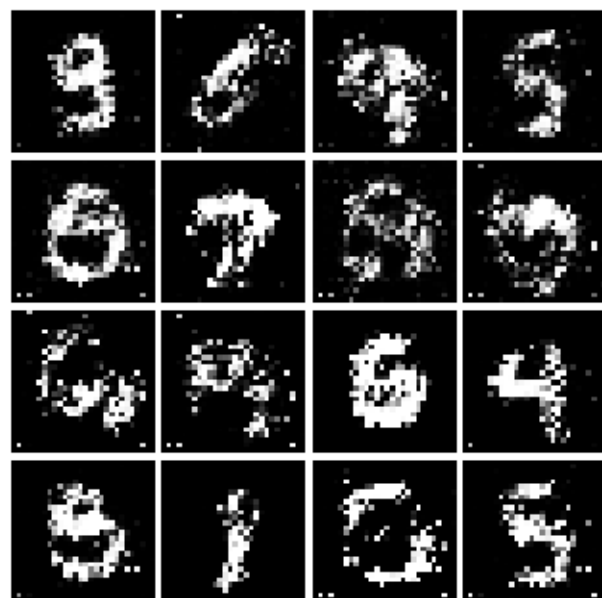
Iter: 1750



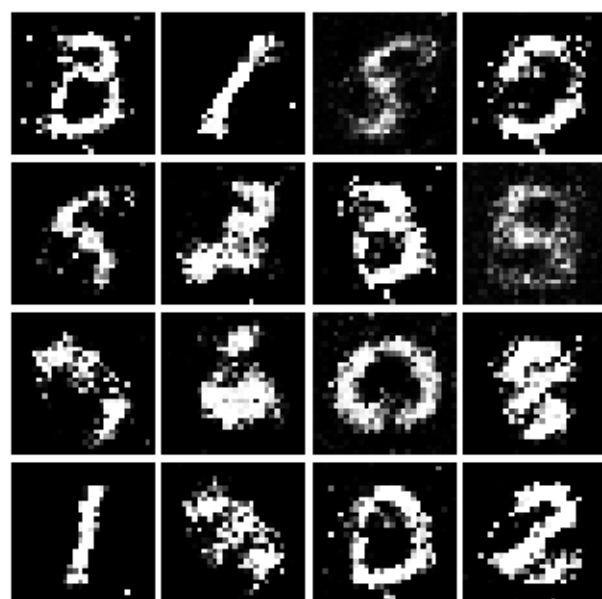
Iter: 2000



Iter: 2250



Iter: 2500



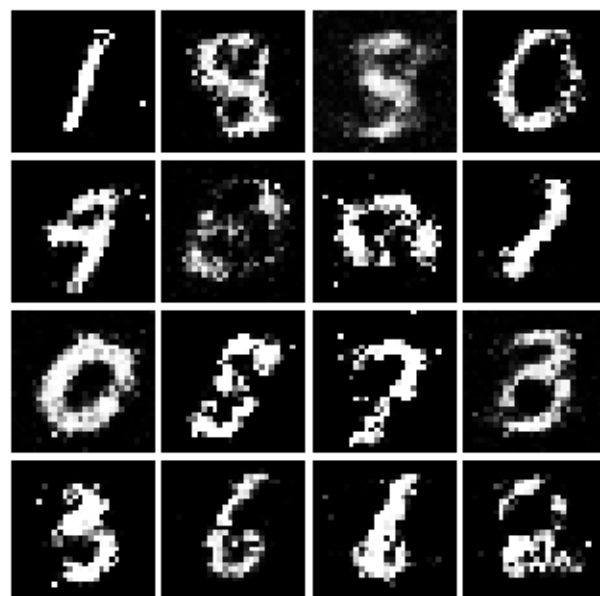
Iter: 2750



Iter: 3000



Iter: 3250



Iter: 3500



Iter: 3750



7.1 Inline Question 2

What does your final LSGAN image look like?

```
[30]: # This output is your answer.  
print("LSGAN final image:")  
show_images(images[-1])  
plt.show()
```

LSGAN final image:



8 Deep Convolutional GAN (2 points)

In the first part of the notebook, we implemented an almost direct copy of the original GAN network from Ian Goodfellow. However, this network architecture allows no real spatial reasoning. It is unable to reason about things like “sharp edges” in general because it lacks any convolutional layers. Thus, in this section, we will implement some of the ideas from [DCGAN](#), where we use convolutional networks

Discriminator We will use a discriminator inspired by the TensorFlow MNIST classification tutorial, which is able to get above 99% accuracy on the MNIST dataset fairly quickly. **Hint:** There is no need to specify padding. **Warning:** If you use LazyLinear I will take off points. * Conv2D: 32 Filters, 5x5, Stride 1 * Leaky ReLU(alpha=0.01) * Max Pool 2x2, Stride 2 * Conv2D: 64 Filters, 5x5, Stride 1 * Leaky ReLU(alpha=0.01) * Max Pool 2x2, Stride 2 * Flatten * Fully Connected with output size 4 x 4 x 64 * Leaky ReLU(alpha=0.01) * Fully Connected with output size 1

Implement `build_dc_classifier` in `gan.py`

```
[31]: from gan import build_dc_classifier

data = next(enumerate(loader_train))[-1][0].type(dtype)
b = build_dc_classifier().type(dtype)
out = b(data)
print(out.size())
```

```
torch.Size([128, 1])
```

Check the number of parameters in your classifier as a sanity check:


```
[32]: def test_dc_classifier(true_count=1102721):
    model = build_dc_classifier()
    cur_count = count_params(model)
    if cur_count != true_count:
        print('Incorrect number of parameters in classifier. Check your_
↪achitecture.')
    else:
        print('Correct number of parameters in classifier.')

test_dc_classifier()
```

Incorrect number of parameters in classifier. Check your achitecture.

Generator For the generator, we will copy the architecture exactly from the [InfoGAN paper](#). See Appendix C.1 MNIST. See the documentation for [nn.ConvTranspose2d](#). We are always “training” in GAN mode. * Fully connected with output size 1024 * ReLU * BatchNorm * Fully connected with output size 7 x 7 x 128 * ReLU * BatchNorm * Use `Unflatten()` to reshape into Image Tensor of shape 7, 7, 128 * `ConvTranspose2d`: 64 filters of 4x4, stride 2, ‘same’ padding (use `padding=1`) * ReLU * BatchNorm * `ConvTranspose2d`: 1 filter of 4x4, stride 2, ‘same’ padding (use `padding=1`) * TanH * Should have a 28x28x1 image, reshape back into 784 vector (using `Flatten()`)

Implement `build_dc_generator` in `gan.py`

```
[33]: from gan import build_dc_generator

test_g_gan = build_dc_generator().type(dtype)
test_g_gan.apply(initialize_weights)

fake_seed = torch.randn(batch_size, NOISE_DIM).type(dtype)
fake_images = test_g_gan.forward(fake_seed)
fake_images.size()
```

```
[33]: torch.Size([128, 1, 28, 28])
```

Check the number of parameters in your generator as a sanity check:

```
[34]: def test_dc_generator(true_count=6580801):
    model = build_dc_generator(4)
    cur_count = count_params(model)
    if cur_count != true_count:
        print('Incorrect number of parameters in generator. Check your_
↪achitecture.')
    else:
        print('Correct number of parameters in generator.')

test_dc_generator()
```

Incorrect number of parameters in generator. Check your achitecture.

Run the following cell to train your DCGAN. If you train with the CPU, it takes about 35 minutes. If you train with the T4 GPU, it takes about 1 minute.

```
[35]: D_DC = build_dc_classifier().type(dtype)
      D_DC.apply(initialize_weights)
      G_DC = build_dc_generator().type(dtype)
      G_DC.apply(initialize_weights)

      D_DC_solver = get_optimizer(D_DC)
      G_DC_solver = get_optimizer(G_DC)

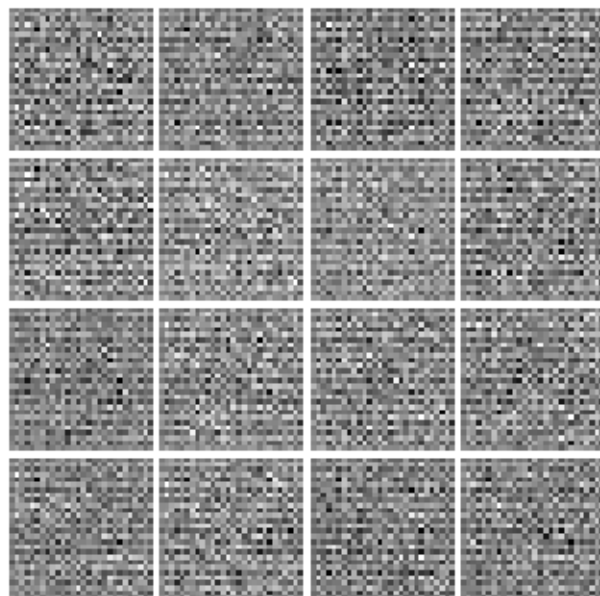
      images = run_a_gan(
          D_DC,
          G_DC,
          D_DC_solver,
          G_DC_solver,
          discriminator_loss,
          generator_loss,
          loader_train,
          num_epochs=5
      )
```

```
Iter: 0, D: 1.495, G:0.5913
Iter: 250, D: 0.2448, G:4.928
Iter: 500, D: 0.5336, G:3.134
Iter: 750, D: 0.3531, G:3.417
Iter: 1000, D: 0.8761, G:2.471
Iter: 1250, D: 0.6755, G:3.443
Iter: 1500, D: 0.7381, G:2.864
Iter: 1750, D: 0.2731, G:3.531
```

Run the cell below to show generated images.

```
[36]: numIter = 0
      for img in images:
          print("Iter: {}".format(numIter))
          show_images(img)
          plt.show()
          numIter += 250
          print()
```

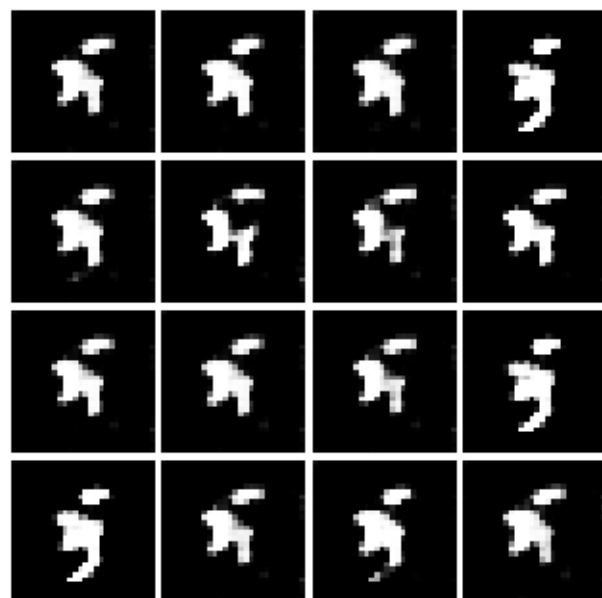
```
Iter: 0
```



Iter: 250



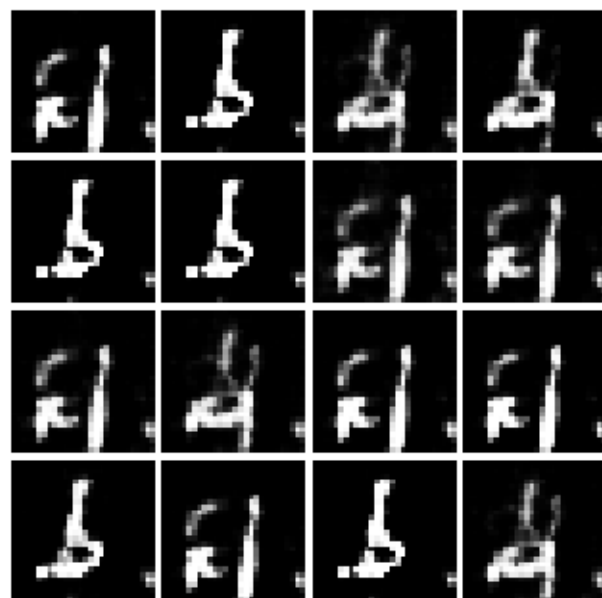
Iter: 500



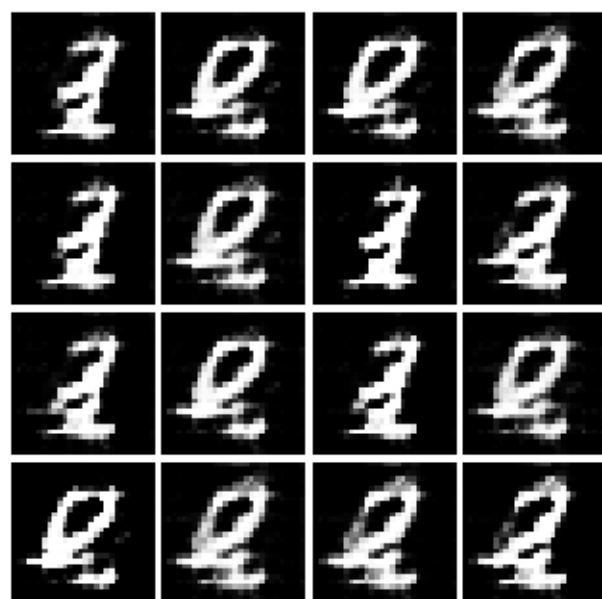
Iter: 750



Iter: 1000



Iter: 1250



Iter: 1500



Iter: 1750

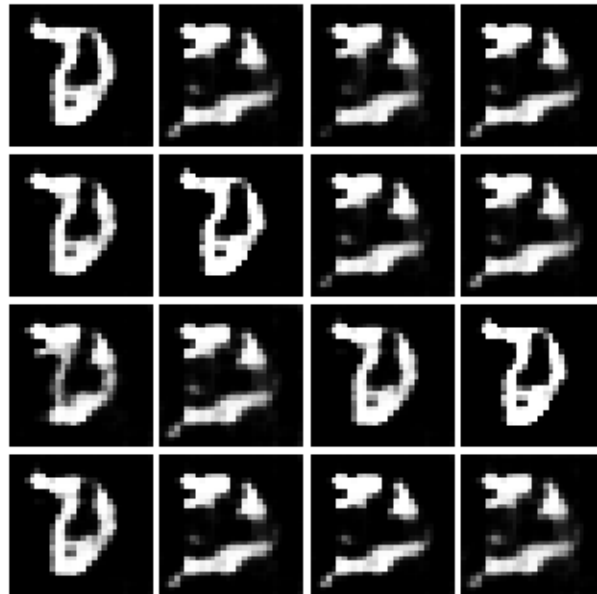


8.1 Inline Question 3

What does your final DCGAN image look like?

```
[37]: # This output is your answer.  
print("DCGAN final image:")  
show_images(images[-1])  
plt.show()
```

DCGAN final image:



8.2 Inline Question 4 (1 point)

We will look at an example to see why alternating minimization of the same objective (like in a GAN) can be tricky business.

Consider $f(x, y) = xy$. What does $\min_x \max_y f(x, y)$ evaluate to? (Hint: minmax tries to minimize the maximum value achievable.)

Now try to evaluate this function numerically for 6 steps, starting at the point $(1, 1)$, by using alternating gradient (first updating y , then updating x using that updated y) with step size 1. **Here step size is the learning_rate, and steps will be learning_rate * gradient.** You'll find that writing out the update step in terms of $x_t, y_t, x_{t+1}, y_{t+1}$ will be useful.

Record the six pairs of explicit values for (x_t, y_t) in the table below and briefly explain what $f(x, y)$ evaluates to.

8.2.1 Your answer:

y_0	y_1	y_2	y_3	y_4	y_5	y_6
1	2	2	0	-4	-8	-8
x_0	x_1	x_2	x_3	x_4	x_5	x_6
1	0	-2	-4	-4	0	8

$f(x, y)$ oscillates and diverge to infinity if iterations go to infinity.

8.3 Inline Question 5 (1 point)

Using this method, will we ever reach the optimal value? Why or why not?

8.3.1 Your answer:

No, because it is unstable that the directions of x/y change oppositely, the system will be cyclic and diverging.

8.4 Inline Question 6 (1 point)

If the generator loss decreases during training while the discriminator loss stays at a constant high value from the start, is this a good sign? Why or why not? A qualitative answer is sufficient.

8.4.1 Your answer:

no. if the discriminator loss stays high, then it may imply the discriminator will never distinguish the real and fake images. So it is a failed discriminator and cannot give valid feedback to the generator. Even under this situation, the generator loss decreases; we cannot tell whether the generator is learning better.


```

import numpy as np

import torch
import torch.nn as nn
import torchvision
import torchvision.transforms as T
import torch.optim as optim
from torch.utils.data import sampler

import PIL

NOISE_DIM = 96

dtype = torch.cuda.FloatTensor if torch.cuda.is_available() else torch.FloatTensor

def sample_noise(batch_size, dim, seed=None):
    """
    Generate a PyTorch Tensor of uniform random noise.

    Input:
    - batch_size: Integer giving the batch size of noise to generate.
    - dim: Integer giving the dimension of noise to generate.

    Output:
    - A PyTorch Tensor of shape (batch_size, dim) containing uniform
      random noise in the range (-1, 1).
    """
    if seed is not None:
        torch.manual_seed(seed)

    # *****START OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****

    return 2 * torch.rand(batch_size, dim) - 1

    # *****END OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****

def discriminator():
    """
    Build and return a PyTorch model implementing the architecture above.
    """
    # *****START OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****
    model = nn.Sequential(
        Flatten(),
        nn.Linear(28*28, 256),
        nn.LeakyReLU(0.01),
        nn.Linear(256, 256),
        nn.LeakyReLU(0.01),
        nn.Linear(256, 1)
    )
    return model
    # *****END OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****

def generator(noise_dim=NOISE_DIM):
    """
    Build and return a PyTorch model implementing the architecture above.
    """
    # *****START OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****
    model = nn.Sequential(
        nn.Linear(noise_dim, 1024),
        nn.ReLU(True),
        nn.Linear(1024, 1024),
        nn.ReLU(True),
        nn.Linear(1024, 784),
        nn.Tanh()
    )
    return model

```

```

# *****END OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****

def bce_loss(input, target):
    """
    Numerically stable version of the binary cross-entropy loss function in PyTorch.

    Inputs:
    - input: PyTorch Tensor of shape (N, 1) giving scores.
    - target: PyTorch Tensor of shape (N, 1) containing 0 and 1 giving targets.
      dtype is float! (a global dtype is defined above).

    Returns:
    - A PyTorch Tensor containing the mean BCE loss over the minibatch of input data.
    """
    bce = nn.BCEWithLogitsLoss()
    return bce(input, target)

def discriminator_loss(logits_real, logits_fake):
    """
    Computes the discriminator loss described above.

    Inputs:
    - logits_real: PyTorch Tensor of shape (N, 1) giving scores for the real data.
    - logits_fake: PyTorch Tensor of shape (N, 1) giving scores for the fake data.

    Returns:
    - loss: PyTorch Tensor containing (scalar) the loss for the discriminator.
    """
    loss = None
    # *****START OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****

    labels_real = torch.ones_like(logits_real)
    labels_fake = torch.zeros_like(logits_fake)
    loss_real = bce_loss(logits_real, labels_real)
    loss_fake = bce_loss(logits_fake, labels_fake)
    return loss_real + loss_fake

    # *****END OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****
    return loss

def generator_loss(logits_fake):
    """
    Computes the generator loss described above.

    Inputs:
    - logits_fake: PyTorch Tensor of shape (N, 1) giving scores for the fake data.

    Returns:
    - loss: PyTorch Tensor containing the (scalar) loss for the generator.
    """
    loss = None
    # *****START OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****
    return bce_loss(logits_fake, torch.ones_like(logits_fake))

    # *****END OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****
    return loss

def get_optimizer(model):
    """
    Construct and return an Adam optimizer for the model with learning rate 1e-3,
    beta1=0.5, and beta2=0.999.

    Input:
    - model: A PyTorch model that we want to optimize.

```

```
- An Adam optimizer for the model with the desired hyperparameters.
"""
```

```
# *****START OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****
```

```
# *****END OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****
```

```
return nn.Sequential(
    nn.Conv2d(1, 32, kernel_size=5, stride=1),
    nn.LeakyReLU(0.01),
    nn.MaxPool2d(2, 2),
```

```

        nn.Conv2d(32, 64, kernel_size=5, stride=1),
        nn.LeakyReLU(0.01),
        nn.MaxPool2d(2, 2),
        Flatten(),
        nn.Linear(1024, 1)
    )

# *****END OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****
#####
#
#                                     END OF YOUR CODE
#
#####

def build_dc_generator(noise_dim=NOISE_DIM):
    """
    Build and return a PyTorch model implementing the DCGAN generator using
    the architecture described above.
    """

    #####
    # TODO: Implement architecture
    #
    # HINT: nn.Sequential might be helpful.
    #####
    # *****START OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****

    model = nn.Sequential(
        nn.Linear(noise_dim, 1024),
        nn.ReLU(True),
        nn.Linear(1024, 7*7*128),
        nn.ReLU(True),
        Unflatten(N=-1, C=128, H=7, W=7),
        nn.ConvTranspose2d(128, 64, kernel_size=4, stride=2, padding=1),
        nn.ReLU(True),
        nn.ConvTranspose2d(64, 1, kernel_size=4, stride=2, padding=1),
        nn.Tanh()
    )

    return model

# *****END OF YOUR CODE (DO NOT DELETE/MODIFY THIS LINE)*****
#####
#
#                                     END OF YOUR CODE
#
#####

def run_a_gan(D, G, D_solver, G_solver, discriminator_loss, generator_loss, loader_train,
show_every=250,
            batch_size=128, noise_size=96, num_epochs=10):
    """
    Train a GAN!

    Inputs:
    - D, G: PyTorch models for the discriminator and generator
    - D_solver, G_solver: torch.optim Optimizers to use for training the
      discriminator and generator.
    - discriminator_loss, generator_loss: Functions to use for computing the generator and
      discriminator loss, respectively.
    - show_every: Show samples after every show_every iterations.
    - batch_size: Batch size to use for training.
    - noise_size: Dimension of the noise to use as input to the generator.
    - num_epochs: Number of epochs over the training dataset to use for training.
    """
    images = []
    iter_count = 0
    for epoch in range(num_epochs):
        for x, _ in loader_train:
            if len(x) != batch_size:

```

```

        continue
    D_solver.zero_grad()
    real_data = x.type(dtype)
    logits_real = D(2* (real_data - 0.5)).type(dtype)

    g_fake_seed = sample_noise(batch_size, noise_size).type(dtype)
    fake_images = G(g_fake_seed).detach()
    logits_fake = D(fake_images.view(batch_size, 1, 28, 28))

    d_total_error = discriminator_loss(logits_real, logits_fake)
    d_total_error.backward()
    D_solver.step()

    G_solver.zero_grad()
    g_fake_seed = sample_noise(batch_size, noise_size).type(dtype)
    fake_images = G(g_fake_seed)

    gen_logits_fake = D(fake_images.view(batch_size, 1, 28, 28))
    g_error = generator_loss(gen_logits_fake)
    g_error.backward()
    G_solver.step()

    if (iter_count % show_every == 0):
        print('Iter: {}, D: {:.4}, G:
{:.4}'.format(iter_count, d_total_error.item(), g_error.item()))
        imgs_numpy = fake_images.data.cpu().numpy()
        images.append(imgs_numpy[0:16])

    iter_count += 1

return images

```

```

class ChunkSampler(sampler.Sampler):

```

"""Samples elements sequentially from some offset.

Arguments:

num_samples: # of desired datapoints

start: offset where we should start selecting from

"""

```

def __init__(self, num_samples, start=0):

```

```

    self.num_samples = num_samples

```

```

    self.start = start

```

```

def __iter__(self):

```

```

    return iter(range(self.start, self.start + self.num_samples))

```

```

def __len__(self):

```

```

    return self.num_samples

```

```

class Flatten(nn.Module):

```

```

    def forward(self, x):

```

```

        N, C, H, W = x.size() # read in N, C, H, W

```

```

        return x.view(N, -1) # "flatten" the C * H * W values into a single vector per image

```

```

class Unflatten(nn.Module):

```

"""

*An Unflatten module receives an input of shape (N, C*H*W) and reshapes it to produce an output of shape (N, C, H, W).*

"""

```

def __init__(self, N=-1, C=128, H=7, W=7):

```

```

    super(Unflatten, self).__init__()

```

```

    self.N = N

```

```

    self.C = C

```

```

    self.H = H

```

```

    self.W = W

```

```

def forward(self, x):
    return x.view(self.N, self.C, self.H, self.W)

def initialize_weights(m):
    if isinstance(m, nn.Linear) or isinstance(m, nn.ConvTranspose2d):
        nn.init.xavier_uniform_(m.weight.data)

def preprocess_img(x):
    return 2 * x - 1.0

def deprocess_img(x):
    return (x + 1.0) / 2.0

def rel_error(x, y):
    return np.max(np.abs(x - y) / (np.maximum(1e-8, np.abs(x) + np.abs(y))))

def count_params(model):
    """Count the number of parameters in the model. """
    param_count = np.sum([np.prod(p.size()) for p in model.parameters()])
    return param_count

```

vae_ipynb

April 21, 2025

We would like to acknowledge University of Michigan's EECS 498-007/598-005 on which we based the development of this project.

1 Variational Autoencoder

In this notebook, you will implement a variational autoencoder and a conditional variational autoencoder with slightly different architectures and apply them to the popular MNIST handwritten dataset. Recall from C147/C247, an autoencoder seeks to learn a latent representation of our training images by using unlabeled data and learning to reconstruct its inputs. The *variational autoencoder* extends this model by adding a probabilistic spin to the encoder and decoder, allowing us to sample from the learned distribution of the latent space to generate new images at inference time.

1.1 Setup Code

Before getting started, we need to run some boilerplate code to set up our environment. You'll need to rerun this setup code each time you start the notebook.

First, run this cell that loads the autoreload extension. This allows us to edit .py source files and re-import them into the notebook for a seamless editing and debugging experience.

```
[12]: %load_ext autoreload
      %autoreload 2

      USE_COLAB = False
```

1.1.1 Google Colab Setup

Next we need to run a few commands to set up our environment on Google Colab. If you are running this notebook on a local machine you can skip this section.

Run the following cell to mount your Google Drive. Follow the link and sign in to your Google account (the same account you used to store this notebook!).

```
[13]: if USE_COLAB:
      from google.colab import drive
      drive.mount('/content/drive')
```

Now recall the path in your Google Drive where you uploaded this notebook and fill it in below. If everything is working correctly then running the following cell should print the filenames from the assignment:

```
['vae.ipynb', 'nnd12', 'vae.py']
```

```
[14]: import os

if USE_COLAB:
    # TODO: Fill in the Google Drive path where you uploaded the assignment
    # Example: '239AS.2/project1/vae'
    GOOGLE_DRIVE_PATH_AFTER_MYDRIVE = '239AS.2/project1/vae'
    GOOGLE_DRIVE_PATH = os.path.join('drive', 'My Drive',
    ↪GOOGLE_DRIVE_PATH_AFTER_MYDRIVE)
    print(os.listdir(GOOGLE_DRIVE_PATH))
```

Once you have successfully mounted your Google Drive and located the path to this assignment, run the following cell to allow us to import from the .py files of this assignment. If it works correctly, it should print the message:

```
Hello from vae.py!
Hello from helper.py!
```

```
[15]: import sys
if USE_COLAB:
    sys.path.append(GOOGLE_DRIVE_PATH)

from vae import hello_vae
hello_vae()

from nnd12.helper import hello_helper
hello_helper()
```

```
Hello from vae.py!
Hello from helper.py!
```

Load several useful packages that are used in this notebook:

```
[16]: from nnd12.grad import rel_error
from nnd12.utils import reset_seed
import math
import torch
import torch.nn as nn
import torch.nn.functional as F
from torch.nn import init
import torchvision
import torchvision.transforms as T
import torch.optim as optim
from torch.utils.data import DataLoader
from torch.utils.data import sampler
```



```

import torchvision.datasets as dset

import matplotlib.pyplot as plt
%matplotlib inline

# for plotting
plt.rcParams['figure.figsize'] = (10.0, 8.0) # set default size of plots
plt.rcParams['font.size'] = 16
plt.rcParams['image.interpolation'] = 'nearest'
plt.rcParams['image.cmap'] = 'gray'

```

We will use GPUs to accelerate our computation in this notebook. Run the following to make sure GPUs are enabled:

```

[17]: if torch.cuda.is_available():
        print('Good to go!')
    else:
        print('Please set GPU via the downward triangle in the top right corner.')

```

Good to go!

1.2 Load MNIST Dataset

VAEs are notoriously finicky with hyperparameters, and also require many training epochs. In order to make this assignment approachable, we will be working on the MNIST dataset, which is 60,000 training and 10,000 test images. Each picture contains a centered image of white digit on black background (0 through 9). This was one of the first datasets used to train convolutional neural networks and it is fairly easy – a standard CNN model can easily exceed 99% accuracy.

To simplify our code here, we will use the PyTorch MNIST wrapper, which downloads and loads the MNIST dataset. See the [documentation](#) for more information about the interface. The default parameters will take 5,000 of the training examples and place them into a validation dataset. The data will be saved into a folder called MNIST.

```

[18]: if USE_COLAB:
        %cd /content/drive/My\ Drive/$GOOGLE_DRIVE_PATH_AFTER_MYDRIVE

batch_size = 128

mnist_train = dset.MNIST('./nndl2', train=True, download=True,
                        transform=T.ToTensor())
loader_train = DataLoader(mnist_train, batch_size=batch_size,
                        shuffle=True, drop_last=True, num_workers=2)

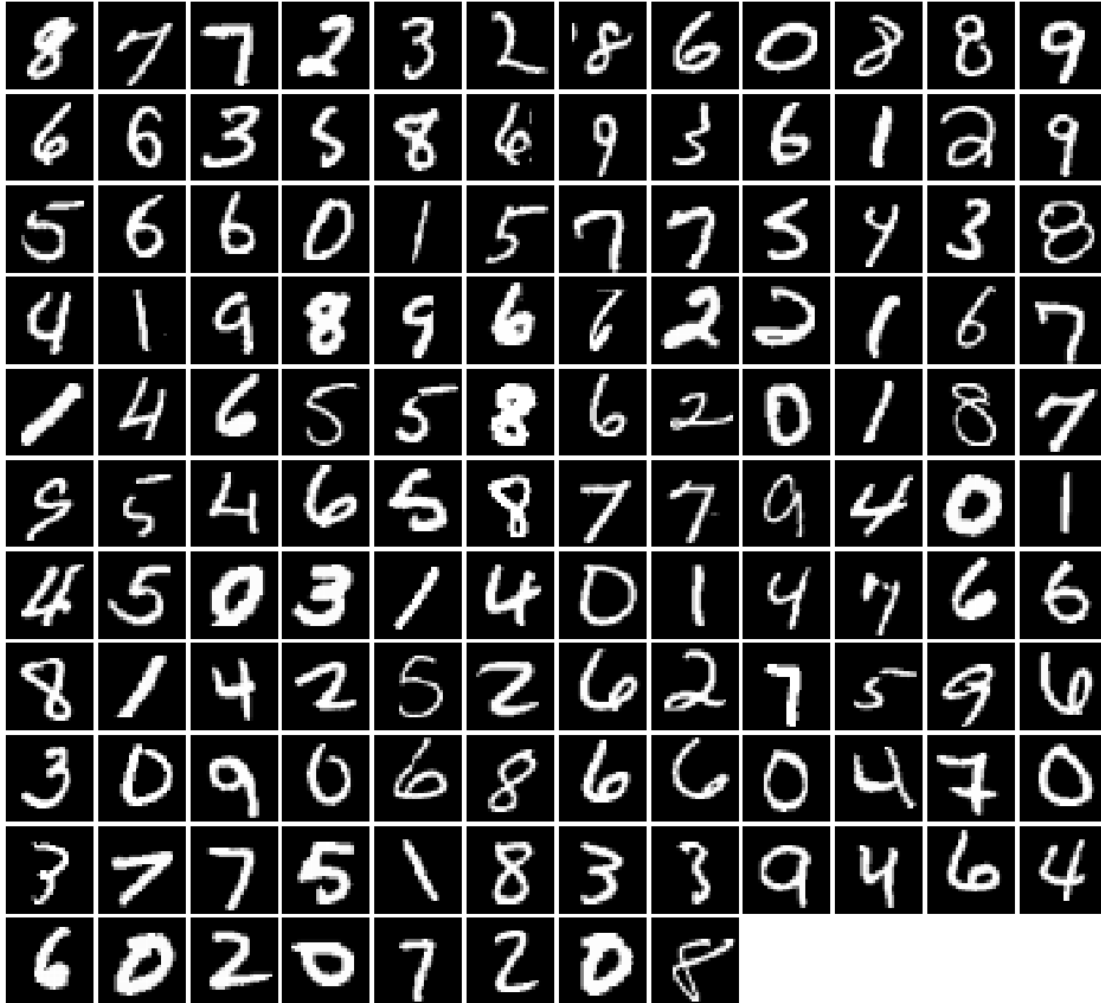
```

1.3 Visualize dataset

It is always a good idea to look at examples from the dataset before working with it. Let's visualize the digits in the MNIST dataset. We have defined the function `show_images` in `helper.py` that we call to visualize the images.

```
[19]: from nndl2.helper import show_images

imgs = next(iter(loader_train))[0].view(batch_size, 784)
show_images(imgs)
```



2 Fully Connected VAE

Our first VAE implementation will consist solely of fully connected layers. We'll take the $1 \times 28 \times 28$ shape of our input and flatten the features to create an input dimension size of 784. In this section you'll define the Encoder and Decoder models in the VAE class of `vae.py` and implement the reparametrization trick, forward pass, and loss function to train your first VAE.

2.1 FC-VAE Encoder (4 points)

Now let's start building our fully-connected VAE network. We'll start with the encoder, which will take our images as input (after flattening C,H,W to D shape) and pass them through a three

Linear+ReLU layers. We'll use this hidden dimension representation to predict both the posterior mu and posterior log-variance using two separate linear layers (both shape (N,Z)).

Note that we are calling this the 'logvar' layer because we'll use the log-variance (instead of variance or standard deviation) to stabilize training. This will specifically matter more when you compute reparametrization and the loss function later.

Define `hidden_dim=400`, `encoder`, `mu_layer`, and `logvar_layer` in the initialization of the VAE class in `vae.py`. Use `nn.Sequential` to define the encoder, and separate `Linear` layers for the `mu` and `logvar` layers. Architecture for the encoder is described below:

- Flatten (Hint: `nn.Flatten`)
- Fully connected layer with input size `input_size` and output size `hidden_dim`
- ReLU
- Fully connected layer with input_size `hidden_dim` and output size `hidden_dim`
- ReLU
- Fully connected layer with input_size `hidden_dim` and output size `hidden_dim`
- ReLU

```
[20]: from vae import VAE
def count_params(model):
    return sum(p.numel() for p in model.parameters())
def test_encoder(model, input_size, hidden_dim, n_encoder_lin_layers):
    """
    model: Model defined as above
    input_size: dimensionality of input
    hidden_dim: dimensionality of hidden state
    n_layers: number of Linear layers
    """
    expected_n_params = (input_size+1)*hidden_dim + \
        (n_encoder_lin_layers-1)*(hidden_dim+1)*hidden_dim
    actual_n_params = count_params(model.encoder)
    if actual_n_params == expected_n_params:
        print('Correct number of parameters in model.encoder.')
        return True
    else:
        print('Incorrect number of parameters in model.encoder.' \
            ' model.encoder does not include mu_layer and the logvar_layer.' \
            ' Check your achitecture.')
        return False
    return
def test_mu_logvar(model, hidden_dim, latent_size):
    """
    model: Model defined as above
    input_size: dimensionality of input
    hidden_dim: dimensionality of hidden state
    n_layers: number of Linear layers
    """
    if count_params(model.mu_layer) == (hidden_dim+1)*latent_size:
```

```

        print('Correct number of parameters in model.mu_layer.')
    else:
        print('Incorrect number of parameters in model.mu_layer.')
    if count_params(model.logvar_layer) == (hidden_dim+1)*latent_size:
        print('Correct number of parameters in model.logvar_layer.')
    else:
        print('Incorrect number of parameters in model.logvar_layer.')
    return
test_encoder(VAE(345, 17), 345, 400, 3)
test_mu_logvar(VAE(345, 17), 400, 17)

```

Correct number of parameters in model.encoder.
 Correct number of parameters in model.mu_layer.
 Correct number of parameters in model.logvar_layer.

2.2 FC-VAE Decoder (1 point)

We'll now define the decoder, which will take the latent space representation and generate a reconstructed image. The architecture is as follows:

- Fully connected layer with input size `latent_size` and output size `hidden_dim`
- ReLU
- Fully connected layer with input_size `hidden_dim` and output size `hidden_dim`
- ReLU
- Fully connected layer with input_size `hidden_dim` and output size `hidden_dim`
- ReLU
- Fully connected layer with input_size `hidden_dim` and output size `input_size`
- Sigmoid
- Unflatten (`nn.Unflatten`)

Define a *decoder* in the initialization of the VAE class in *vae.py*. Like the encoding step, use *nn.Sequential*

```

[21]: from vae import VAE
def count_params(model):
    return sum(p.numel() for p in model.parameters())
def test_decoder(model, input_size, hidden_dim, latent_size,
    ↪n_decoder_lin_layers):
    """
    model: Model defined as above
    input_size: dimensionality of input
    hidden_dim: dimensionality of hidden state
    latent_size: dimensionality of latent space
    n_layers: number of Linear layers in model.decoder
    """
    expected_n_params = (latent_size+1)*hidden_dim + \
        (n_decoder_lin_layers-2)*(hidden_dim+1)*hidden_dim + \
        (hidden_dim+1)*input_size
    actual_n_params = count_params(model.decoder)

```

```

    if actual_n_params == expected_n_params:
        print('Correct number of parameters in model.decoder.')
    else:
        print('Incorrect number of parameters in model.decoder.')
    return
test_decoder(VAE(345, 17), 345, 400, 17, 4)

```

Correct number of parameters in model.decoder.

2.3 Reparametrization (2 points)

Now we'll apply a reparametrization trick in order to estimate the posterior z during our forward pass, given the μ and σ^2 estimated by the encoder. A simple way to do this could be to simply generate a normal distribution centered at our μ and having a std corresponding to our σ^2 . However, we would have to backpropagate through this random sampling that is not differentiable. Instead, we sample initial random data ϵ from a fixed distribution, and compute z as a function of $(\epsilon, \sigma^2, \mu)$. Specifically:

$$z = \mu + \sigma\epsilon$$

We can easily find the partial derivatives w.r.t μ and σ^2 and backpropagate through z . If $\epsilon = \mathcal{N}(0, 1)$, then it's easy to verify that the result of our forward pass calculation will be a distribution centered at μ with variance σ^2 .

Implement reparametrization in `vae.py` and verify your mean and std error are at or less than $1e-4$.

```

[22]: reset_seed(0)
      from vae import reparametrize
      latent_size = 15
      size = (1, latent_size)
      mu = torch.zeros(size)
      logvar = torch.ones(size)

      z = reparametrize(mu, logvar)

      expected_mean = torch.FloatTensor([-0.4363])
      expected_std = torch.FloatTensor([1.6860])
      z_mean = torch.mean(z, dim=-1)
      z_std = torch.std(z, dim=-1)
      assert z.size() == size

      print('Mean Error', rel_error(z_mean, expected_mean))
      print('Std Error', rel_error(z_std, expected_std))

```

Mean Error 5.639056398351415e-05

Std Error 7.1412955526273885e-06

2.4 FC-VAE Forward (1 point)

Complete the VAE class by writing the forward pass. The forward pass should pass the input image through the encoder to calculate the estimation of mu and logvar, reparametrize to estimate the latent space z, and finally pass z into the decoder to generate an image.

```
[23]: from vae import VAE
def test_VAE_shapes():
    all_shapes_correct = True
    with torch.no_grad():
        batch_size = 3
        latent_size = 17
        x_hat, mu, logvar = VAE(28*28, latent_size)(torch.ones(batch_size, 1, 28, 28))
        if x_hat.shape != (batch_size, 1, 28, 28):
            print(f'x_hat has incorrect shape. Expected (batch_size, 1, 28, 28)')
            f' Got {tuple(x_hat.shape)}.'
            all_shapes_correct = False
        if mu.shape != (batch_size, latent_size):
            print(f'mu has incorrect shape. Expected (batch_size, latent_size)')
            f' Got {tuple(mu.shape)}.'
            all_shapes_correct = False
        if logvar.shape != (batch_size, latent_size):
            print(f'logvar has incorrect shape. Expected (batch_size, latent_size)')
            f' Got {tuple(logvar.shape)}.'
            all_shapes_correct = False
        if all_shapes_correct:
            print('Shapes of x_hat, mu, and logvar are correct.')
        if batch_size > 1 and x_hat.std(0).mean() == 0:
            print('x_hat has no randomness.')
    return
test_VAE_shapes()
```

```
torch.Size([3, 17])
torch.Size([3, 1, 28, 28])
Shapes of x_hat, mu, and logvar are correct.
```

2.5 Loss Function (1 point)

Before we're able to train our final model, we'll need to define our loss function. As seen below, the loss function for VAEs contains two terms: A reconstruction loss term (left) and KL divergence term (right).

$$-E_{Z \sim q_\phi(z|x)}[\log p_\theta(x|z)] + D_{KL}(q_\phi(z|x), p(z))$$

Note that this is the negative of the variational lowerbound shown in lecture—this ensures that when we are minimizing this loss term, we're maximizing the variational lowerbound. The reconstruction

loss term can be computed by simply using the binary cross entropy loss between the original input pixels and the output pixels of our decoder (Hint: `nn.functional.binary_cross_entropy`). The KL divergence term works to force the latent space distribution to be close to a prior distribution (we're using a standard normal gaussian as our prior).

To help you out, we've derived an unvectorized form of the KL divergence term for you. Suppose that $q_\phi(z|x)$ is a Z -dimensional diagonal Gaussian with mean $\mu_{z|x}$ of shape $(Z,)$ and standard deviation $\sigma_{z|x}$ of shape $(Z,)$, and that $p(z)$ is a Z -dimensional Gaussian with zero mean and unit variance. Then we can write the KL divergence term as:

$$D_{KL}(q_\phi(z|x), p(z)) = -\frac{1}{2} \sum_{j=1}^J (1 + \log(\sigma_{z|x}^2)_j - (\mu_{z|x})_j^2 - (\sigma_{z|x})_j^2)$$

It's up to you to implement a vectorized version of this loss that also operates on minibatches. You should average the loss across samples in the minibatch.

Implement `loss_function` in `vae.py` and verify your implementation below. Your relative error should be less than or equal to `1e-5`

```
[24]: from vae import loss_function
size = (1,15)

image_hat = torch.sigmoid(torch.FloatTensor([[2,5], [6,7]]).unsqueeze(0).
    ↪unsqueeze(0))
image = torch.sigmoid(torch.FloatTensor([[1,10], [9,3]]).unsqueeze(0).
    ↪unsqueeze(0))

expected_out = torch.tensor(8.5079)
mu, logvar = torch.ones(size), torch.zeros(size)
out = loss_function(image_hat, image, mu, logvar)
print('Loss error', rel_error(expected_out,out))
```

Loss error 2.1297676389877955e-06

2.6 Train a model

Now that we have our VAE defined and loss function ready, lets train our model! Our training script is provided in `nnd12/helper.py`, and we have pre-defined an Adam optimizer, learning rate, and # of epochs for you to use.

Training for 10 epochs should take ~2 minutes and your loss should be less than 120.

```
[25]: num_epochs = 10
latent_size = 15
from vae import VAE
from nnd12.helper import train_vae
input_size = 28*28
device = 'cuda' if torch.cuda.is_available() else 'cpu'
if device == 'cpu':
    print(f'Warning: using device {device} may take longer.')
vae_model = VAE(input_size, latent_size=latent_size)
vae_model.to(device)
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
Train Epoch: 0 Loss: 150.360550
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
Train Epoch: 2 Loss: 123.949516
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])

```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```

torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
Train Epoch: 4 Loss: 121.911789
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
Train Epoch: 8 Loss: 116.603508
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
torch.Size([128, 15])
torch.Size([128, 1, 28, 28])
Train Epoch: 9  Loss: 114.974007

```

2.7 Visualize results

After training our VAE network, we're able to take advantage of its power to generate new training examples. This process simply involves the decoder: we initialize some random distribution for our latent spaces z , and generate new examples by passing these latent space into the decoder.

Run the cell below to generate new images! You should be able to visually recognize many of the digits, although some may be a bit blurry or badly formed. Our next model will see improvement in these results.

```

[26]: device = next(vae_model.parameters()).device
z = torch.randn(10, latent_size).to(device=device)
import matplotlib.gridspec as gridspec
vae_model.eval()
samples = vae_model.decoder(z).data.cpu().numpy()

fig = plt.figure(figsize=(10, 1))
gspec = gridspec.GridSpec(1, 10)
gspec.update(wspace=0.05, hspace=0.05)
for i, sample in enumerate(samples):
    ax = plt.subplot(gspec[i])
    plt.axis('off')
    ax.set_xticklabels([])
    ax.set_yticklabels([])
    ax.set_aspect('equal')
    plt.imshow(sample.reshape(28,28), cmap='Greys_r')

```

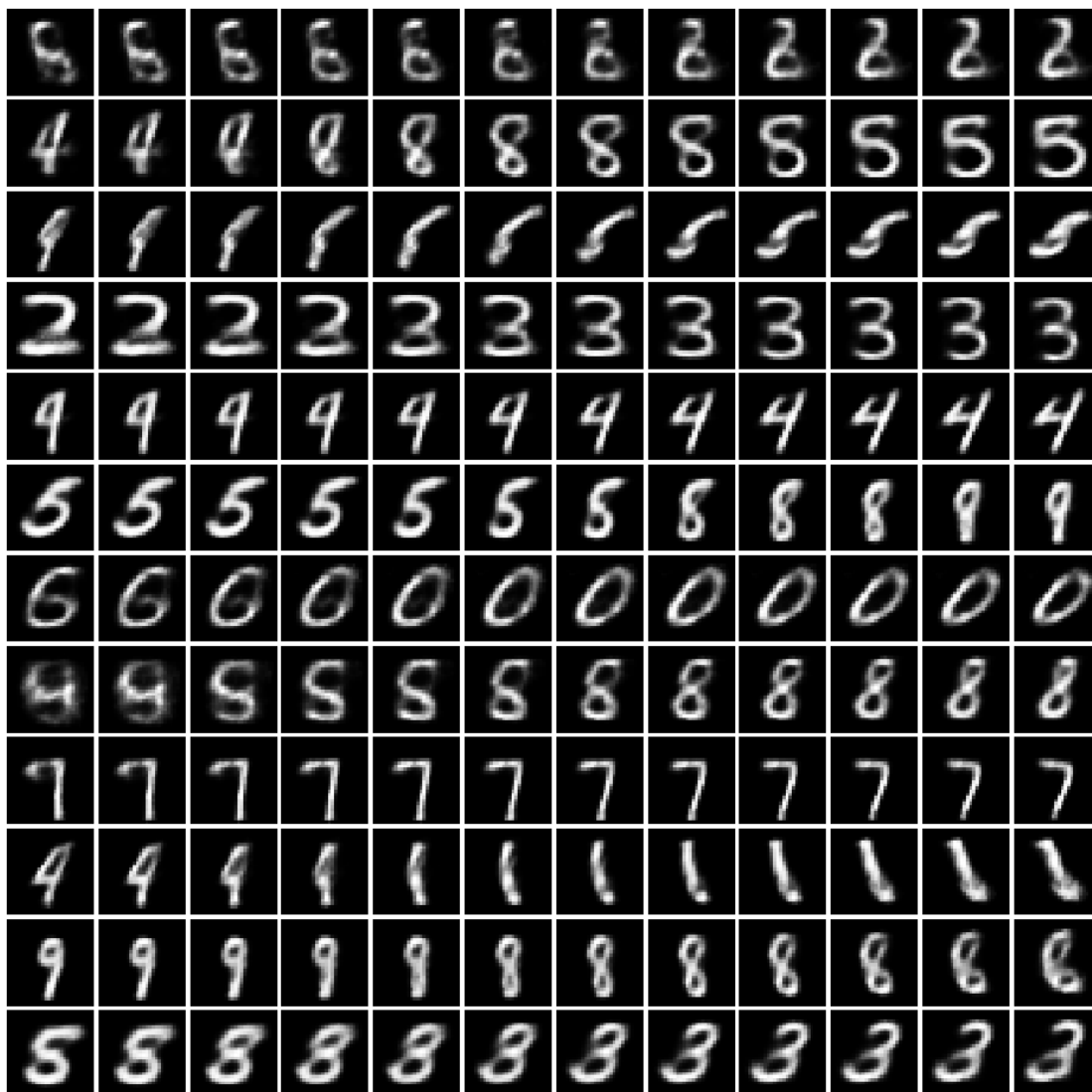


2.8 Latent Space Interpolation

As a final visual test of our trained VAE model, we can perform interpolation in latent space. We generate random latent vectors z_0 and z_1 , and linearly interpolate between them; we run each interpolated vector through the trained generator to produce an image.

Each row of the figure below interpolates between two random vectors. For the most part the model should exhibit smooth transitions along each row, demonstrating that the model has learned something nontrivial about the underlying spatial structure of the digits it is modeling.

```
[27]: S = 12
latent_size = 15
device = next(vae_model.parameters()).device
z0 = torch.randn(S, latent_size, device=device)
z1 = torch.randn(S, latent_size, device=device)
w = torch.linspace(0, 1, S, device=device).view(S, 1, 1)
z = (w * z0 + (1 - w) * z1).transpose(0, 1).reshape(S * S, latent_size)
x = vae_model.decoder(z)
show_images(x.data.cpu())
```



3 Conditional FC-VAE

The second model you'll develop will be very similar to the FC-VAE, but with a slight conditional twist to it. We'll use what we know about the labels of each MNIST image, and *condition* our latent space and image generation on the specific class. Instead of $q_\phi(z|x)$ and $p_\phi(x|z)$ we have $q_\phi(z|x, c)$ and $p_\phi(x|z, c)$

This will allow us to do some powerful conditional generation at inference time. We can specifically choose to generate more 1s, 2s, 9s, etc. instead of simply generating new digits randomly.

3.1 Define Network with class input (3 points)

Our CVAE architecture will be the same as our FC-VAE architecture, except we'll now add a one-hot label vector to both the x input (in our case, the flattened image dimensions) and the z

latent space.

If our one-hot vector is called `c`, then `c[label] = 1` and `c = 0` elsewhere.

For the CVAE class in `vae.py` use the same FC-VAE architecture implemented in the last network with the following modifications:

1. Modify the first linear layer of your `encoder` to take in not only the flattened input image, but also the one-hot label vector `c`. The CVAE `encoder` should not have a `Flatten` layer.
2. Modify the first layer of your `decoder` to project the latent space + one-hot vector to the `hidden_dim`
3. Lastly, implement the `forward` pass to combine the flattened input image with the one-hot vectors (`torch.cat`) before passing them to the `encoder` and combine the latent space with the one-hot vectors (`torch.cat`) before passing them to the `decoder`. You should flatten the image before concatenation (e.g. with `torch.flatten` or `torch.reshape`).

```
[30]: from vae import CVAE
def test_CVAE_shapes():
    all_shapes_correct = True
    with torch.no_grad():
        batch_size = 3
        num_classes = 10
        latent_size = 17
        cls = nn.functional.one_hot(torch.tensor([3]*batch_size, dtype=torch.
↳long), num_classes=num_classes)
        x_hat, mu, logvar = CVAE(28*28,
↳num_classes=num_classes,latent_size=latent_size)(
            torch.ones(batch_size, 1, 28, 28), cls)
        if x_hat.shape != (batch_size, 1, 28, 28):
            print(f'x_hat has incorrect shape. Expected (batch_size, 1, 28, 28)
↳= ({batch_size}, 1, 28, 28).')
            f' Got {tuple(x_hat.shape)}.'
            all_shapes_correct = False
        if mu.shape != (batch_size, latent_size):
            print(f'mu has incorrect shape. Expected (batch_size, latent_size)
↳= ({batch_size}, {latent_size}).')
            f' Got {tuple(mu.shape)}.'
            all_shapes_correct = False
        if logvar.shape != (batch_size, latent_size):
            print(f'logvar has incorrect shape. Expected (batch_size,
↳latent_size) = ({batch_size}, {latent_size}).')
            f' Got {tuple(logvar.shape)}.'
            all_shapes_correct = False
        if all_shapes_correct:
            print('Shapes of x_hat, mu, and logvar are correct.')
        if batch_size > 1 and x_hat.std(0).mean() == 0:
            print('x_hat has no randomness.')
    return
```

```
test_CVAE_shapes()
```

Shapes of `x_hat`, `mu`, and `logvar` are correct.

3.2 Train model

Using the same training script, let's now train our CVAE!

Training for 10 epochs should take ~2 minutes and your loss should be less than 120.

```
[31]: from vae import CVAE
num_epochs = 10
latent_size = 15
from nndl2.helper import train_vae
input_size = 28*28
device = 'cuda' if torch.cuda.is_available() else 'cpu'
if device == 'cpu':
    print(f'Warning: using device {device} may take longer.')

cvae = CVAE(input_size, latent_size=latent_size)
cvae.to(device)
for epoch in range(0, num_epochs):
    train_vae(epoch, cvae, loader_train, cond=True)
```

```
Train Epoch: 0  Loss: 135.927734
Train Epoch: 1  Loss: 122.679947
Train Epoch: 2  Loss: 112.725868
Train Epoch: 3  Loss: 109.824448
Train Epoch: 4  Loss: 107.520432
Train Epoch: 5  Loss: 108.468788
Train Epoch: 6  Loss: 105.107819
Train Epoch: 7  Loss: 108.116425
Train Epoch: 8  Loss: 104.715103
Train Epoch: 9  Loss: 101.960243
```

3.3 Visualize Results

We've trained our CVAE, now let's conditionally generate some new data! This time, we can specify the class we want to generate by adding our one hot matrix of class labels. We use `torch.eye` to create an identity matrix, which effectively gives us one label for each digit. When you run the cell below, you should get one example per digit. Each digit should be reasonably distinguishable (it is ok to run this cell a few times to save your best results).

```
[33]: device = next(cvae.parameters()).device
z = torch.randn(10, latent_size)
c = torch.eye(10, 10) # [one hot labels for 0-9]
import matplotlib.gridspec as gridspec
z = torch.cat((z,c), dim=-1).to(device=device)
cvae.eval()
```

```
samples = cvae.decoder(z).data.cpu().numpy()

fig = plt.figure(figsize=(10, 1))
gspec = gridspec.GridSpec(1, 10)
gspec.update(wspace=0.05, hspace=0.05)
for i, sample in enumerate(samples):
    ax = plt.subplot(gspec[i])
    plt.axis('off')
    ax.set_xticklabels([])
    ax.set_yticklabels([])
    ax.set_aspect('equal')
    plt.imshow(sample.reshape(28, 28), cmap='Greys_r')
```




```

# %load vae.py
from __future__ import print_function
import matplotlib.pyplot as plt
import matplotlib.gridspec as gridspec
import numpy as np
import torch
import torch.utils.data
from torch import nn, optim
from torch.autograd import Variable
from torch.nn import functional as F
from torchvision import datasets, transforms
from torchvision.utils import save_image

def hello_vae():
    print("Hello from vae.py!")

class VAE(nn.Module):
    def __init__(self, input_size, latent_size=15):
        super(VAE, self).__init__()
        self.input_size = input_size # 784 # H*W
        self.latent_size = latent_size # Z
        self.hidden_dim = 400 # H_d
        self.encoder = None
        self.mu_layer = None
        self.logvar_layer = None
        self.decoder = None

#####
# TODO: Implement the fully-connected encoder architecture described in the notebook.
#
# Specifically, self.encoder should be a network that inputs a batch of input images
of #
# shape (N, 1, H, W) into a batch of hidden features of shape (N, H_d). Set up
#
# self.mu_layer and self.logvar_layer to be a pair of linear layers that map the hidden
#
# features into estimates of the mean and log-variance of the posterior over the latent
#
# vectors; the mean and log-variance estimates will both be tensors of shape (N, Z).
#

#####
# Replace "pass" statement with your code
self.encoder = nn.Sequential(
    nn.Flatten(),
    nn.Linear(self.input_size, self.hidden_dim),
    nn.LeakyReLU(0.01),
    nn.Linear(self.hidden_dim, self.hidden_dim),
    nn.LeakyReLU(0.01),
    nn.Linear(self.hidden_dim, self.hidden_dim),
    nn.LeakyReLU(0.01),
)
self.mu_layer = nn.Linear(self.hidden_dim, self.latent_size)

self.logvar_layer = nn.Linear(self.hidden_dim, self.latent_size)

#####
# TODO: Implement the fully-connected decoder architecture described in the notebook.
#
# Specifically, self.decoder should be a network that inputs a batch of latent vectors
of #
# shape (N, Z) and outputs a tensor of estimated images of shape (N, 1, H, W).
#

```



```
#####  
    return x_hat, mu, logvar  
  
class CVAE(nn.Module):  
    def __init__(self, input_size, num_classes=10, latent_size=15):  
        super(CVAE, self).__init__()  
        self.input_size = input_size # H*W  
        self.latent_size = latent_size # Z  
        self.num_classes = num_classes # K  
        self.hidden_dim = 400 # H_d  
        self.encoder = None  
        self.mu_layer = None  
        self.logvar_layer = None  
        self.decoder = None  
  
#####  
        # TODO: Define a FC encoder as described in the notebook that transforms the image--  
after #  
        # flattening and now adding our one-hot class vector (N, H*W + K)--into a  
hidden_dimension #  
        # (N, H_d) feature space, and a final two layers that project that feature space  
#  
        # to posterior mu and posterior log-variance estimates of the latent space (N, Z)  
#  
#####  
        # Replace "pass" statement with your code  
        self.encoder = nn.Sequential(  
            nn.Linear(self.input_size + self.num_classes, self.hidden_dim),  
            nn.LeakyReLU(0.01),  
            nn.Linear(self.hidden_dim, self.hidden_dim),  
            nn.LeakyReLU(0.01),  
            nn.Linear(self.hidden_dim, self.hidden_dim),  
            nn.LeakyReLU(0.01),  
        )  
        self.mu_layer = nn.Linear(self.hidden_dim, self.latent_size)  
  
        self.logvar_layer = nn.Linear(self.hidden_dim, self.latent_size)  
  
#####  
        # TODO: Define a fully-connected decoder as described in the notebook that transforms  
the #  
        # latent space (N, Z + K) to the estimated images of shape (N, 1, H, W).  
#  
#####  
        # Replace "pass" statement with your code  
        self.decoder = nn.Sequential(  
            nn.Linear(self.latent_size + self.num_classes, self.hidden_dim),  
            nn.LeakyReLU(0.01),  
            nn.Linear(self.hidden_dim, self.hidden_dim),  
            nn.LeakyReLU(0.01),  
            nn.Linear(self.hidden_dim, self.hidden_dim),  
            nn.LeakyReLU(0.01),  
            nn.Linear(self.hidden_dim, self.input_size),  
            nn.Sigmoid(),  
            nn.Unflatten(dim =1 ,unflattened_size=(1, 28, 28) )  
        )  
  
#####  
        #  
END OF YOUR CODE  
#
```

```
def forward(self, x, labels):
    """
    Performs forward pass through FC-CVAE model by passing image through
    encoder, reparametrize trick, and decoder models

    Inputs:
    - x: Input data for this timestep of shape (N, 1, H, W)
    - labels: One hot vector representing the input class (0-9) (N, K)

    Returns:
    - x_hat: Reconstructed input data of shape (N, 1, H, W)
    - mu: Matrix representing estimated posterior mu (N, Z), with Z latent space dimension
    - logvar: Matrix representing estimated variance in log-space (N, Z), with Z latent
    space dimension
    """
    x_hat = None
    mu = None
    logvar = None

#####
    # TODO: Implement the forward pass by following these steps
#
# (1) Pass the concatenation of input batch and one hot vectors through the encoder
model #
# to get posterior mu and logvariance
#
# (2) Reparametrize to compute the latent vector z
#
# (3) Pass concatenation of z and one hot vectors through the decoder to reconstruct x
#
#####
    # Replace "pass" statement with your code
    # step 1:
    N = x.size(0)

    x_flat = x.view(N, -1) # (N, H*W)
    enc_input = torch.cat([x_flat, labels], dim=1) # (N, H*W + K)
    h = self.encoder(enc_input) # (N, hidden_dim)
    mu = self.mu_layer(h) # (N, Z)
    logvar = self.logvar_layer(h) # (N, Z)

    # step 2:
    z = reparametrize(mu, logvar) # (N, Z)

    # step 3:
    dec_input = torch.cat([z, labels], dim=1) # (N, Z + K)
    x_hat = self.decoder(dec_input) # (N, 1, H, W)

#####
    #
#
#####
    return x_hat, mu, logvar

def reparametrize(mu, logvar):
    """
    Differentiably sample random Gaussian data with specified mean and variance using the
    reparameterization trick.

    Suppose we want to sample a random number z from a Gaussian distribution with mean mu and
    standard deviation sigma, such that we can backpropagate from the z back to mu and sigma.
    """
```

We can achieve this by first sampling a random value ϵ from a standard Gaussian distribution with zero mean and unit variance, then setting $z = \sigma * \epsilon + \mu$.

For more stable training when integrating this function into a neural network, it helps to pass this function the log of the variance of the distribution from which to sample, rather than specifying the standard deviation directly.

Inputs:

- μ : Tensor of shape (N, Z) giving means
- logvar: Tensor of shape (N, Z) giving log-variances

Returns:

- z: Estimated latent vectors, where $z[i, j]$ is a random value sampled from a Gaussian with mean $\mu[i, j]$ and log-variance $\logvar[i, j]$.

"""

z = None

#####

TODO: Reparametrize by initializing epsilon as a normal distribution and scaling by

#

posterior mu and sigma to estimate z

#

#####

Replace "pass" statement with your code

sigma = torch.exp(0.5 * logvar)

epsilon = torch.randn_like(sigma)

z = sigma*epsilon + mu

#####

END OF YOUR CODE

#

#####

return z

def loss_function(x_hat, x, mu, logvar):

"""

Computes the negative variational lower bound loss term of the VAE (refer to formulation in notebook).

Inputs:

- x_hat: Reconstructed input data of shape (N, 1, H, W)
- x: Input data for this timestep of shape (N, 1, H, W)
- mu: Matrix representing estimated posterior μ (N, Z), with Z latent space dimension
- logvar: Matrix representing estimated variance in log-space (N, Z), with Z latent space dimension

Returns:

- loss: Tensor containing the scalar loss for the negative variational lowerbound

"""

loss = None

#####

TODO: Compute negative variational lowerbound loss as described in the notebook

#

#####

Replace "pass" statement with your code

N = x.size(0)

```
# flatten images to (N, D) for BCE
x_hat_flat = x_hat.view(N, -1)
x_flat      = x.view(N, -1)

# Reconstruction term:
recon_loss = F.binary_cross_entropy(
    x_hat_flat, x_flat,
    reduction='sum'
) / N

# KL divergence term:
kl_per_sample = -0.5 * torch.sum(
    1 + logvar - mu.pow(2) - logvar.exp(),
    dim=1
)
kl_loss = kl_per_sample.mean()

loss = recon_loss + kl_loss
```

#####

```
#                                     END OF YOUR CODE
#
```

#####

```
return loss
```