Generative Adversarial Networks (GANs)

So far in CS231N, 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. This has ranged from straightforward classification of image categories to sentence generation (which was still phrased as a classification problem, our labels were in vocabulary space and we'd learned a recurrence to capture multi-word labels). In this notebook, we will expand our repetoire, 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.

What is a GAN?

In 2014, [Goodfellow et al.](https://arxiv.org/abs/1406.2661) 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: $$\underset{G}{\text{minimize}}\; \underset{D}{\text{maximize}}\; \mathbb{E}_{x \sim p_\text{data}}\left[\log D(x)\right] + \mathbb{E}_{z \sim p(z)}\left[\log \left(1-D(G(z))\right)\right]$$ where $x \sim p_\text{data}$ are samples from the input data, $z \sim p(z)$ 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.](https://arxiv.org/abs/1406.2661), 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 aternate 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 allevaiate 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.](https://arxiv.org/abs/1406.2661).

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: $$\underset{G}{\text{maximize}}\; \mathbb{E}_{z \sim p(z)}\left[\log D(G(z))\right]$$
2. Update the discriminator (D), to maximize the probability of the discriminator making the

2. Update the discriminator (D), to maximize the probability of the discriminator making the correct choice on real and generated data: $$\underset{D}{\text{maximize}}\; \mathbb{E}_{x \sim p_\text{data}}\left[\log D(x)\right] + \mathbb{E}_{z \sim p(z)}\left[\log \left(1-D(G(z))\right)\right]$$

What else is there?

Since 2014, GANs have exploded into a huge research area, with massive workshops, and hundreds of new papers. Compared to other approaches for generative models, they often produce the highest quality samples but are some of the most difficult and finicky models to train (see this github repo that contains a set of 17 hacks that are useful for getting models working). Improving the stabiilty and robustness of GAN training is an open research question, with new papers coming out every day! For a more recent tutorial on GANs, see here. There is also some even more recent exciting work that changes the objective function to Wasserstein distance and yields much more stable results across model architectures: WGAN, WGAN-GP.

GANs are not the only way to train a generative model! For other approaches to generative modeling check out the deep generative model chapter of the Deep Learning book. Another popular way of training neural networks as generative models is Variational Autoencoders (co-discovered here and here). Variational autoencoders combine neural networks with variational inference to train deep generative models. These models tend to be far more stable and easier to train but currently don't produce samples that are as pretty as GANs.

Example pictures of what you should expect (yours might look slightly different):

Setup

In [212]:

```
import tensorflow as tf
import numpy as np
import os

import matplotlib.pyplot as plt
import matplotlib.gridspec as gridspec

%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'

# A bunch of utility functions

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

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():
    """Count the number of parameters in the current TensorFlow graph """
    param_count = np.sum([np.prod(x.get_shape().as_list()) for x in
tf.global_variables()])
    return param_count


def get_session():
    config = tf.ConfigProto()
    config.gpu_options.allow_growth = True
    session = tf.Session(config=config)
    return session

answers = np.load('gan-checks-tf.npz')
```

Dataset

GANs are notoriously finicky with hyperparameters, and also require many training epochs. In order to make this assignment approachable without a GPU, 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.

Heads-up: Our MNIST wrapper returns images as vectors. That is, they're size (batch, 784). If you want to treat them as images, we have to resize them to (batch,28,28) or (batch,28,28,1). They are also type np.float32 and bounded [0,1].

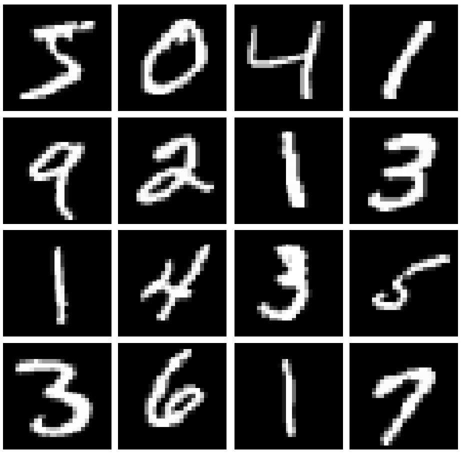
In [213]:

```
class MNIST(object):
    def __init__(self, batch_size, shuffle=False):
        """
        Construct an iterator object over the MNIST data

        Inputs:
        - batch_size: Integer giving number of elements per minibatch
        - shuffle: (optional) Boolean, whether to shuffle the data on each
epoch
        """
        train, _ = tf.keras.datasets.mnist.load_data()
```

```
        X, y = train
        X = X.astype(np.float32)/255
        X = X.reshape((X.shape[0], -1))
        self.X, self.y = X, y
        self.batch_size, self.shuffle = batch_size, shuffle

    def __iter__(self):
        N, B = self.X.shape[0], self.batch_size
        idxs = np.arange(N)
        if self.shuffle:
            np.random.shuffle(idxs)
        return iter((self.X[i:i+B], self.y[i:i+B]) for i in range(0, N, B))
```

In [214]:

```
# show a batch
mnist = MNIST(batch_size = 16)
show_images(mnist.X[:16])
```



LeakyReLU

In the cell below, you should implement a LeakyReLU. See the [class notes](#) (where alpha is small number) or equation (3) in [this paper](#). LeakyReLUs keep ReLU units from dying and are often used in GAN methods (as are maxout units, however those increase model size and therefore are not used in this notebook).

HINT: You should be able to use `tf.maximum`

In [215]:

```
def leaky_relu(x, alpha=0.01):
    """Compute the leaky ReLU activation function.

    Inputs:
    - x: TensorFlow Tensor with arbitrary shape
    - alpha: leak parameter for leaky ReLU

    Returns:
    TensorFlow Tensor with the same shape as x
    """
    # TODO: implement leaky ReLU
    #pass

    return tf.maximum(x, alpha*x)
```

```
    return tf.maximum(x,alpha*x)
```

Test your leaky ReLU implementation. You should get errors < 1e-10

In [216]:

```
def test_leaky_relu(x, y_true):
    tf.reset_default_graph()
    with get_session() as sess:
        y_tf = leaky_relu(tf.constant(x))
        y = sess.run(y_tf)
        print('Maximum error: %g'%rel_error(y_true, y))

test_leaky_relu(answers['lrelu_x'], answers['lrelu_y'])
```

```
Maximum error: 0
```

Random Noise

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

In [217]:

```
def sample_noise(batch_size, dim):
    """Generate random uniform noise from -1 to 1.

    Inputs:
    - batch_size: integer giving the batch size of noise to generate
    - dim: integer giving the dimension of the the noise to generate

    Returns:
    TensorFlow Tensor containing uniform noise in [-1, 1] with shape [batch
_size, dim]
    """
    # TODO: sample and return noise
    #pass

    return tf.random_uniform(shape = [batch_size, dim],  minval = -1, maxval
= 1)
```

Make sure noise is the correct shape and type:

In [218]:

```
def test_sample_noise():
    batch_size = 3
    dim = 4
    tf.reset_default_graph()
    with get_session() as sess:
        z = sample_noise(batch_size, dim)
        # Check z has the correct shape
        assert z.get_shape().as_list() == [batch_size, dim]
        # Make sure z is a Tensor and not a numpy array
        assert isinstance(z, tf.Tensor)
        # Check that we get different noise for different evaluations
```

```
        z1 = sess.run(z)
        z2 = sess.run(z)
        assert not np.array_equal(z1, z2)
        # Check that we get the correct range
        assert np.all(z1 >= -1.0) and np.all(z1 <= 1.0)
        print("All tests passed!")

test_sample_noise()
```

```
All tests passed!
```

Discriminator

Our first step is to build a discriminator. You should use the layers in `tf.layers` to build the model. All fully connected layers should include bias terms. For initialization, just use the default initializer used by the `tf.layers` functions.

Architecture:

- Fully connected layer with input size 784 and output size 256
- LeakyReLU with alpha 0.01
- Fully connected layer with output size 256
- LeakyReLU with alpha 0.01
- Fully connected layer with output size 1

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

In [219]:

```
def discriminator(x):
    """Compute discriminator score for a batch of input images.

    Inputs:
    - x: TensorFlow Tensor of flattened input images, shape [batch_size, 78
4]

    Returns:
    TensorFlow Tensor with shape [batch_size, 1], containing the score
    for an image being real for each input image.
    """
    with tf.variable_scope("discriminator"):
        # TODO: implement architecture
        #pass
        fc1 = tf.layers.dense(inputs = x, units = 256, activation = leaky_re
lu, use_bias=True)
        fc2 = tf.layers.dense(inputs = fc1, units = 256, activation = leaky_
relu, use_bias=True)
        logits = tf.layers.dense(inputs = fc2, units = 1)


        return logits
```

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

In [220]:

```python
def test_discriminator(true_count=267009):
    tf.reset_default_graph()
    with get_session() as sess:
        y = discriminator(tf.ones((2, 784)))
        cur_count = count_params()
        if cur_count != true_count:
            print('Incorrect number of parameters in discriminator. {0}
instead of {1}. Check your achitecture.'.format(cur_count,true_count))
        else:
            print('Correct number of parameters in discriminator.')

test_discriminator()
```

```
Correct number of parameters in discriminator.
```

In [142]:

```python

```

Out[142]:

```
<tf.Tensor 'ones_1:0' shape=(2, 784) dtype=float32>
```

Generator

Now to build a generator. You should use the layers in `tf.layers` to construct the model. All fully connected layers should include bias terms. Note that you can use the tf.nn module to access activation functions. Once again, use the default initializers for parameters.

Architecture:

- Fully connected layer with inupt size tf.shape(z)[1] (the number of noise dimensions) and output size 1024
- `ReLU`
- Fully connected layer with output size 1024
- `ReLU`
- Fully connected layer with output size 784
- `TanH` (To restrict every element of the output to be in the range [-1,1])

In [221]:

```python
def generator(z):
    """Generate images from a random noise vector.

    Inputs:
    - z: TensorFlow Tensor of random noise with shape [batch_size, noise_di
m]

    Returns:
    TensorFlow Tensor of generated images, with shape [batch_size, 784].
    """
    with tf.variable_scope("generator"):
        # TODO: implement architecture
        #pass
        fc1 = tf.layers.dense(inputs = z, units = 1024, activation = tf.nn.r
elu, use_bias = True)
        fc2 = tf.layers.dense(inputs = fc1, units = 1024, activation = tf.nn
.relu, use_bias = True)
```

```
        img = tf.layers.dense(inputs = fc2, units = 784, activation = tf.nn.
tanh, use_bias = True)



        return img
```

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

In [222]:

```
def test_generator(true_count=1858320):
    tf.reset_default_graph()
    with get_session() as sess:
        y = generator(tf.ones((1, 4)))
        cur_count = count_params()
        if cur_count != true_count:
            print('Incorrect number of parameters in generator. {0} instead
of {1}. Check your achitecture.'.format(cur_count,true_count))
        else:
            print('Correct number of parameters in generator.')

test_generator()
```

```
Correct number of parameters in generator.
```

GAN Loss

Compute the generator and discriminator loss. The generator loss is: $$\ell_G = -\mathbb{E}_{z \sim p(z)}\left[\log D(G(z))\right]$$ and the discriminator loss is: $$ \ell_D = -\mathbb{E}_{x \sim p_\text{data}}\left[\log D(x)\right] - \mathbb{E}_{z \sim p(z)}\left[\log \left(1-D(G(z))\right)\right]$$ Note that these are negated from the equations presented earlier as we will be *minimizing* these losses.

HINTS: Use tf.ones_like and tf.zeros_like to generate labels for your discriminator. Use tf.nn.sigmoid_cross_entropy_with_logits to help compute your loss function. 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.

In [263]:

```
def gan_loss(logits_real, logits_fake):
    """Compute the GAN loss.

    Inputs:
    - logits_real: Tensor, shape [batch_size, 1], output of discriminator
        Unnormalized score that the image is real for each real image
    - logits_fake: Tensor, shape[batch_size, 1], output of discriminator
        Unnormalized score that the image is real for each fake image

    Returns:
    - D_loss: discriminator loss scalar
    - G_loss: generator loss scalar

    HINT: for the discriminator loss, you'll want to do the averaging separ
ately for
    its two components, and then add them together (instead of averaging
```

```
once at the very end).
    """
    # TODO: compute D_loss and G_loss
    #D_loss = None
    #G_loss = None
    #pass

    labels_real = tf.ones_like(logits_real)
    labels_fake = tf.zeros_like(logits_fake)

    D_loss = tf.reduce_mean(tf.nn.sigmoid_cross_entropy_with_logits(labels=
labels_real, logits=logits_real))
    D_loss += tf.reduce_mean(tf.nn.sigmoid_cross_entropy_with_logits(labels
=labels_fake, logits=logits_fake))

    labels_fake = tf.ones_like(logits_fake)
    G_loss = tf.reduce_mean(tf.nn.sigmoid_cross_entropy_with_logits(labels=
labels_fake, logits=logits_fake))
    return D_loss, G_loss
```

Test your GAN loss. Make sure both the generator and discriminator loss are correct. You should see errors less than 1e-5.

In [264]:

```
def test_gan_loss(logits_real, logits_fake, d_loss_true, g_loss_true):
    tf.reset_default_graph()
    with get_session() as sess:
        d_loss, g_loss = sess.run(gan_loss(tf.constant(logits_real), tf.cons
tant(logits_fake)))
    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_gan_loss(answers['logits_real'], answers['logits_fake'],
              answers['d_loss_true'], answers['g_loss_true'])
```

```
Maximum error in d_loss: 6.02597e-17
Maximum error in g_loss: 7.19722e-17
```

In [265]:

```
type(answers['d_loss_true'])
```

Out[265]:

```
numpy.ndarray
```

Optimizing our loss

Make an `AdamOptimizer` with a 1e-3 learning rate, beta1=0.5 to mininize G_loss and D_loss separately. The trick of decreasing beta was shown to be effective in helping GANs converge in the [Improved Techniques for Training GANs](#) paper. In fact, with our current hyperparameters, if you set beta1 to the Tensorflow default of 0.9, there's a good chance your discriminator loss will go to zero and the generator will fail to learn entirely. In fact, this is a common failure mode in GANs; if your D(x) learns to be too fast (e.g. loss goes near zero), your G(z) is never able to learn. Often D(x) is trained with SGD with Momentum or RMSProp instead of Adam, but here we'll use Adam for both D(x) and

G(z).

In [266]:

```
# TODO: create an AdamOptimizer for D_solver and G_solver
def get_solvers(learning_rate=1e-3, beta1=0.5):
    """Create solvers for GAN training.

    Inputs:
    - learning_rate: learning rate to use for both solvers
    - beta1: beta1 parameter for both solvers (first moment decay)

    Returns:
    - D_solver: instance of tf.train.AdamOptimizer with correct learning_ra
te and beta1
    - G_solver: instance of tf.train.AdamOptimizer with correct learning_ra
te and beta1
    """
    #D_solver = None
    #G_solver = None
    #pass
    D_solver = tf.train.AdamOptimizer(learning_rate, beta1)
    G_solver = tf.train.AdamOptimizer(learning_rate, beta1)
    return D_solver, G_solver
```

Putting it all together

Now just a bit of Lego Construction.. Read this section over carefully to understand how we'll be composing the generator and discriminator

In [267]:

```
tf.reset_default_graph()

# number of images for each batch
batch_size = 128
# our noise dimension
noise_dim = 96

# placeholder for images from the training dataset
x = tf.placeholder(tf.float32, [None, 784])
# random noise fed into our generator
z = sample_noise(batch_size, noise_dim)
# generated images
G_sample = generator(z)

with tf.variable_scope("") as scope:
    #scale images to be -1 to 1
    logits_real = discriminator(preprocess_img(x))
    # Re-use discriminator weights on new inputs
    scope.reuse_variables()
    logits_fake = discriminator(G_sample)

# Get the list of variables for the discriminator and generator
D_vars = tf.get_collection(tf.GraphKeys.TRAINABLE_VARIABLES,
'discriminator')
G_vars = tf.get_collection(tf.GraphKeys.TRAINABLE_VARIABLES, 'generator')
```

```
# get our solver
D_solver, G_solver = get_solvers()

# get our loss
D_loss, G_loss = gan_loss(logits_real, logits_fake)

# setup training steps
D_train_step = D_solver.minimize(D_loss, var_list=D_vars)
G_train_step = G_solver.minimize(G_loss, var_list=G_vars)
D_extra_step = tf.get_collection(tf.GraphKeys.UPDATE_OPS, 'discriminator')
G_extra_step = tf.get_collection(tf.GraphKeys.UPDATE_OPS, 'generator')
```

Training a GAN!

Well that wasn't so hard, was it? After the first epoch, you should see fuzzy outlines, clear shapes as you approach epoch 3, and decent shapes, about half of which will be sharp and clearly recognizable as we pass epoch 5. In our case, we'll simply train D(x) and G(z) with one batch each every iteration. However, papers often experiment with different schedules of training D(x) and G(z), sometimes doing one for more steps than the other, or even training each one until the loss gets "good enough" and then switching to training the other.

In [268]:

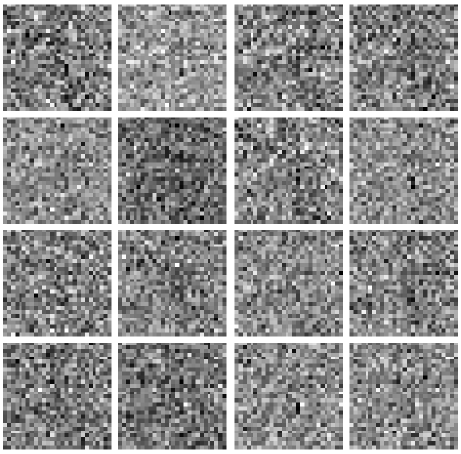
```
# a giant helper function
def run_a_gan(sess, G_train_step, G_loss, D_train_step, D_loss,
G_extra_step, D_extra_step,\
              show_every=2, print_every=1, batch_size=128, num_epoch=10):
    """Train a GAN for a certain number of epochs.

    Inputs:
    - sess: A tf.Session that we want to use to run our data
    - G_train_step: A training step for the Generator
    - G_loss: Generator loss
    - D_train_step: A training step for the Generator
    - D_loss: Discriminator loss
    - G_extra_step: A collection of tf.GraphKeys.UPDATE_OPS for generator
    - D_extra_step: A collection of tf.GraphKeys.UPDATE_OPS for discriminat
or
    Returns:
        Nothing
    """
    # compute the number of iterations we need
    mnist = MNIST(batch_size=batch_size, shuffle=True)
    for epoch in range(num_epoch):
        # every show often, show a sample result
        if epoch % show_every == 0:
            samples = sess.run(G_sample)
            fig = show_images(samples[:16])
            plt.show()
            print()
        for (minibatch, minbatch_y) in mnist:
            # run a batch of data through the network
            _, D_loss_curr = sess.run([D_train_step, D_loss], feed_dict={x:
minibatch})
            _, G_loss_curr = sess.run([G_train_step, G_loss])

        # print loss every so often.
```

```
        # we want to make sure D_loss doesn't go to 0
        if epoch % print_every == 0:
            print('Epoch: {}, D: {:.4}, G:{:.4}'.format(epoch,D_loss_curr,G_
loss_curr))
    print('Final images')
    samples = sess.run(G_sample)

    fig = show_images(samples[:16])
    plt.show()
```

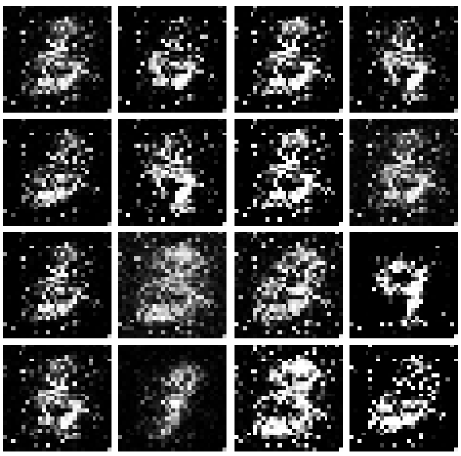
Train your GAN! This should take about 10 minutes on a CPU, or less than a minute on GPU.

In [269]:

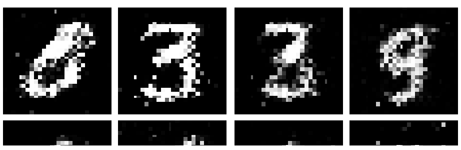
```
with get_session() as sess:
    sess.run(tf.global_variables_initializer())
    run_a_gan(sess,G_train_step,G_loss,D_train_step,D_loss,G_extra_step,D_e
xtra_step)
```

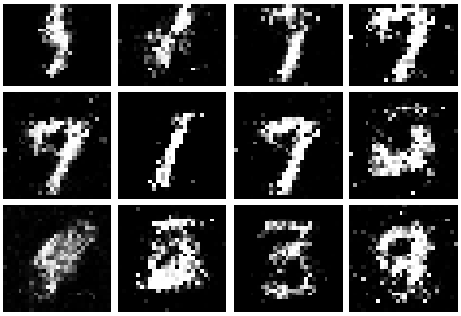


```
Epoch: 0, D: 1.231, G:0.9907
Epoch: 1, D: 1.352, G:1.047
```

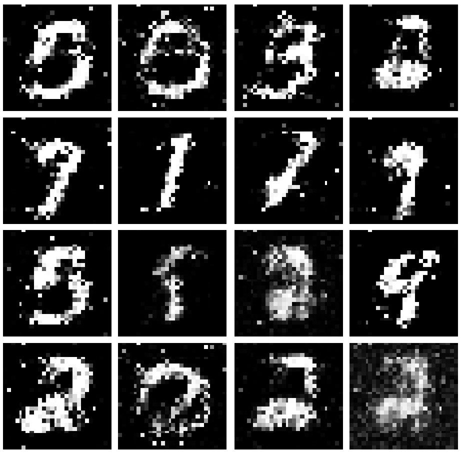


```
Epoch: 2, D: 1.141, G:2.549
Epoch: 3, D: 1.322, G:0.9026
```

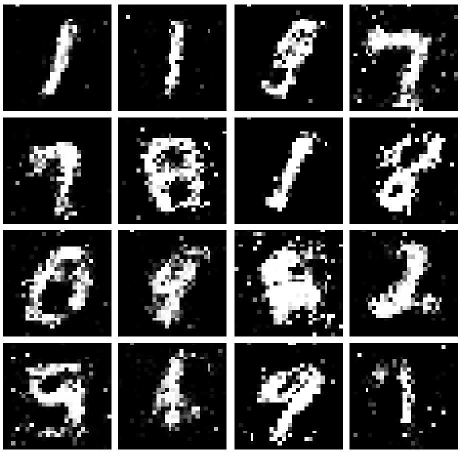




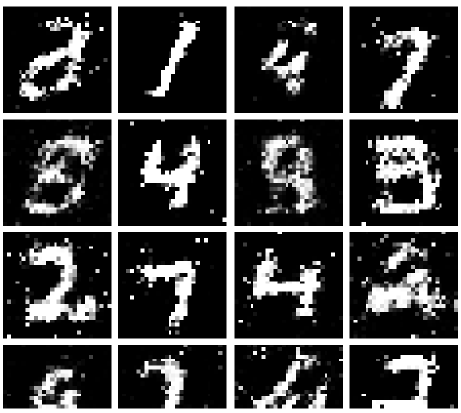
```
Epoch: 4, D: 1.182, G:0.9019
Epoch: 5, D: 1.497, G:0.8338
```

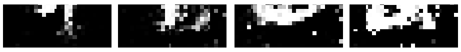


```
Epoch: 6, D: 1.325, G:0.8605
Epoch: 7, D: 1.403, G:0.8567
```



```
Epoch: 8, D: 1.473, G:0.8296
Epoch: 9, D: 1.205, G:0.8509
Final images
```





Least Squares GAN

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: $$\ell_G = \frac{1}{2}\mathbb{E}_{z \sim p(z)}\left[\left(D(G(z))-1\right)^2\right]$$ and the discriminator loss: $$ \ell_D = \frac{1}{2}\mathbb{E}_{x \sim p_\text{data}}\left[\left(D(x)-1\right)^2\right] + \frac{1}{2}\mathbb{E}_{z \sim p(z)}\left[\left(D(G(z))\right)^2\right]$$

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 (`score_real` and `score_fake`).

In [271]:

```
def lsgan_loss(scores_real, scores_fake):
    """Compute the Least Squares GAN loss.

    Inputs:
    - scores_real: Tensor, shape [batch_size, 1], output of discriminator
        The score for each real image
    - scores_fake: Tensor, shape[batch_size, 1], output of discriminator
        The score for each fake image

    Returns:
    - D_loss: discriminator loss scalar
    - G_loss: generator loss scalar
    """
    # TODO: compute D_loss and G_loss
    #D_loss = None
    #G_loss = None
    #pass
    true_labels = tf.ones_like(scores_fake)
    fake_image_loss = tf.reduce_mean((scores_real - true_labels)**2)
    real_image_loss = tf.reduce_mean(scores_fake**2)
    D_loss = 0.5*(fake_image_loss + real_image_loss)
    G_loss = 0.5*tf.reduce_mean((scores_fake - true_labels)**2)
    return D_loss, G_loss
```

Test your LSGAN loss. You should see errors less than 1e-7.

In [272]:

```
def test_lsgan_loss(score_real, score_fake, d_loss_true, g_loss_true):
    with get_session() as sess:
        d_loss, g_loss = sess.run(
            lsgan_loss(tf.constant(score_real), tf.constant(score_fake)))
    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: 0
```

```
Maximum error in g_loss: 0
```

Create new training steps so we instead minimize the LSGAN loss:

In [273]:

```
D_loss, G_loss = lsgan_loss(logits_real, logits_fake)
D_train_step = D_solver.minimize(D_loss, var_list=D_vars)
G_train_step = G_solver.minimize(G_loss, var_list=G_vars)
```

Run the following cell to train your model!

In [274]:

```
with get_session() as sess:
    sess.run(tf.global_variables_initializer())
    run_a_gan(sess, G_train_step, G_loss, D_train_step, D_loss,
G_extra_step, D_extra_step)
```

```
---------------------------------------------------------------------------
InvalidArgumentError                      Traceback (most recent call last)
/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/client/session.py in _do_call(self, fn, *args)
   1326     try:
-> 1327       return fn(*args)
   1328     except errors.OpError as e:

/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/client/session.py in _run_fn(feed_dict, fetch_li
st, target_list, options, run_metadata)
   1311       return self._call_tf_sessionrun(
-> 1312           options, feed_dict, fetch_list, target_list,
run_metadata)
   1313

/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/client/session.py in _call_tf_sessionrun(self, o
ptions, feed_dict, fetch_list, target_list, run_metadata)
   1419             self._session, options, feed_dict, fetch_list, target_li
st,
-> 1420             status, run_metadata)
```

```
-> 1420                  status, run_metadata)
   1421

/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/framework/errors_impl.py in __exit__(self, type_
arg, value_arg, traceback_arg)
    515             compat.as_text(c_api.TF_Message(self.status.status)),
--> 516             c_api.TF_GetCode(self.status.status))
    517       # Delete the underlying status object from memory otherwise it
stays alive

InvalidArgumentError: Incompatible shapes: [96,1] vs. [128,1]
  [[Node: sub_3 = Sub[T=DT_FLOAT,
_device="/job:localhost/replica:0/task:0/device:CPU:0"]
(discriminator/dense_2/BiasAdd, ones_like_3)]]

During handling of the above exception, another exception occurred:

InvalidArgumentError                      Traceback (most recent call last)
<ipython-input-274-bdbd7efbaf56> in <module>()
      1 with get_session() as sess:
      2     sess.run(tf.global_variables_initializer())
----> 3     run_a_gan(sess, G_train_step, G_loss, D_train_step, D_loss,
G_extra_step, D_extra_step)

<ipython-input-268-65d217674af1> in run_a_gan(sess, G_train_step, G_loss, D
_train_step, D_loss, G_extra_step, D_extra_step, show_every, print_every, b
atch_size, num_epoch)
     25         for (minibatch, minbatch_y) in mnist:
     26             # run a batch of data through the network
---> 27             _, D_loss_curr = sess.run([D_train_step, D_loss], feed_d
ict={x: minibatch})
     28             _, G_loss_curr = sess.run([G_train_step, G_loss])
     29

/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/client/session.py in run(self, fetches, feed_dic
t, options, run_metadata)
    903     try:
    904       result = self._run(None, fetches, feed_dict, options_ptr,
--> 905                          run_metadata_ptr)
    906       if run_metadata:
    907         proto_data = tf_session.TF_GetBuffer(run_metadata_ptr)

/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/client/session.py in _run(self, handle, fetches,
feed_dict, options, run_metadata)
   1138     if final_fetches or final_targets or (handle and feed_dict_tens
r):
   1139       results = self._do_run(handle, final_targets, final_fetches,
-> 1140                              feed_dict_tensor, options,
run_metadata)
   1141     else:
   1142       results = []

/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/client/session.py in _do_run(self, handle, targe
```

```
packages/tensorflow/python/client/session.py in _do_run(self, handle, targe
t_list, fetch_list, feed_dict, options, run_metadata)
   1319     if handle is None:
   1320       return self._do_call(_run_fn, feeds, fetches, targets, option
s,
-> 1321                            run_metadata)
   1322     else:
   1323       return self._do_call(_prun_fn, handle, feeds, fetches)

/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/client/session.py in _do_call(self, fn, *args)
   1338         except KeyError:
   1339           pass
-> 1340       raise type(e)(node_def, op, message)
   1341 
   1342   def _extend_graph(self):

InvalidArgumentError: Incompatible shapes: [96,1] vs. [128,1]
	 [[Node: sub_3 = Sub[T=DT_FLOAT,
_device="/job:localhost/replica:0/task:0/device:CPU:0"]
(discriminator/dense_2/BiasAdd, ones_like_3)]]

Caused by op 'sub_3', defined at:
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/runpy.py", line 193, in _run_module_as_main
    "__main__", mod_spec)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/runpy.py", line 85, in _run_code
    exec(code, run_globals)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/ipykernel_launcher.py", line
16, in <module>
    app.launch_new_instance()
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/traitlets/config/application.py", line 658, in launch_instance
    app.start()
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/ipykernel/kernelapp.py", line
477, in start
    ioloop.IOLoop.instance().start()
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/zmq/eventloop/ioloop.py", lin
e 177, in start
    super(ZMQIOLoop, self).start()
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/tornado/ioloop.py", line 888,
in start
    handler_func(fd_obj, events)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/tornado/stack_context.py", li
ne 277, in null_wrapper
    return fn(*args, **kwargs)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/zmq/eventloop/zmqstream.py",
line 440, in _handle_events
    self._handle_recv()
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/zmq/eventloop/zmqstream.py",
line 472, in _handle_recv
```

```
    self._run_callback(callback, msg)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/zmq/eventloop/zmqstream.py",
line 414, in _run_callback
    callback(*args, **kwargs)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/tornado/stack_context.py", li
ne 277, in null_wrapper
    return fn(*args, **kwargs)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/ipykernel/kernelbase.py", lin
e 283, in dispatcher
    return self.dispatch_shell(stream, msg)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/ipykernel/kernelbase.py", lin
e 235, in dispatch_shell
    handler(stream, idents, msg)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/ipykernel/kernelbase.py", lin
e 399, in execute_request
    user_expressions, allow_stdin)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/ipykernel/ipkernel.py", line
196, in do_execute
    res = shell.run_cell(code, store_history=store_history, silent=silent)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/ipykernel/zmqshell.py", line
533, in run_cell
    return super(ZMQInteractiveShell, self).run_cell(*args, **kwargs)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/IPython/core/interactiveshell.py", line 2717, in run_cell
    interactivity=interactivity, compiler=compiler, result=result)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/IPython/core/interactiveshell.py", line 2821, in run_ast_nodes
    if self.run_code(code, result):
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/IPython/core/interactiveshell.py", line 2881, in run_code
    exec(code_obj, self.user_global_ns, self.user_ns)
  File "<ipython-input-273-0b8053754f81>", line 1, in <module>
    D_loss, G_loss = lsgan_loss(logits_real, logits_fake)
  File "<ipython-input-271-be47cf6d18a2>", line 19, in lsgan_loss
    fake_image_loss = tf.reduce_mean((scores_real - true_labels)**2)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/ops/math_ops.py", line 971, in binary_op_wrapper
    return func(x, y, name=name)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/ops/gen_math_ops.py", line 7933, in sub
    "Sub", x=x, y=y, name=name)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/framework/op_def_library.py", line 787, in _appl
y_op_helper
    op_def=op_def)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
```

```
packages/tensorflow/python/framework/ops.py", line 3290, in create_op
    op_def=op_def)
  File "/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-
packages/tensorflow/python/framework/ops.py", line 1654, in __init__
    self._traceback = self._graph._extract_stack()  # pylint:
disable=protected-access

InvalidArgumentError (see above for traceback): Incompatible shapes: [96,1]
vs. [128,1]
  [[Node: sub_3 = Sub[T=DT_FLOAT,
_device="/job:localhost/replica:0/task:0/device:CPU:0"]
(discriminator/dense_2/BiasAdd, ones_like_3)]]
```

Deep Convolutional GANs

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 as our discriminators and generators.

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. *Be sure to check the dimensions of x and reshape when needed*, fully connected blocks expect [N,D] Tensors while conv2d blocks expect [N,H,W,C] Tensors. Please use `tf.layers` to define the following architecture:

Architecture:

- Conv2D: 32 Filters, 5x5, Stride 1, padding 0
- Leaky ReLU(alpha=0.01)
- Max Pool 2x2, Stride 2
- Conv2D: 64 Filters, 5x5, Stride 1, padding 0
- 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

Once again, please use biases for all convolutional and fully connected layers, and use the default parameter initializers. Note that a padding of 0 can be accomplished with the 'VALID' padding option.

In [275]:

```
def discriminator(x):
    """Compute discriminator score for a batch of input images.

    Inputs:
    - x: TensorFlow Tensor of flattened input images, shape [batch_size, 78
4]
```

```
    Returns:
    TensorFlow Tensor with shape [batch_size, 1], containing the score
    for an image being real for each input image.
    """
    with tf.variable_scope("discriminator"):
        # TODO: implement architecture
        #pass
        x1 = tf.reshape(x, shape = [-1, 28, 28, 1])
        conv1 = tf.layers.conv2d(inputs = x1, kernel_size = 5, strides = 1, 
filters = 32 ,activation = leaky_relu)
        maxpool1 = tf.layers.max_pooling2d(inputs = conv1, pool_size = 2, st
rides = 2)
        conv2 = tf.layers.conv2d(inputs = maxpool1, kernel_size = 5, strides
 = 1, filters = 64, activation = leaky_relu)
        maxpool2 = tf.layers.max_pooling2d(inputs=conv2, pool_size=2, stride
s=2)
        flatten = tf.reshape(maxpool2, shape = [-1, 1024])
        fc1 = tf.layers.dense(inputs = flatten, units = 1024, activation = l
eaky_relu)
        logits = tf.layers.dense(inputs = fc1, units = 1)


        return logits
test_discriminator(1102721)
```

```
Correct number of parameters in discriminator.
```

Generator

For the generator, we will copy the architecture exactly from the InfoGAN paper. See Appendix C.1 MNIST. Please use `tf.layers` for your implementation. You might find the documentation for tf.layers.conv2d_transpose useful. The architecture is as follows.

Architecture:

- Fully connected with output size 1024
- `ReLU`
- BatchNorm
- Fully connected with output size 7 x 7 x 128
- `ReLU`
- BatchNorm
- Resize into Image Tensor of size 7, 7, 128
- Conv2D^T (transpose): 64 filters of 4x4, stride 2
- `ReLU`
- BatchNorm
- Conv2d^T (transpose): 1 filter of 4x4, stride 2
- `TanH`

Once again, use biases for the fully connected and transpose convolutional layers. Please use the default initializers for your parameters. For padding, choose the 'same' option for transpose convolutions. For Batch Normalization, assume we are always in 'training' mode.

In [276]:

```
def generator(z):
```

```
    """Generate images from a random noise vector.

    Inputs:
    - z: TensorFlow Tensor of random noise with shape [batch_size, noise_di
m]

    Returns:
    TensorFlow Tensor of generated images, with shape [batch_size, 784].
    """
    with tf.variable_scope("generator"):
        # TODO: implement architecture
        #pass
        # Dense layer #1
        fc1 = tf.layers.dense(inputs = z, units = 1024, activation = tf.nn.r
elu)
        fc1_bn = tf.layers.batch_normalization(inputs = fc1, training = True
)
        # Dense layer #2
        fc2 = tf.layers.dense(inputs = fc1_bn, units = 7*7*128, activation =
tf.nn.relu)
        fc2_bn = tf.layers.batch_normalization(inputs = fc2, training = True
)

        reshaped_tensor = tf.reshape(fc2_bn, shape = [-1, 7, 7, 128])
        # Transposed convolutional layer #1:
        #trans_conv1 =  tf.nn.conv2d_transpose(value = reshaped_tensor, fil
ters = 64, kernel_size = 4, strides = 2,padding = 'SAME', activation =
tf.nn.relu)
        #trans_conv1 =  tf.nn.conv2d_transpose(value = reshaped_tensor, fil
ters = 64, strides = 2, padding = 'SAME', activation = tf.nn.relu)

        conv_t1 = tf.layers.conv2d_transpose(inputs = reshaped_tensor, filte
rs = 64, kernel_size = 4, strides = 2, padding = 'SAME', activation = tf.nn.
relu)
        conv_t1_bn = tf.layers.batch_normalization(inputs = conv_t1, trainin
g = True)
        # Transposed convolutional layer #2:
        conv_t2 = tf.layers.conv2d_transpose(inputs = conv_t1_bn, filters =
1, kernel_size = 4, strides = 2, padding = 'SAME', activation = tf.nn.tanh)
        img = tf.reshape(conv_t2, shape=[-1, 784])
        return img
test_generator(6595521)
```

```
Correct number of parameters in generator.
```

We have to recreate our network since we've changed our functions.

In [277]:

```
tf.reset_default_graph()

batch_size = 128
# our noise dimension
noise_dim = 96

# placeholders for images from the training dataset
x = tf.placeholder(tf.float32, [None, 784])
```

```
x = tf.placeholder(tf.float32, [None, 784])
z = sample_noise(batch_size, noise_dim)
# generated images
G_sample = generator(z)

with tf.variable_scope("") as scope:
    #scale images to be -1 to 1
    logits_real = discriminator(preprocess_img(x))
    # Re-use discriminator weights on new inputs
    scope.reuse_variables()
    logits_fake = discriminator(G_sample)

# Get the list of variables for the discriminator and generator
D_vars = tf.get_collection(tf.GraphKeys.TRAINABLE_VARIABLES,'discriminator'
)
G_vars = tf.get_collection(tf.GraphKeys.TRAINABLE_VARIABLES,'generator')

D_solver,G_solver = get_solvers()
D_loss, G_loss = gan_loss(logits_real, logits_fake)
D_train_step = D_solver.minimize(D_loss, var_list=D_vars)
G_train_step = G_solver.minimize(G_loss, var_list=G_vars)
D_extra_step = tf.get_collection(tf.GraphKeys.UPDATE_OPS,'discriminator')
G_extra_step = tf.get_collection(tf.GraphKeys.UPDATE_OPS,'generator')
```

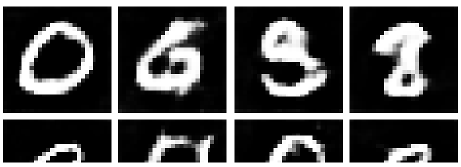
Train and evaluate a DCGAN

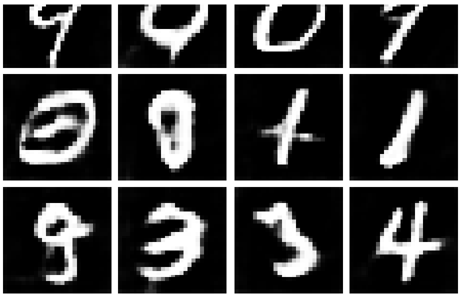
This is the one part of A3 that significantly benefits from using a GPU. It takes 3 minutes on a GPU for the requested five epochs. Or about 50 minutes on a dual core laptop on CPU (feel free to use 3 epochs if you do it on CPU).

In [280]:

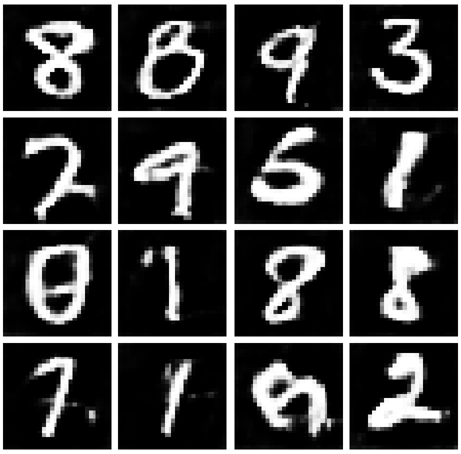
```
with get_session() as sess:
    sess.run(tf.global_variables_initializer())
    run_a_gan(sess,G_train_step,G_loss,D_train_step,D_loss,G_extra_step,D_e
xtra_step,num_epoch=3)
```

```
Epoch: 0, D: 1.017, G:1.476
Epoch: 1, D: 1.134, G:0.8794
```





```
Epoch: 2, D: 1.193, G:0.6598
Final images
```



In [ ]:

```

```

INLINE QUESTION 1

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) with step size 1. 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.

Your answer:

y_0	**y_1**	**y_2**	**y_3**	**y_4**	**y_5**	**y_6**
1						
x_0	x_1	x_2	x_3	x_4	x_5	x_6
1						

INLINE QUESTION 2

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

Your answer:

INLINE QUESTION 3

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

Your answer:

That is not a good sign. Ideally both G and D loss should decrease and converge to a similar value. We want to maintain some degree of balance between a generator and a discriminator such that both parties should win roughly half of the time (eg Nash equilibrium). the fact that G loss is decreasing is expected. The fact that D loss stays constant implies that the discriminator is not learning anything. we should make the discriminator more powerful.