HOMEWORK 5

Hafeez Ali Anees Ali
908 460 3423

Instructions:

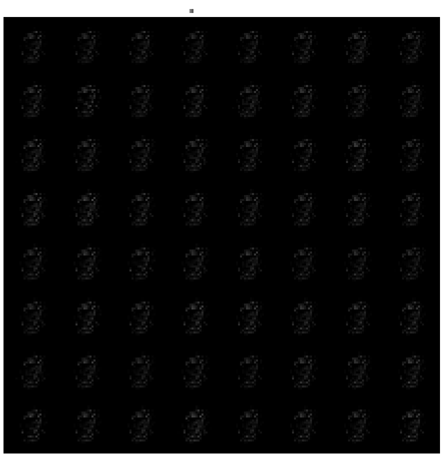
- Please submit your answers in a single pdf file and your code in a zip file. pdf preferably made using latex. No need to submit latex code.
- Submit code for programming exercises. Though we provide a base code with python (jupyter notebook, you can import it in colab and choose GPU as the runtime environment), you can use any programming language you like as long as you use the same model and dataset.

1 Implementation: GAN (55 pts)

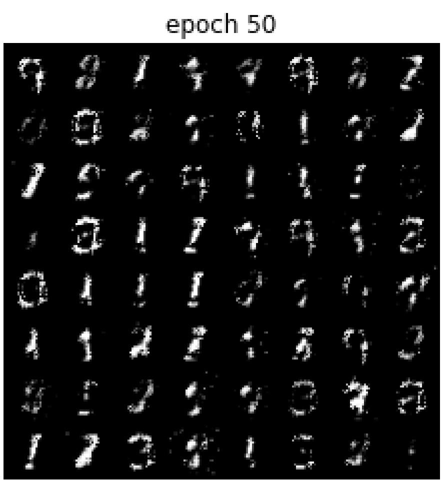
In this part, you are expected to implement GAN with MNIST dataset. We have provided a base jupyter notebook (gan-base.ipynb) for you to start with, which provides a model setup and training configurations to train GAN with MNIST dataset.

(a) Implement training loop and report learning curves and generated images in epochs 1, 50, and 100. Note that drawing learning curves and visualization of images are already implemented in the provided jupyter notebook. (20 pts)

Epoch 1



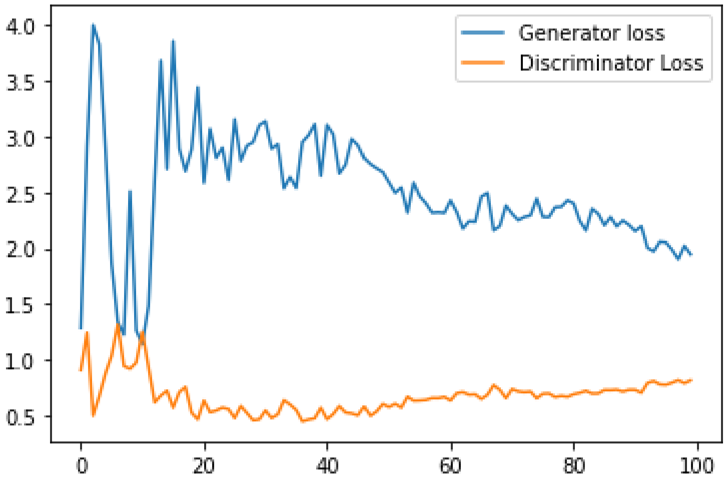
Epoch 50



Epoch 100

epoch 100

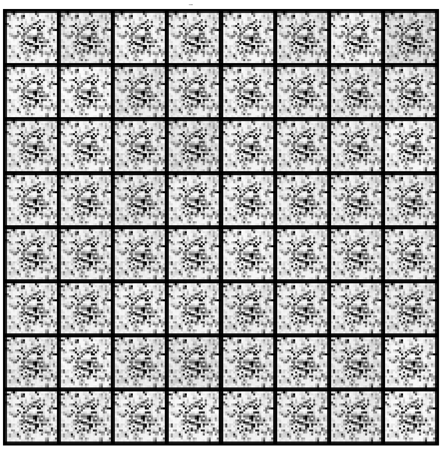
Learning Curve



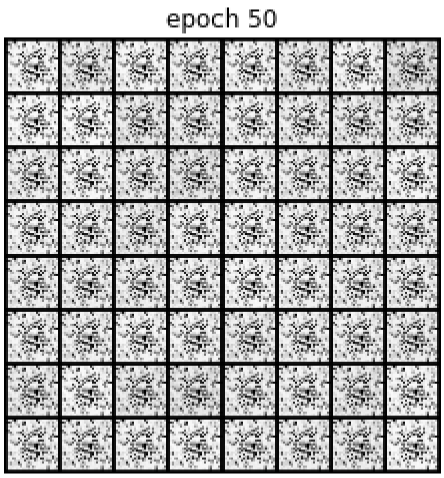
(b) Replace the generator update rule as the original one in the slide,
"Update the generator by descending its stochastic gradient:"

and report learning curves and generated images in epochs 1, 50, and 100. Compare the result with (a). Note that it may not work. If training does not work, explain why it does not work. (10 pts)

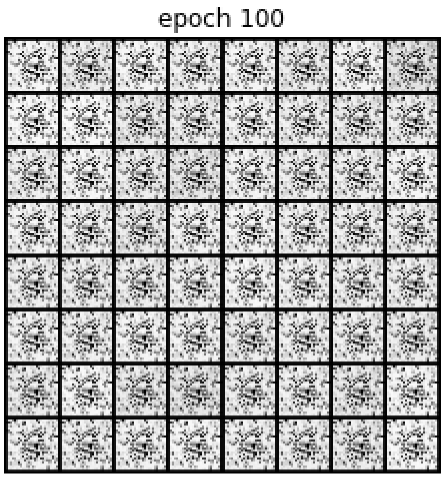
Epoch 1



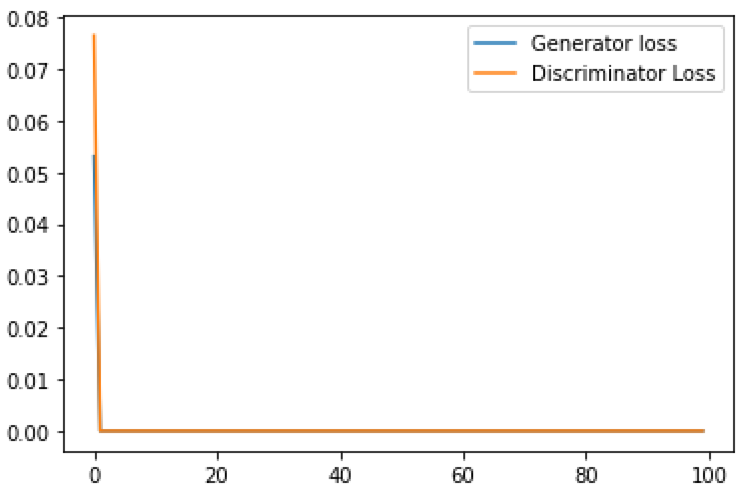
Epoch 50



Epoch 100



Learning Curve



Comparing the images with the results of 1.a, we see that the new update rule does not work.

This is because of the gradient saturation problem of the generator loss function. This happens because early in the learning process the discriminator can reject the generated samples with high probability leading to the saturation of the log(1-D(G(z)))

(c) Except for the method that we used in (a), how can we improve training for GAN? Implement that and report learning curves and generated images in epochs 1, 50, and 100. (10 pts)

Modifications to the hyperparameters and label smoothing are used as potential approaches to improve training for GAN.

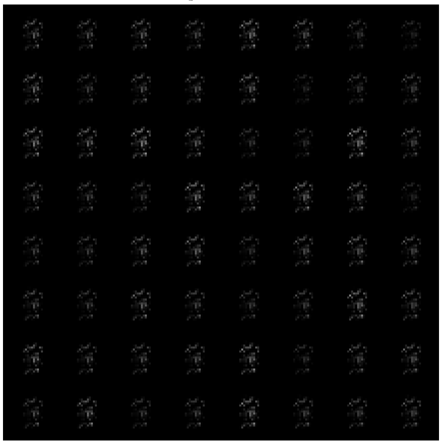
Training discriminator more times than the generator: training the generator depends on how well the discriminator is able to identify fake images from real images. So, in some cases the training of GAN can be improved by training the discriminator more than the generator.

Label smoothing: In GAN, if the discriminator starts to depend on only a small set of features to detect real images, the generator will produce those features only. To avoid this problem the discriminator is penalized when the prediction for any real image goes beyond 0.9

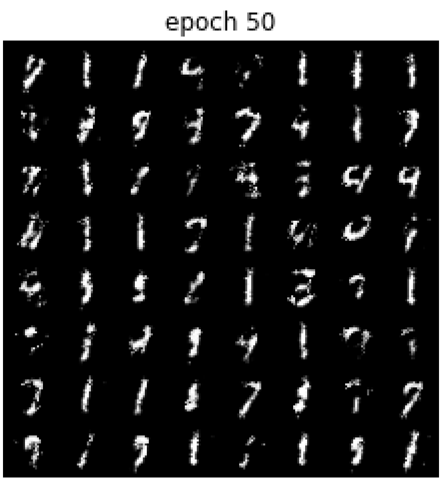
The results are shown below in the form of images generated after epoch 1, 50 and 100. The learning curves are also shown

1. Increasing number of steps for the discriminator in the first epoch, starting with 32 steps in the first epoch and then 1 step for every subsequent epoch

Epoch 1

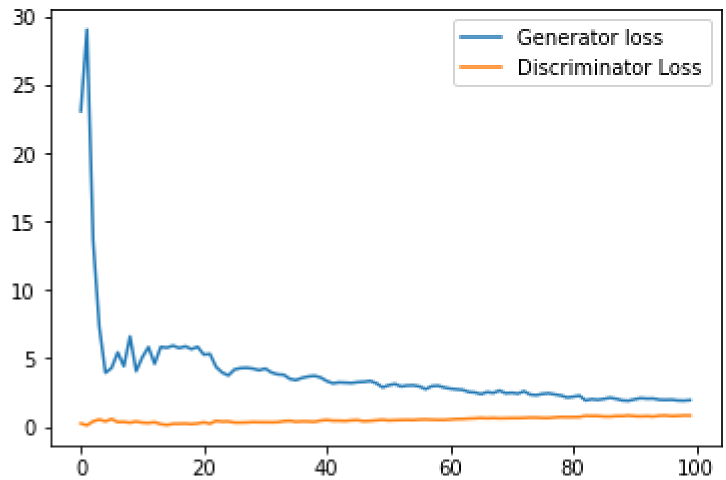


Epoch 50



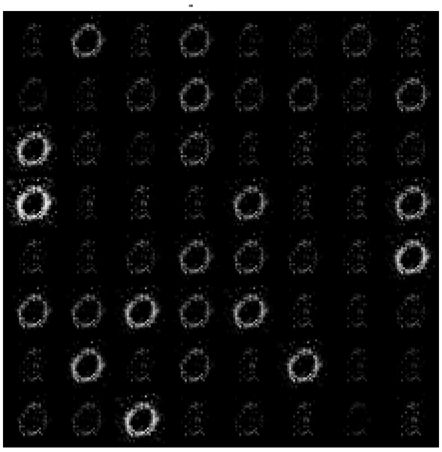
Epoch 100

Learning Curve

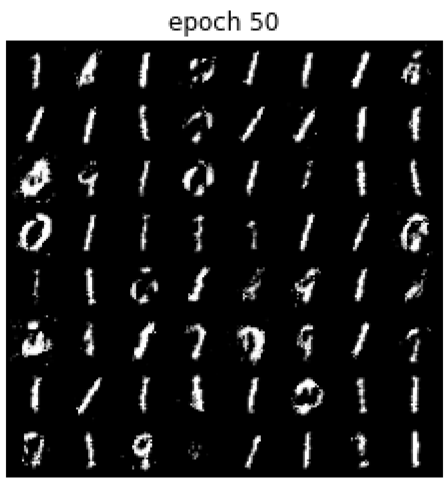


2. Increasing number of steps for the discriminator in the first few epochs, starting with 16 for the first epoch, 8, 4 and 2 for the subsequent 1-10, 11-20 and 21-30 epochs respectively.

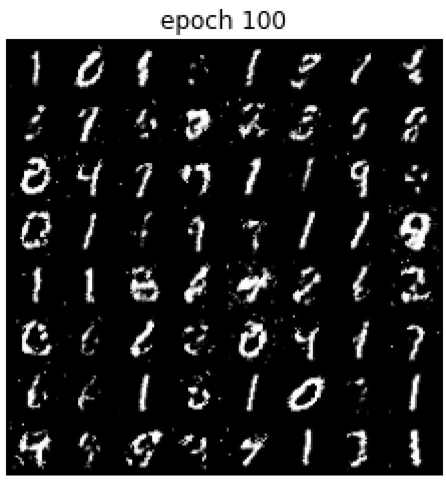
Epoch 1



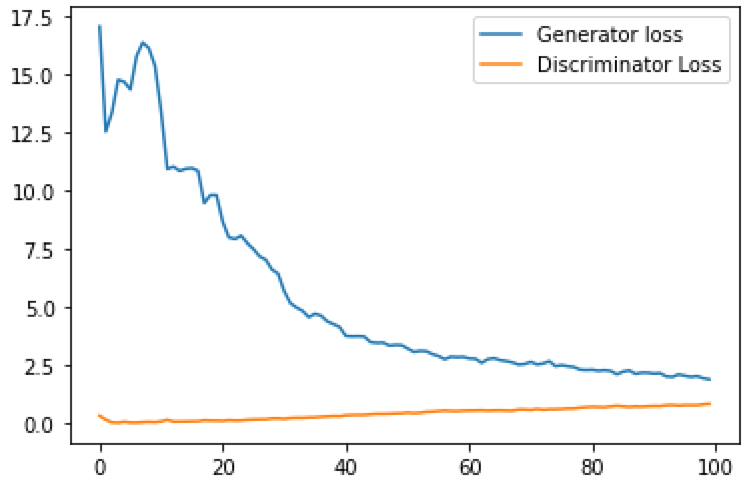
Epoch 50



Epoch 100

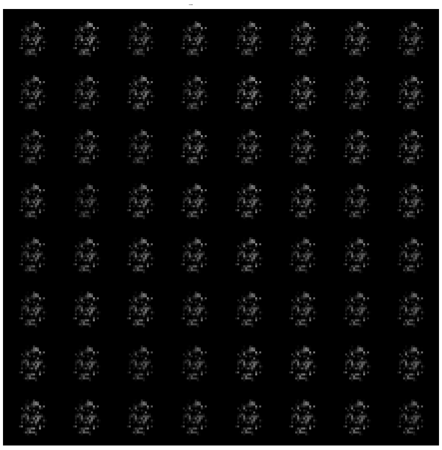


Learning Curve

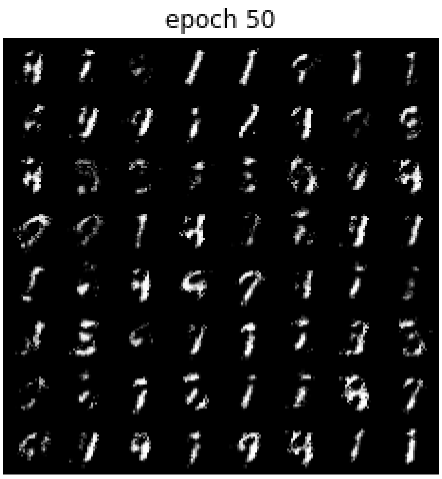


3. Increasing number of steps for the discriminator in the second 50 epochs, alternating between 1 and 2 steps.

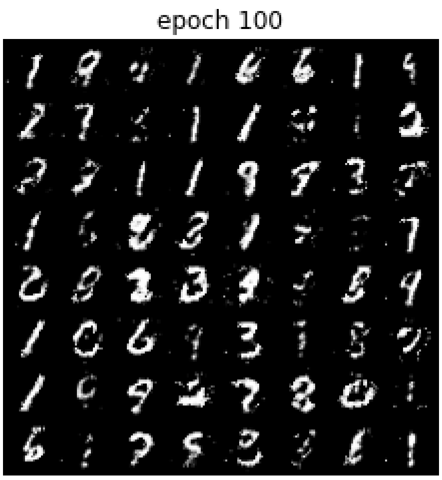
Epoch 1



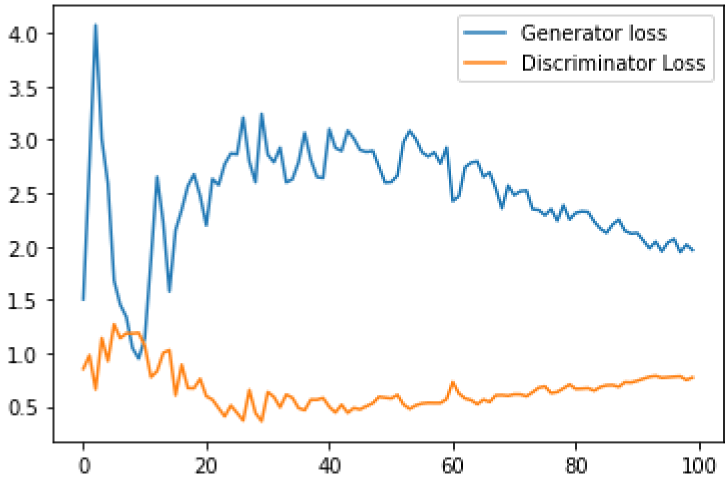
Epoch 50



Epoch 100

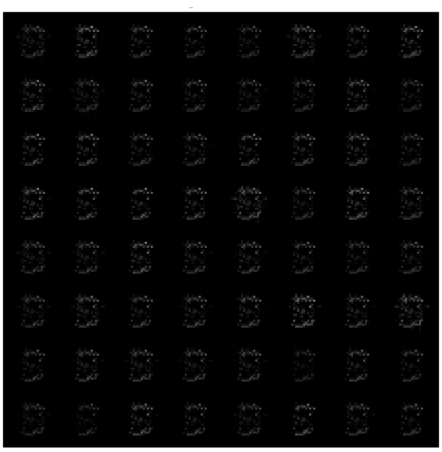


Learning Curve

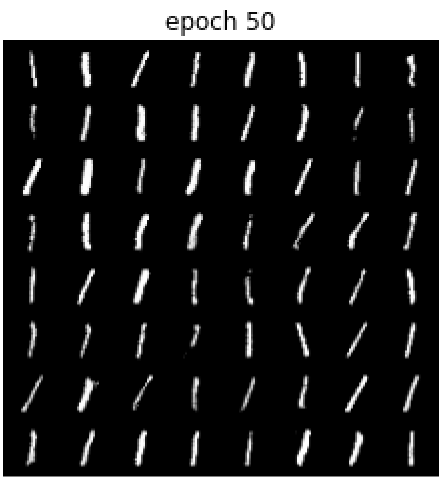


4. Keeping a consistent number of 4 steps for the discriminator for all epochs: Here, it looks like the discriminator is has learnt only 1 to be ‘real’ thereby the generator started generating only 1s

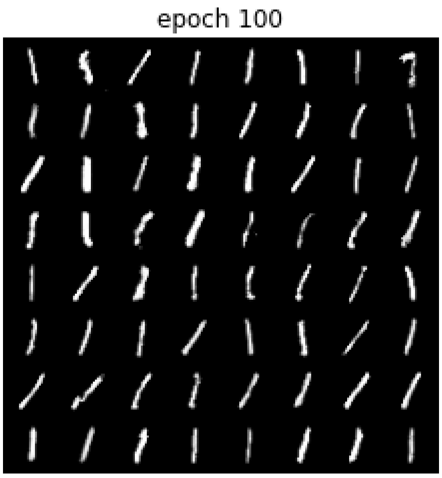
Epoch 1



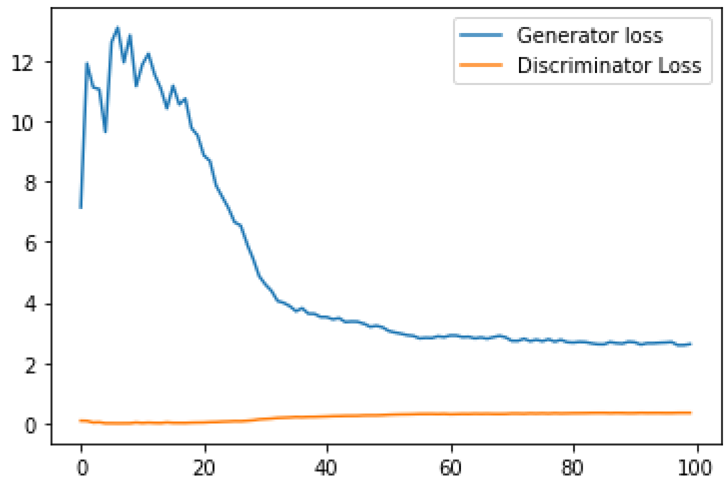
Epoch 50



Epoch 100

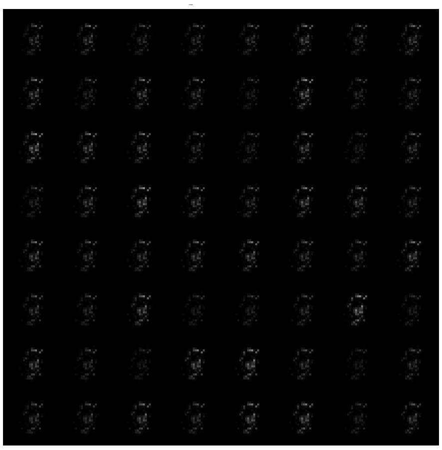


Learning Curve

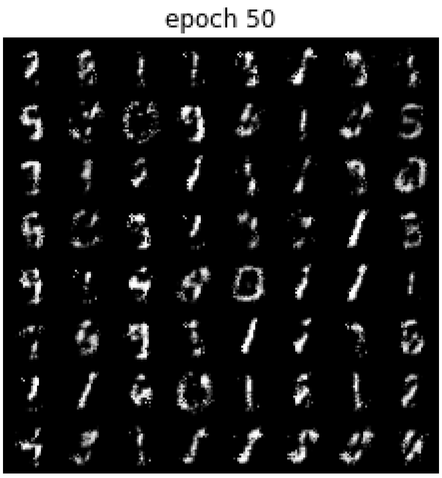


5. Implementing label smoothing for both real and fake images for the generator and the discriminator

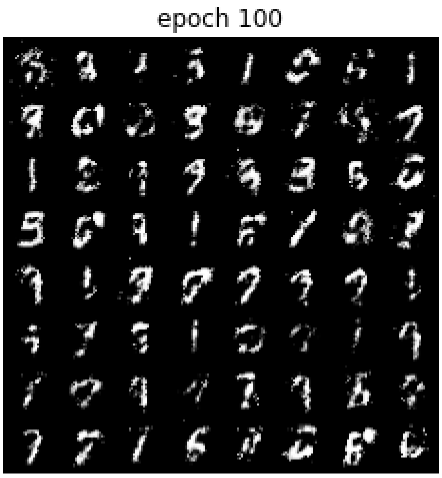
Epoch 1



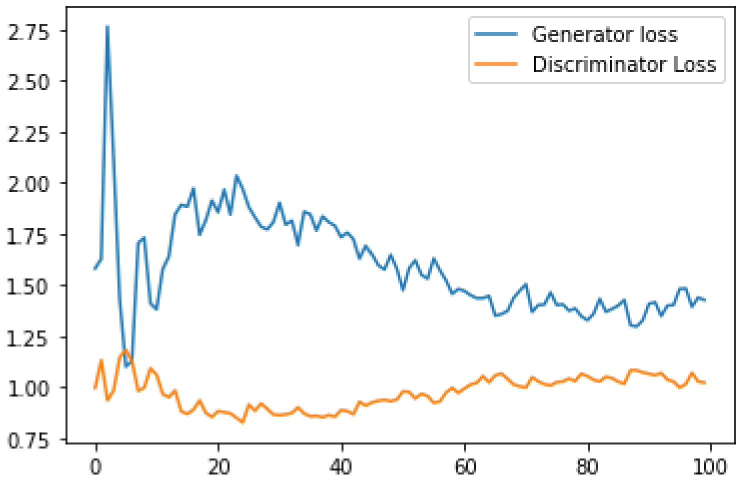
Epoch 50



Epoch 100

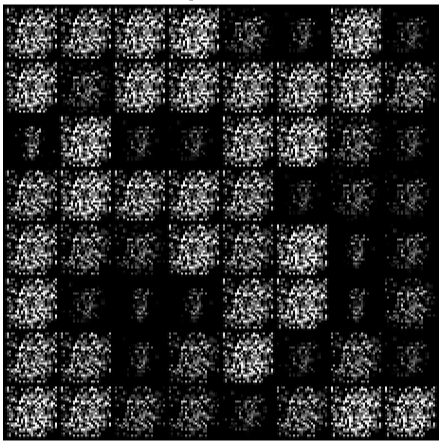


Learning Curve

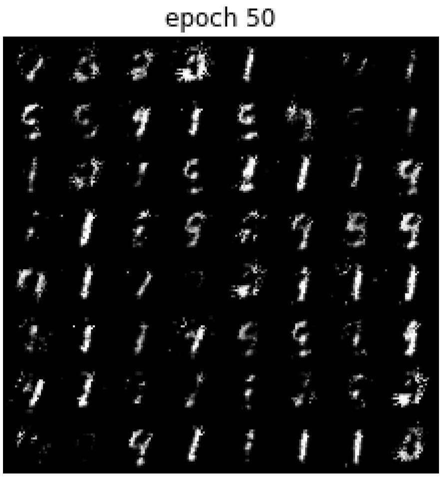


6. Implementing one-side smoothing for real images for the discriminator only

Epoch 1

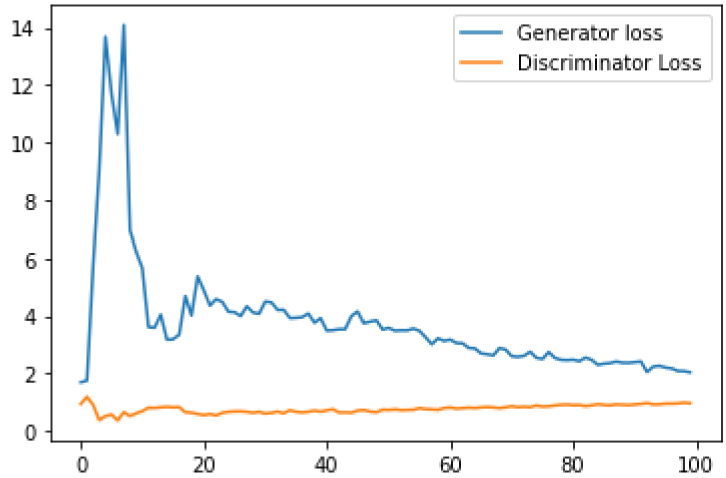


Epoch 50



Epoch 100

Learning Curve


Generator loss
Discriminator Loss
0
2
4
6
8
10
12
14
0
20
40
60
80
100

2 Ridge regression [20 pts]

Derive the closed-form solution in matrix form for the ridge regression problem:

This A matrix has the effect of NOT regularizing the bias β_0, which is standard practice in ridge regression. Note: Derive the closed-form solution, do not blindly copy lecture notes.

2] Ridge Regression:

$$\min_{\beta} \left(\frac{1}{n} \sum_{i=1}^{n} (z_i^T \beta - y_i)^2 \right) + \lambda \|\beta\|_A^2$$

$$\text{where } \|\beta\|_A^2 := \beta^T A \beta$$

$$\text{and } A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Taking the derivative & setting to 0, we have:

$$\nabla_\beta \left[\frac{1}{n} \sum_{i=1}^{n} (z_i^T \beta - y_i)^2 + \lambda \|\beta\|_A^2 \right] = 0$$

Representing this in matrix form, we have:

$$\nabla_\beta \left[\frac{1}{n} \|Z\beta - Y\|_2^2 + \lambda \|\beta\|_A^2 \right] = 0$$

$$\nabla_\beta \left[\frac{1}{n} \left[(Z\beta - Y)^T (Z\beta - Y) \right] + \lambda \beta^T A \beta \right] = 0$$

$$\nabla_\beta \left[\frac{(Z\beta - Y)^T (Z\beta - Y)}{n} + \lambda \beta^T A \beta \right] = 0$$

$$\nabla_\beta \left[\frac{\beta^T Z^T Z \beta - 2\beta^T Z^T Y + Y^T Y}{n} + \lambda \beta^T A \beta \right] = 0$$

$$\frac{2Z^T Z \beta - 2Z^T Y}{n} + 2\lambda A \beta = 0$$

$$\frac{Z^T Z \beta}{n} - \frac{Z^T Y}{n} + \lambda A \beta = 0$$

$$\beta \left(\frac{Z^T Z}{n} + \lambda A \right) = \frac{Z^T Y}{n} \Rightarrow \beta (Z^T Z + n\lambda A) = Z^T Y$$

$$\Rightarrow \boxed{\beta = (Z^T Z + n\lambda A)^{-1} Z^T Y}$$

3 Review the change of variable in probability density function [25 pts]

In Flow based generative model, we have seen $p_\theta(x) = p(f_\theta(x))|\frac{\partial f_\theta(x)}{\partial x}|$.

As a hands-on (fixed parameter) example, consider the following setting.

Let X and Y be independent, standard normal random variables. Consider the transformation $U = X + Y$ and $V = X - Y$. In the notation used above, $U = g_1(X, Y)$ where $g_1(X, Y)$ where $g_1(x, y) = x + y$ and $V = g_2(X, Y)$ where $g_2(x, y) = x - y$. The joint pdf of X and Y is $f_{X,Y} = (2\pi)^{-1} exp(-x^2/2)exp(-y^2/2)$,

(a) Compute Jacobian matrix (5 pts)

3] given: X & Y are independent, standard normal random variables

$U = X + Y = g_1(X, Y)$ $\quad$ $g_1(x, y) = x + y$

$V = X - Y = g_2(X, Y)$ $\quad$ $g_2(x, y) = x - y$

$f_{X,Y} = \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2}$, $\quad -\infty < x < \infty$

$-\infty < y < \infty$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

a) Jacobian matrix

$$J = \begin{bmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{bmatrix}$$

$u = x + y$

$v = x - y$

$2x = u + v$

$\boxed{x = \frac{u+v}{2}}$

$\boxed{y = \frac{u-v}{2}}$

$\partial x/\partial u = 1/2$

$\partial x/\partial v = 1/2$

$\partial y/\partial u = 1/2$

$\partial y/\partial v = -1/2$

$$\boxed{J = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}}$$

(b) (Forward) Show the joint pdf of U,V

b) For the rectangle from (u, v) to $(u+\Delta u, v+\Delta v)$, we have the joint density function $f_{U,V}(u,v)$ and probability $f_{U,V}(u,v)\,\Delta u\,\Delta v$

Since $(U,V) = g(X,Y)$,
Having $(X,Y) = g^{-1}(U,V)$, then this probability is equal to the area of image of the rectangle from (u,v) to $(u+\Delta u, v+\Delta v)$ under the map g^{-1} times the density $f_{X,Y}(x,y)$

The linear approximations for g^{-1} give, in vector form, the two sides in the parallelogram that approximates the image of the rectangle.

$$g^{-1}(u+\Delta u, v) \approx g^{-1}(u,v) + \frac{\partial}{\partial u} g^{-1}(u,v)\Delta u = g^{-1}(u,v) + \left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}\right)\Delta u$$

and $$g^{-1}(u, v+\Delta v) \approx g^{-1}(u,v) + \frac{\partial}{\partial v} g^{-1}(u,v)\Delta v = g^{-1}(u,v) + \left(\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}\right)\Delta v$$

The area of the rectangle is given by the norm of cross product $\left|\left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}\right) \times \left(\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}\right)\right| \Delta u\,\Delta v$

The determinant of the Jacobian matrix is

$$\det(J) = \det\begin{pmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{pmatrix}$$

Thus, $f_{U,V}(u,v)\,\Delta u\,\Delta v \approx f_{X,Y}(g^{-1}(u,v))\,|\det(J)|\,\Delta u\,\Delta v$

$$\Rightarrow f_{U,V}(u,v) = f_{X,Y}(g^{-1}(u,v))\,|\det(J)|$$

$$f_{U,V}(u,v) = \frac{1}{2} f_{X,Y}\left(g^{-1}(u,v)\right)$$

$$= \frac{1}{2}\left[\frac{1}{2\pi} e^{-\frac{\left(\frac{u+v}{2}\right)^2}{2}} e^{-\frac{\left(\frac{u-v}{2}\right)^2}{2}}\right]$$

$$= \frac{1}{2\times 2\pi}\left[e^{-\left(\frac{u^2+v^2+2uv+u^2+v^2-2uv}{2^2\times 2}\right)}\right]$$

$$= \frac{1}{2\times 2\pi} e^{-\frac{(u^2+v^2)}{4}}$$

$$\boxed{f_{U,V}(u,v) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-u^2/4} \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-v^2/4}}$$

$$|\det(J)| = 1/2$$

$$J = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$$

$$\det(J) = -1/2$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{-1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}$$

$$x = \frac{u+v}{2} \quad y = \frac{u-v}{2}$$

(c) (Inverse) Check whether the inverse holds

c) given $f_{U,V}(u,v) = f_{X,Y}\left(g^{-1}(u,v)\right) |\det(J)|$

$$f_{U,V}(u,v) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-u^2/4} \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-v^2/4}$$

$$\Rightarrow f_{U,V}(x+y, x-y) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{(x+y)^2}{4}} \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{(x-y)^2}{4}}$$

$$= \frac{1}{2\times 2\pi} e^{-\left(\frac{x^2+y^2+2xy+x^2+y^2-2xy}{4}\right)}$$

$$= \frac{1}{2\times 2\pi} e^{-\frac{(x^2+y^2)}{2}}$$

$$f_{U,V}(x+y, x-y) = \frac{1}{2} \times \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2}$$

Since

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

$u = x + y$

$v = x - y$

$$f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2} \quad \text{— ①}$$

$$f_{U,V}(x+y, x-y)\,|\det(J)|^{-1} = \frac{1}{\cancel{2}} \times \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2} \times \cancel{2}$$

$$= \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2} \quad \text{— ②}$$

From ① & ②,

$$\boxed{f_{X,Y}(x,y) = f_{U,V}(x+y, x-y)\,|\det(J)|^{-1}}$$