Problems of GANs

① Vanishing Gradient : Mathematematical Reasoning

G_θ :

$\hookrightarrow (w, b)$: derivative with respect to the weight and biases

$\frac{\partial}{\partial w}, \frac{\partial}{\partial b} \longrightarrow 0$ at the trainig process.

at the beginning the generator is lousy

$\hookrightarrow$ gradient at the back propagation becomes 0.

$$L = \min_G \max_D \left\{ \mathbb{E}_{x \sim P_{data}(x)} \left[\log D_\phi(x) \right] + \mathbb{E}_{z \sim P_g(z)} \left[\log \left(1 - D(G_\theta(z)) \right) \right] \right.$$

generator update step :

$$\nabla_\theta L = \nabla_\theta \left[\mathbb{E}_{z \sim P_z(z)} \log \left(1 - D_\phi(G_\theta(z)) \right) \right]$$

$$= \mathbb{E}_{z \sim P_z(z)} \nabla_\theta \left[\log \left(1 - D_\phi(G_\theta(z)) \right) \right]$$

$$= \mathbb{E}_{z \sim P_z(z)} \frac{1}{\left(1 - D_\phi(G_\theta(z))\right)} \cdot \frac{(-)\partial D_\phi(G_\theta(z))}{\partial G_\theta(z)} \cdot \frac{\partial G_\theta(z)}{\partial \theta}$$

$$= \mathbb{E}_{z \sim P_z(z)} \frac{\partial G_\theta(z)}{\partial \theta} \frac{\partial D_\phi G_\theta(z)}{\partial G_\theta(z)} \frac{1}{\left(D_\phi(G_\theta(z)) - 1\right)}$$

Here we are taking $z \sim P_z(z)$: random distribution

But we can do the transformation as

$x = G(z)$ to transform the distribution

from $P_g(z)$ to $P_g(x)$.

Therefore considering $x = G_\theta(z)$ we can move forward

as
$$\nabla_\theta L = \mathbb{E}_{x \sim P_g(x)} \frac{\delta x}{\delta \theta} \cdot \frac{\delta D_\phi(x)}{\delta x} \cdot \frac{1}{D_\phi(x) - 1}$$

generator distribution.

From the previous page

$$\mathbb{E}_{x\sim P_g(x)}\left\{\frac{1}{(D(x)-1)}\cdot\frac{\delta D(x)}{\delta x}\cdot\frac{\delta x}{\delta \theta}\right\}$$

where $\nabla_\theta = \frac{\delta}{\delta\theta}$, $\nabla_x = \frac{\delta}{\delta x}$.

$$= \mathbb{E}_{x\sim P_g(x)}\left\{\frac{1}{(D(x)-1)}\cdot\nabla_x D(x)\cdot\nabla_\theta x\right\}$$

$$= \mathbb{E}_{x\sim P_g(x)}\ \underline{\frac{\nabla_x D(x)\cdot\nabla_\theta x}{(D_\theta(x)-1)}}$$

(X) we are now training generator after the discriminator.

we consider that the discriminator is trained very well so that it can detect fake and real images.

consider $\tilde{D}(x)$: well trained discriminator. Then the output will be $\tilde{D}(x) = 1$ if $x \sim P_{data}(x)$

and $\tilde{D}(x) = 0$ if $x \sim P_g(x)$.

in this case when $D \to \tilde{D}$ [move towards the perfect discriminator then $x \sim P_g$]

then it can be expressed as

$$\boxed{\lim_{D\to\tilde{D}} D(x) = 0} \longrightarrow (A)$$

and for the derivative $\boxed{\lim_{D\to\tilde{D}} \nabla_x D(x) = 0} \longrightarrow (B)$

now we apply the limit in (B) in the equation of (X) then we get it as

$$\nabla_\theta L = \lim_{D\to\tilde{D}} \mathbb{E}_{x\sim P_g(x)}\left\{\frac{\overbrace{\boxed{\nabla_x D(x)}}^{=0}\ \nabla_\theta x}{(D(x)-1)}\right\} = 0$$

Because $\lim_{D\to\tilde{D}} \nabla_x D(x) = 0$

For the same reason also a moderately trained discriminator at the very beginning of the training of the generator will act as a perfect discriminator and as a result it will make like (A) which is $\lim_{D\to\tilde{D}} D(x) = 0$

as it will take all of the samples from x to $P_g(x)$.

Solution to Vanishing Gradient Problem:

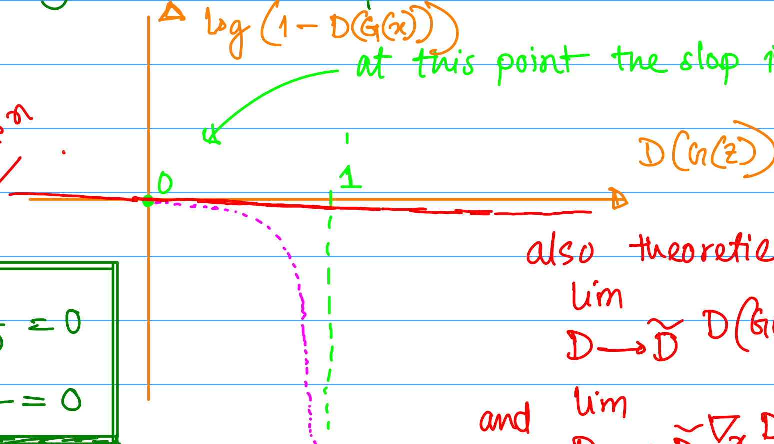
Here the original loss function of the generator that is being targeted is

$$\min_{G} \mathbb{E}_{z \sim p_g(z)} \left[\log\left(1 - D_{\cancel{\theta}}(G_{\cancel{\theta}}(x))\right) \right]$$

Here the training of the generator will be done after the training of the discriminator is done. So at this point the discriminator is good but the generator is lousy.

So it will immediately produce $D(G(z)) = 0$ → Lousy.

So the plot of the above equation will be



So at this situation it is difficult to train the function at this point

$$\boxed{\begin{array}{l} \dfrac{\partial}{\partial w} = 0 \\ \dfrac{\partial}{\partial b} = 0 \end{array}}$$

also theoretically

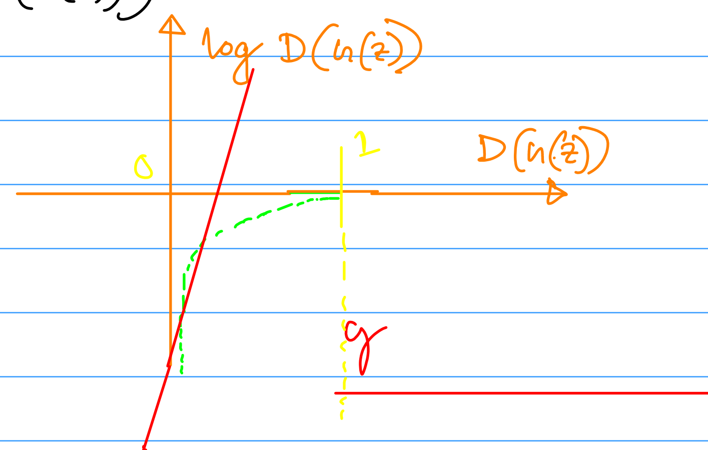
$$\lim_{D \to \tilde{D}} D(G(x)) = 0$$

and

$$\lim_{D \to \tilde{D}} \nabla_x D(G(z)) = 0$$

Thus we change the optimization function into a maximization.

$$\max_{G} \mathbb{E}_{z \sim p_g(z)} \log\left(D(G(z))\right)$$



initial high slope good for training.

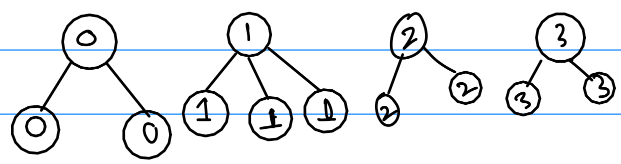
Problem 2: Mode Collapse

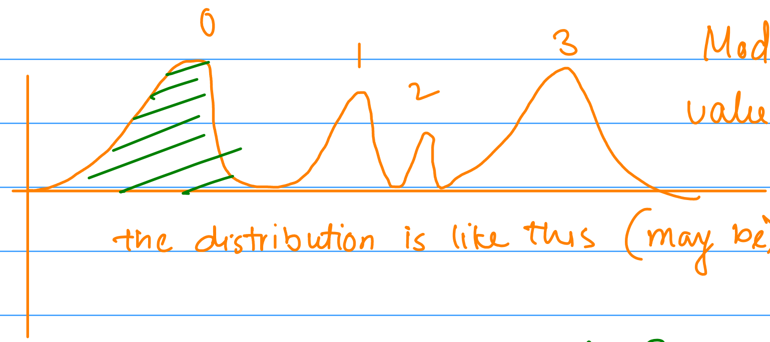
During the training process, the generator sometimes collapse to a setting where it always produces the same output. This is called mode collapse.

Reasons of mode collapse:

- Most of the high level distribution of the real world are highly complex and they are multimodal i.e. their distribution has multiple modes / peaks in the probability distribution.

for example in the MNIST dataset





Mode collapse will produce output values from either of the two modes.

the distribution is like this (may be)

why it is not learning all modes?

- aim, as long the generator is able to fool the discriminator.

- when the discriminator gets better on distinguishing one mode, the generator sometimes concentrate more on that mode instead of producing / sampling data from the other modes.

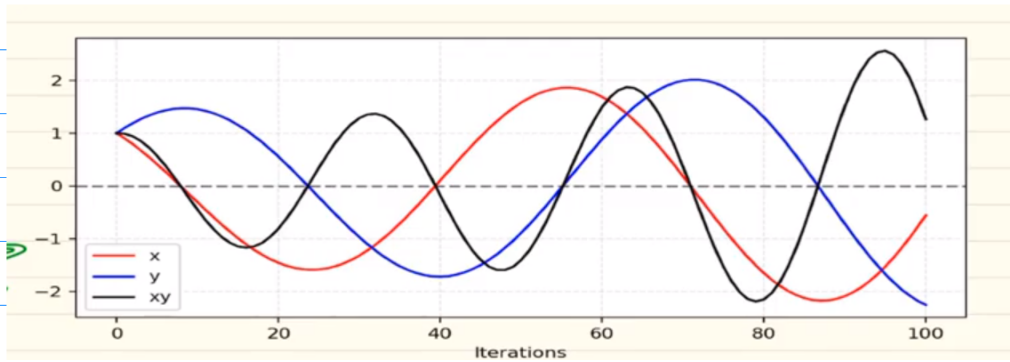
Problem: Hard to achieve Nash Equilibrium
(Convergence of both G_θ and D_ϕ is not guaranted.).

$\boxed{\text{GAN:}}$ a non cooperative two player game.

generator and discriminator are not coooerating each other

For example let us just look at the following functions while f_1 is trying to maximize and f_2 is trying to minimize.

Functions	Derivative	parameter update learning state
$f_1(x) = xy$	$\frac{\delta}{\delta x} f_1(x) = y$	$x' : x - \eta y$
$f_2(y) = -xy$	$\frac{\delta}{\delta y} f_2(y) = -x$	$y' : y + \eta x$



So here as more and moro iteratin goes by the relation and combination of variables x and y becomes more and more complex.

Therefore in this process it is very difficult to obtain the Nash Equilibrium.

Problem : Problem with counting

5 Problem : Problem with Perspective

- Front end object and back end object
-

6 Problem: Problem in understanding the global structure.