Deep Learning 2024011303 Liu Hanzuo

HW4

True or False

P1 False; No, we are expecting a lower FID score.

P2 False; We need to update the generator and discriminator step by step, one by one.

QA

P3 1.

$$JSD(p\|q) = \frac{1}{2}\left(KL(p||\frac{p+q}{2}) + KL(q||\frac{p+q}{2})\right)$$

$$= \frac{1}{2}\left(\sum_x p(x)\log\frac{p(x)}{\frac{p(x)+q(x)}{2}} + \sum_x q(x)\log\frac{q(x)}{\frac{p(x)+q(x)}{2}}\right)$$

$$= \frac{1}{2}\left(-\sum_x (p(x)+q(x))\log\frac{p(x)+q(x)}{2} + \sum_x p(x)\log p(x) + \sum_x q(x)\log q(x)\right)$$

$$= -H(p+q) + \frac{1}{2}(H(p)+H(q))$$

2. Note that $H(x)$ is convex, thus the zero side of the inequality holds.

$$JSD(p\|q) = \frac{1}{2}\left(\sum_x p(x)\log\frac{p(x)}{\frac{p(x)+q(x)}{2}} + \sum_x q(x)\log\frac{q(x)}{\frac{p(x)+q(x)}{2}}\right)$$

$$\leq \frac{1}{2}\left(\sum_x p(x)\log 2 + \sum_x q(x)\log 2\right) \leq log2$$

Thus we have proved two side of the inequality.

3.

Theorem 1

$$\sqrt{\mathbb{E}_{x\sim p}[f(x)^2]} + \sqrt{\mathbb{E}_{x\sim p}[g(x)^2]} \geq \sqrt{\mathbb{E}_{x\sim p}[f(x)+g(x)]^2}$$

Proof. Square two side of the inequality, we only need to show that

$$\sqrt{\mathbb{E}_{x\sim p}[f(x)^2]}\sqrt{\mathbb{E}_{x\sim p}[g(x)^2]} \geq \mathbb{E}_{x\sim p}[f(x)g(x)]$$

Note that from Cauchy-Schwarz inequality, we have

$$\sum_x p(x)f(x)^2 \sum_x p(x)g(x)^2 \geq \left(\sum_x p(x)f(x)g(x)\right)^2$$

Which indicate that

$$\mathbb{E}_{x\sim p}[f(x)^2]\mathbb{E}_{x\sim p}[g(x)^2] \geq (\mathbb{E}_{x\sim p}[f(x)g(x)])^2$$

Bring this inequality back to the initial statement and we have proved the theorem. $\square$

Now we go back to the proof of initial statement.
We note that $a = p_1(x), b = p_2(x), c = p_3(x)$, from the theorem we proved(also note that the element under square is a non-negative number thus do not need to consider whether it is largher than 0 or not), we only need that:

$$\sqrt{\log b + \frac{a}{b}\log a - \frac{a+b}{b}\log\frac{a+b}{2}} + \sqrt{\log b + \frac{c}{b}\log c - \frac{b+c}{b}\log\frac{b+c}{2}}$$
$$\geq \sqrt{\frac{a}{b}\log a + \frac{c}{b}\log c - \frac{a+c}{b}\log\frac{a+c}{2}}$$

sqaure two side and we deduce the problem to

$$\left(\log\frac{a+b}{2b} + \log\frac{b+c}{2b} + \frac{a}{b}\log\frac{a+b}{a+c} + \frac{c}{b}\log\frac{b+c}{a+c}\right)$$

$$\leq 2\sqrt{\log b + \frac{a}{b}\log a - \frac{a+b}{b}\log\frac{a+b}{2}}\sqrt{\log b + \frac{c}{b}\log c - \frac{b+c}{b}\log\frac{b+c}{2}}$$

let $x = \frac{a+b}{2b}, y = \frac{c+b}{2b}$, we can rewrite the inequality as

$$\left(\log x + \log y + (2x-1)\log\frac{x}{x+y-1} + (2y-1)\log\frac{y}{x+y-1}\right)^2$$
$$\leq 4\left((2x-1)\log(2x-1) - 2x\log x\right)\left((2y-1)\log(2y-1) - 2y\log y\right)$$

derivate two part, we can gain the condition that the inequation has its local minimum from Lagrange multiplier(here if $x = 1$ or $y = 1$ then we already prove the inequality, thus we assume they are not equal to 1 to make the derivative meaningful)

$$\log\frac{x}{x+y-1}\left(\log x + \log y + (2x-1)\log\frac{x}{x+y-1} + (2y-1)\log\frac{y}{x+y-1}\right)$$

$$= 2\log\frac{2x-1}{x}\left((2y-1)\log(2y-1) - 2y\log y\right)$$

From the simlilar derivative to y, we combine them together and gain that:

$$\frac{\log x - \log(x+y-1)}{\log y - \log(x+y-1)} = \frac{f(x)}{f(y)} \tag{1}$$

$$f(x) = \frac{\log(2x-1) - \log x}{(2x-1)\log(2x-1) - 2x\log x}$$

Note that

$$f'(x) < 0 (x \geq 1)$$

bring this back to the eqution 1, we have that with x increasing, LHS increase while RHS decrease, thus the solution of x is unique. Since $x = 1$ is a solution, thus the only solution is $x = 1$, which shows that the only local minimum of this inequation is at $x = y = 1$. Thus, the initial statement is proved.

P4 **1.** Note that from Kantorovich-Rubinstein duality, we have

$$W(p,q) = \sup_{\|f\|_L \leq 1} \|\mathbb{E}_{x\sim p}[f(x)] - \mathbb{E}_{x\sim q}[f(x)]\|$$

thus for any f that $\|f\|_L \leq 1$, we have

$$W(p,r) + W(r,q) \geq \|\mathbb{E}_{x\sim p}[f(x)] - \mathbb{E}_{x\sim r}[f(x)]\| + \|\mathbb{E}_{x\sim r}[f(x)] - \mathbb{E}_{x\sim q}[f(x)]\|$$

$$\geq \|\mathbb{E}_{x\sim p}[f(x)] - \mathbb{E}_{x\sim q}[f(x)]\|$$

thus $W(p,r) + W(r,q) \geq W(p,q)$

2. Note that from Cauchy-Schwarz inequality, we have

$$W(p_x, p_{x+\epsilon}) \leq \mathbb{E}_{x\sim p_x, \epsilon\sim N(0,\sigma^2 I)} \|x - (x+\epsilon)\|_2$$

$$= \mathbb{E}[\|\epsilon\|_2] \leq \sqrt{\mathbb{E}[\|\epsilon\|_2^2]} = \sqrt{V}$$

3.

Lemma 1 Pinsker's inequality: for any two probability distribution p, q, we have

$$\delta(p,q) \leq \sqrt{\frac{1}{2} D_{KL}(p\|q)}$$

Proof. let $A = \{x | p(x) > q(x)\}$, then $\delta(p,q) = \sup_U \|p(U) - q(U)\| = p(A) - q(A)$

$$D_{KL}(p\|q) = \sum_x p(x)\log\frac{p(x)}{q(x)} \geq \sum_x p(x)(1 - \frac{q(x)}{p(x)}) = p(A) - q(A) = \delta(p,q)$$

From the conclusion we proved in 1, we have

$$W(p_r, p_q) \leq W(p_r, p_{r+\epsilon}) + W(p_{r+\epsilon}, p_{q+\epsilon}) + W(p_{q+\epsilon}, p_q)$$

From the conclusion we proved in 2, we have

$$W(p_r, p_{r+\epsilon}) \leq \sqrt{V}, W(p_q, p_{q+\epsilon}) \leq \sqrt{V}$$

And from hint 1 we have:

$$W(p_{r+\epsilon}, p_{q+\epsilon}) \leq C\delta(p_x, p_y) \leq C\sqrt{\frac{1}{2}D_{KL}(p_x\|p_y)}$$

Where the first inequality is gained from: any point in the support set has a variance of at most C, thus can be easily proved from the definition. The second inequality is gained from Pinsker's inequality(lemma).

4. Here are some possible tricks for training GANs:

- Add Gaussian Noise to the input(from 3, the Wesserstein distance is bounded by the variance of the input, thus adding noise can help to stabilize the training process)
- Add a gradient penalty, since we need a f that $\|f\|_L \leq 1$, we can add a penalty term to the loss function to make sure that the gradient is bounded.

These might also cause some potential problems:

- Add noises to the pictures might degrade the quality of the generated pictures.
- If σ is too small, the approximation that using JSD term to approximate the Wesserstein distance might not be accurate.
- Unlike Wasserstein distance, JSD does not provide meaningful gradients when distributions are disjoint, leading to training instability