Deep Learning 2023040163 Zhao Han Hong

Homework 4

1 True or False Questions

Problem 1

False.

Problem 2

False.

2 Q & A

Problem 3

1. We can directly write out

$$\begin{aligned}
\mathrm{JSD}(p||q) &= \frac{1}{2}\left(\sum_z p(z)\log\frac{2p(z)}{p(z)+q(z)} + \sum_z q(z)\log\frac{2q(z)}{p(z)+q(z)}\right) \\
&= -\sum_z \frac{p(z)+q(z)}{2}\log\frac{p(z)+q(z)}{2} + \frac{1}{2}\sum_z p(z)\log p(z) + \frac{1}{2}\sum_z q(z)\log q(z) \\
&= H\left(\frac{p+q}{2}\right) - \frac{1}{2}H(p) - \frac{1}{2}H(q).
\end{aligned}$$

2. Since KL divergence is non-negative, we know that JSD is also non-negative. On the other hand, we claim that

$$\frac{p(z)}{2}\log p(z) + \frac{q(z)}{2}\log q(z) - \frac{p(z)+q(z)}{2}\log\frac{p(z)+q(z)}{2} \le \frac{p(z)+q(z)}{2}\log 2.$$

In fact, since $x\log x$ is a convex function, we have

$$p(z)\log p(z) + q(z)\log q(z) \le (p(z)+q(z))\log(p(z)+q(z)).$$

Thus, we are done.

3. We first show that

$$\sqrt{f(x,y)} + \sqrt{f(y,z)} \geq \sqrt{f(z,x)},$$

where the function f is defined as

$$f(x,y) = x\log\frac{2x}{x+y} + y\log\frac{2y}{x+y}.$$

This can be proved by taking a derivative of y. We have

$$\frac{\partial}{\partial y}\sqrt{f(x,y)} = \frac{\log 2 + \log y - \log(x+y)}{2\sqrt{f(x,y)}}.$$

Now, we analyze the function

$$g(x,y) = \frac{\log\frac{2y}{x+y}}{2\sqrt{f(x,y)}} = \frac{\log\frac{2y}{x+y}}{2\sqrt{x\log\frac{2x}{x+y} + y\log\frac{2y}{x+y}}}.$$

First notice that $g(x,y) > 0$ for $x < y$, and $g(x,y) < 0$ for $x > y$. Moreover,

$$\frac{\partial}{\partial x}(g(x,y))^2 = \frac{\log\frac{2y}{x+y}}{4} \cdot \frac{2\left(-\frac{1}{x+y}\right)\left(x\log\frac{2x}{x+y} + y\log\frac{2y}{x+y}\right) - \log\frac{2y}{x+y}\log\frac{2x}{x+y}}{\left(x\log\frac{2x}{x+y} + y\log\frac{2y}{x+y}\right)^2}.$$

We then define

$$h(x,y) = -\frac{2x}{x+y}\log\frac{2x}{x+y} - \frac{2y}{x+y}\log\frac{2y}{x+y} - \log\frac{2y}{x+y}\log\frac{2x}{x+y}.$$

Notice that we can let $p = \frac{2x}{x+y}$, so it becomes

$$h_1(p) = -p\log p - (2-p)\log(2-p) - \log p\log(2-p).$$

We then take the derivative of p:

$$h_1'(p) = \log\frac{2-p}{p} - \frac{1}{p}\log(2-p) + \frac{1}{2-p}\log p,$$

$$\begin{aligned}\frac{d}{dp}\left(h_1'(p)\cdot p(2-p)\right) &= p(2-p)\left(-\frac{1}{p} - \frac{1}{2-p}\right) + (2-2p)\log\frac{2-p}{p} + \log(2-p) + 1 + \log p + 1 \\ &= (3-2p)\log(2-p) + (2p-1)\log p,\end{aligned}$$

$$\begin{aligned}
\frac{d^2}{dp^2}\left(h_1'(p)\cdot p(2-p)\right) &= -2\log(2-p)+2\log p-\frac{3-2p}{2-p}+\frac{2p-1}{p}\\
&= 2\log\frac{p}{2-p}-\frac{1}{p}+\frac{1}{2-p},
\end{aligned}$$

$$\frac{d^3}{dp^3}\left(h_1'(p)\cdot p(2-p)\right)=2\left(\frac{1}{p}+\frac{1}{2-p}\right)+\frac{1}{p^2}+\frac{1}{(2-p)^2}>0.$$

Thus, we know that $\frac{d}{dp}\left(h_1'(p)\cdot p(2-p)\right)\geq 0$ always hold, so $h_1'(p)$ is a strictly increasing function, implying that $h_1(p)\geq 0$ always hold. We then know that $g(x,y)$ is strictly increasing both when $x<y$ and $x>y$. Moreover, when $x\to y$, we can find

$$\begin{aligned}
\lim_{x\to y} g(x,y)^2 &= \lim_{x\to y}\frac{1}{4}\cdot\frac{\left(\frac{y-x}{x+y}\right)^2}{x\log\frac{2x}{x+y}+y\log\frac{2y}{x+y}}\\
&= \frac{1}{4}\lim_{x\to y}\frac{\frac{1}{2y^2}(x-y)}{\log\frac{2x}{x+y}}=\frac{1}{4y}.
\end{aligned}$$

Thus, $g(x,y)$ **increases from** $\frac{1}{2}\sqrt{\frac{\log 2}{y}}$ **to** $\frac{1}{2\sqrt{y}}$ **as** x **increases from** 0 **to** y**, and increases from** $-\frac{1}{2\sqrt{y}}$ **to 0 as** x **increases from** y **to** $+\infty$**.**

We can then go back to the derivative of the original equation: the derivative of LHS w.r.t. y is

$$D=\frac{\partial\sqrt{f(x,y)}}{\partial y}+\frac{\partial\sqrt{f(y,z)}}{\partial y}=g(x,y)+g(z,y).$$

Without loss of generality, we can assume that $x\leq z$. Then, when $y<x$, we have $D<0$; when $y>z$, we have $D>0$. When $x\leq y\leq z$, we can notice that D is strictly decreasing, and since

$$\begin{aligned}
\lim_{y\to x^+} g(x,y)+g(z,y) &= g(z,x)+\frac{1}{2\sqrt{x}}>0,\\
\lim_{y\to z^-} g(x,y)+g(z,y) &= -\frac{1}{2\sqrt{z}}+g(x,z)<0,
\end{aligned}$$

we know that there exists $y_0\in[x,z]$, such that $D(y_0)=0$; moreover, when $y>y_0$, $D(y)<0$, and when $y<y_0$, $D(y)>0$. Thus, the function LHS(y) first decreases until $y=x$, then increases until $y=y_0$, then decreases again until $y=z$, then increases again. Thus, it suffices to verify that LHS$(y=x)$ and LHS$(y=z)$ are both no lesser than RHS, which is clearly the case. That finishes the proof of

$$\sqrt{f(x,y)}+\sqrt{f(y,z)}\geq\sqrt{f(z,x)}.$$

Finally, we may come back to the original problem. The inequality we have to prove is

$$\sqrt{\sum_z \frac{1}{2} f(p_1(z), p_2(z))} + \sqrt{\sum_z \frac{1}{2} f(p_2(z), p_3(z))} \geq \sqrt{\sum_z \frac{1}{2} f(p_3(z), p_1(z))}.$$

However, notice the Minkowski's inequality

$$\begin{aligned}\sqrt{\sum_z f(p_1(z), p_2(z))} + \sqrt{\sum_z f(p_2(z), p_3(z))} &\geq \sqrt{\sum_z \left(\sqrt{f(p_1(z), p_2(z))} + \sqrt{f(p_2(z), p_3(z))}\right)^2} \\ &\geq \sqrt{\sum_z f(p_1(z), p_3(z))},\end{aligned}$$

so we are done.

Problem 4

1. Using the Kantorovich-Rubinstein duality, we have

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

Thus,

$$\begin{aligned}W(p, q) + W(q, r) &= \sup_{\|f\|_L \leq 1} [\mathbb{E}_{x\sim p}[f(x)] - \mathbb{E}_{x\sim q}[f(x)]] + \sup_{\|f\|_L \leq 1} [\mathbb{E}_{x\sim q}[f(x)] - \mathbb{E}_{x\sim r}[f(x)]] \\ &\geq \sup_{\|f\|_L \leq 1} [\mathbb{E}_{x\sim p}[f(x)] - \mathbb{E}_{x\sim r}[f(x)]] = W(p, r).\end{aligned}$$

2. We have

$$\begin{aligned}W(p_x, p_{x+\epsilon}) &= \inf_{\gamma\in\Gamma} \iint_{\mathcal{X}\times\mathcal{X}} \|x - y\| \gamma(x, y) dx dy \\ &= \inf_{\gamma\in\Gamma} \iint_{\mathcal{X}\times\mathbb{R}^n} \|\epsilon\| \gamma(x, x + \epsilon) dx d\epsilon,\end{aligned}$$

and the constraints become

$$\int \gamma(x, x + \epsilon) d\epsilon = p_x(x)$$

$$\int \gamma(x, y) dx = p_{x+\epsilon}(y) = \mathbb{E}_{x\sim p_x}[p_\epsilon(y - x)].$$

To prove that $W(p_x, p_{x+\epsilon}) \leq \sqrt{\mathbb{E}[||\epsilon||_2^2]}$, we only have to construct a proper γ. We can simply pick $\gamma(x, x + \epsilon) = p_x(x)p_\epsilon(\epsilon)$, then the first constraint directly holds; the second constraint holds since

$$\int \gamma(x,y)dx = \int p_x(x)p_\epsilon(y-x)dx = \mathbb{E}_{x\sim p_x}[p_\epsilon(y-x)].$$

Finally, we compute

$$W(p_x, p_{x+\epsilon}) \leq \iint_{\mathcal{X}\times\mathbb{R}^n} \|\epsilon\| p_x(x)p_\epsilon(\epsilon)dxd\epsilon = \mathbb{E}[||\epsilon||] \leq \sqrt{\mathbb{E}[||\epsilon||_2^2]}.$$

3. We first assume that the hints are true, then by triangle inequality and the hint,

$$W(p_r,p_g) \leq W(p_r,p_{r+\epsilon}) + W(p_g,p_{g+\epsilon}) + W(p_{r+\epsilon},p_{g+\epsilon}) \leq 2V^{\frac{1}{2}} + C\delta\left(p_{r+\epsilon},p_{g+\epsilon}\right).$$

Moreover, notice that

$$\begin{aligned}
2\sqrt{\mathrm{JSD}(p_{r+\epsilon}||p_{g+\epsilon})} &= \sqrt{2\mathrm{KL}\left(p_{r+\epsilon}||\frac{p_{r+\epsilon}+p_{g+\epsilon}}{2}\right) + 2\mathrm{KL}\left(p_{g+\epsilon}||\frac{p_{r+\epsilon}+p_{g+\epsilon}}{2}\right)} \\
&\geq \sqrt{\mathrm{KL}\left(p_{r+\epsilon}||\frac{p_{r+\epsilon}+p_{g+\epsilon}}{2}\right)} + \sqrt{\mathrm{KL}\left(p_{g+\epsilon}||\frac{p_{r+\epsilon}+p_{g+\epsilon}}{2}\right)} \\
&\geq \sqrt{2}\left(\delta\left(p_{r+\epsilon},\frac{p_{r+\epsilon}+p_{g+\epsilon}}{2}\right) + \delta\left(p_{g+\epsilon},\frac{p_{r+\epsilon}+p_{g+\epsilon}}{2}\right)\right) \\
&\geq \sqrt{2}\delta\left(p_{r+\epsilon},p_{g+\epsilon}\right),
\end{aligned}$$

so we are done.

4. The trick is that we may add noise ϵ to both the real images and the generated images, and gradually decrease the variance of the noise. In this way, we may optimize the upper bound of $W(p_r,p_g)$, so the Wasserstein distance can be optimized.

The potential issue is that the noise may make the training more unstable, and the generator may go to unespected minimum since the objective $\mathrm{JSD}(p_{r+\epsilon}||p_{g+\epsilon})$ is different from the original objective $\mathrm{JSD}(p_r||p_g)$.