Deep Learning 2023040163 Zhao Han Hong

Homework 4

1 True or False Questions

Problem 1

True.

2 Q & A

Problem 2

1. We can directly get

$$x_t = \sqrt{\alpha_t\alpha_{t-1}\cdots\alpha_1}x_0 + \sqrt{1-\alpha_t}\epsilon_{t-1} + \sqrt{\alpha_t(1-\alpha_{t-1})}\epsilon_{t-2} + \cdots + \sqrt{\alpha_t\alpha_{t-1}\cdots\alpha_2(1-\alpha_1)}\epsilon_0.$$

Next, we use the moment generating function to get the distribution of

$$\epsilon = \frac{1}{\sqrt{1-\bar{\alpha}_t}}\left(\sqrt{1-\alpha_t}\epsilon_{t-1} + \sqrt{\alpha_t(1-\alpha_{t-1})}\epsilon_{t-2} + \cdots + \sqrt{\alpha_t\alpha_{t-1}\cdots\alpha_2(1-\alpha_1)}\epsilon_0\right).$$

Since the linear combinations of independent Gaussian variables are still Gaussian, we know that ϵ is Gaussian. Moreover, we can compute the mean and variance of ϵ: $\mathbb{E}(\epsilon) = 0$,

$$\begin{aligned}\mathbb{V}(\epsilon) &= \frac{(1-\alpha_t)\mathbf{I} + \alpha_t(1-\alpha_{t-1})\mathbf{I} + \cdots + \alpha_t\alpha_{t-1}\cdots\alpha_2(1-\alpha_1)\mathbf{I}}{1-\bar{\alpha}_t}\\ &= \frac{1-\alpha_t\alpha_{t-1}\cdots\alpha_1}{1-\alpha_t\alpha_{t-1}\cdots\alpha_1}\mathbf{I} = \mathbf{I}.\end{aligned}$$

Thus, we know that $\epsilon \sim \mathcal{N}(0, \mathbf{I})$, so we are done.

2. Given x_t and x_0, ϵ is then fixed by

$$\epsilon = \frac{x_t - \sqrt{\bar{\alpha}_t}x_0}{\sqrt{1-\bar{\alpha}_t}}.$$

The conditional probability is given by

$$q(x_{t-1}|x_t, x_0) = \frac{p_{\epsilon'}(\epsilon')p_{\epsilon_{t-1}}(\epsilon_{t-1})}{p_\epsilon(\epsilon)},$$

where

$$\epsilon_{t-1} = \frac{x_t - \sqrt{\alpha_t}x_{t-1}}{\sqrt{1-\alpha_t}},$$

$$\epsilon' = \frac{x_{t-1} - \sqrt{\alpha_{t-1}\cdots\alpha_1}x_0}{\sqrt{1-\alpha_{t-1}\cdots\alpha_1}} = \frac{\sqrt{1-\alpha_{t-1}}\epsilon_{t-2} + \sqrt{\alpha_{t-1}(1-\alpha_{t-2})}\epsilon_{t-3} + \cdots + \sqrt{\alpha_{t-2}\cdots\alpha_2(1-\alpha_1)}\epsilon_0}{\sqrt{1-\alpha_{t-1}\cdots\alpha_1}}.$$

From the arguments in 1 we know that both ϵ_{t-1}, ϵ and ϵ' are Gaussian variable with mean 0 and variance $\mathbf{I}$. Thus, we have

$$q(x_{t-1}|x_t, x_0) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(\frac{1}{2}\left(\frac{||x_t - \sqrt{\bar{\alpha}_t}x_0||^2}{1-\bar{\alpha}_t} - \frac{||x_t - \sqrt{\alpha_t}x_{t-1}||^2}{1-\alpha_t} - \frac{||x_{t-1} - \sqrt{\overline{\alpha_{t-1}}}x_0||^2}{1-\overline{\alpha_{t-1}}}\right)\right).$$

This is a Gaussian distribution with respect to x_{t-1}, and its mean is given by

$$\begin{aligned}
\tilde{\mu_t} &= \frac{\frac{2\sqrt{\alpha_t}}{1-\alpha_t}x_t + \frac{2\sqrt{\overline{\alpha_{t-1}}}}{1-\overline{\alpha_{t-1}}}x_0}{2\left(\frac{\alpha_t}{1-\alpha_t} + \frac{1}{1-\overline{\alpha_{t-1}}}\right)} \\
&= \frac{\sqrt{\alpha_t}(1-\frac{\bar{\alpha}_t}{\alpha_t})x_t + (1-\alpha_t)\cdot\sqrt{\frac{\bar{\alpha}_t}{\alpha_t}}x_0}{1-\bar{\alpha}_t} \\
&= \frac{(\alpha_t - \bar{\alpha}_t)x_t + (1-\alpha_t)\cdot(-\sqrt{1-\bar{\alpha}_t}\epsilon + x_t)}{\sqrt{\alpha_t}(1-\bar{\alpha}_t)} \\
&= \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha_t}}}\epsilon\right).
\end{aligned}$$

3. For the first inequality, we have

$$\begin{aligned}
\mathbb{E}_{q(x_{0:T})}\left[\log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right] &= \sum_{x_0} q(x_0) \sum_{x_{1:T}} q(x_{1:T}|x_0)\log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} \\
&= \sum_{x_0} q(x_0)\left(\sum_{x_{1:T}} q(x_{1:T}|x_0)\log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{1:T}|x_0)} - \log p_\theta(x_0)\right) \\
&\geq \sum_{x_0} q(x_0)(-\log p_\theta(x_0)) \\
&= -\mathbb{E}_{q(x_0)}[\log p_\theta(x_0)].
\end{aligned}$$

For the second equality, we have:

RHS$-$LHS

$$
\begin{aligned}
&= \mathbb{E}_q\Bigg[\sum_{x_T} q(x_T|x_0)\log\frac{q(x_T|x_0)}{p_\theta(x_T)} + \sum_{t=2}^{T}\sum_{x_{t-1}} q(x_{t-1}|x_t,x_0)\log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)} - \log p_\theta(x_0|x_1) \\
&\quad - \log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\Bigg] \\
&= \mathbb{E}_q\Bigg[\sum_{x_T} q(x_T|x_0)\log\frac{q(x_T|x_0)}{p_\theta(x_T)} + \sum_{x_{T-1}} q(x_{T-1}|x_T,x_0)\log\frac{q(x_{T-1}|x_T,x_0)}{p_\theta(x_{T-1}|x_T)} \\
&\quad + \sum_{x_{T-2}} q(x_{T-2}|x_{T-1},x_0)\log\frac{q(x_{T-2}|x_T,x_0)}{p_\theta(x_{T-2}|x_{T-1})} + \cdots + \sum_{x_1} q(x_1|x_2,x_0)\log\frac{q(x_1|x_2,x_0)}{p_\theta(x_1|x_2)} \\
&\quad - \log p_\theta(x_0|x_1) - \log\frac{q(x_T|x_0)q(x_{T-1}|x_T,x_0)\cdots q(x_1|x_2,x_0)}{p_\theta(x_T)p_\theta(x_{T-1}|x_T)\cdots p_\theta(x_0|x_1)}\Bigg] \\
&= \sum_{x_T,x_0} q(x_T,x_0)\log\frac{q(x_T|x_0)}{p_\theta(x_T)} + \sum_{x_{T-1},x_T,x_0} q(x_{T-1},x_T,x_0)\log\frac{q(x_{T-1}|x_T,x_0)}{p_\theta(x_{T-1}|x_T)} \\
&\quad + \sum_{x_{T-2},x_{T-1},x_0} q(x_{T-2},x_{T-1},x_0)\log\frac{q(x_{T-2}|x_T,x_0)}{p_\theta(x_{T-2}|x_{T-1})} + \cdots + \sum_{x_1,x_2,x_0} q(x_1|x_2,x_0)\log\frac{q(x_1,x_2,x_0)}{p_\theta(x_1|x_2)} \\
&\quad - \Bigg(\sum_{x_T,x_0} q(x_T,x_0)\log\frac{q(x_T|x_0)}{p_\theta(x_T)} + \sum_{x_{T-1},x_T,x_0} q(x_{T-1},x_T,x_0)\log\frac{q(x_{T-1}|x_T,x_0)}{p_\theta(x_{T-1}|x_T)} + \cdots \\
&\quad + \sum_{x_1,x_2,x_0} q(x_1,x_2,x_0)\log\frac{q(x_1|x_2,x_0)}{p_\theta(x_1|x_2)}\Bigg) = 0,
\end{aligned}
$$

so we are done.

4. We plug in the expression of $\tilde{\mu_t}$ and μ_θ to get

$$
\begin{aligned}
L_t &= \frac{1}{2||\Sigma_\theta||_2^2}\mathbb{E}_{x_0,\epsilon}\left[\left|\left|\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha_t}}}\epsilon\right) - \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,x)\right)\right|\right|^2\right] \\
&= \frac{1}{2||\Sigma_\theta||_2^2}\cdot\frac{(1-\alpha_t)^2}{\alpha_t(1-\bar{\alpha}_t)}\mathbb{E}_{x_0,\epsilon}\left[||\epsilon_\theta(x_t,x)-\epsilon||^2\right].
\end{aligned}
$$

Now, using that $x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$, we are done.

Problem 3

We begin from the second expression:

$$\begin{aligned}
&\mathbb{E}_{x\sim p_{\text{data}}(x),\tilde{x}\sim q_\sigma(\tilde{x}|x)}[\nabla_{\tilde{x}}\log q_\sigma(\tilde{x}|x)^T s_\theta(\tilde{x})] \\
&= \iint p_{\text{data}}(x) q_\sigma(\tilde{x}|x)\nabla_{\tilde{x}}\log q_\sigma(\tilde{x}|x)^T s_\theta(\tilde{x}) d\tilde{x}dx \\
&= \iint p_{\text{data}}(x)\nabla_{\tilde{x}} q_\sigma(\tilde{x}|x)^T s_\theta(\tilde{x}) d\tilde{x}dx \\
&= \iint p_{\text{data}}(x)\nabla_{\tilde{x}} \left(\frac{q_\sigma(\tilde{x},x)}{p_{\text{data}}(x)}\right)^T s_\theta(\tilde{x}) d\tilde{x}dx \\
&= \iint \nabla_{\tilde{x}} \left(q_\sigma(\tilde{x},x)\right)^T s_\theta(\tilde{x}) d\tilde{x}dx \\
&= \int q_\sigma(\tilde{x})\nabla_{\tilde{x}}\log q_\sigma(\tilde{x})^T s_\theta(\tilde{x}) d\tilde{x}.
\end{aligned}$$

Thus, we are done.

Problem 4

In the diffusion process, we have

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \epsilon \sim \mathcal{N}(0, \mathbf{I}),$$

and our model aims to estimate

$$\epsilon \approx \epsilon_\theta(x_t, t).$$

On the other hand, the score-based model gives

$$s_\theta(x_t, t) \approx \nabla \log p(x_t),$$

so what we need is to relate ϵ with $\nabla \log p(x_t)$. We first show a lemma.

Lemma (Tweedie's formula). Let σ be a fixed value and θ be drawn from $p(\theta)$. Next, we pick several $\theta_1, \cdots, \theta_n$ and randomly draw x_i from each θ_i such that

$$p(x|\theta) \sim \mathcal{N}(\theta, \sigma^2).$$

Then, we claim that

$$\mathbb{E}_\theta[\theta|x] = x + \sigma^2 \frac{d}{dx}\log p(x).$$

Proof. We can prove it using the definition.

$$
\begin{aligned}
\mathbb{E}_\theta[\theta|x] &= \int \theta p(\theta|x) d\theta \\
&= \int \theta \frac{p(x|\theta)p(\theta)}{p(x)} d\theta \\
&= \frac{1}{p(x)} \int \theta \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\theta)^2} p(\theta) d\theta \\
&= \frac{1}{p(x)} \left(x \int p(x|\theta)p(\theta)d\theta + \int \sigma^2 \frac{d}{dx} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\theta)^2} \right) p(\theta) d\theta \right) \\
&= \frac{1}{p(x)} \left(x \int p(x|\theta)p(\theta)d\theta + \int \sigma^2 \frac{d}{dx} \left(p(x|\theta) \right) p(\theta) d\theta \right) \\
&= x + \sigma^2 \frac{d}{dx} \log p(x).
\end{aligned}
$$

Now, using the **Lemma**, we can yield

$$\sqrt{\bar{\alpha}_t} x_0 \approx x_t + (1 - \bar{\alpha}_t) \nabla \log p(x_t),$$

where the approximate sign means that the right-hand side is the mean value (i.e. the best estimation) of x_0 given that we only know x_t. Thus, we can also use the gradient to estimate the noise:

$$\epsilon_\theta(x_t, t) = \epsilon = -\sqrt{1 - \bar{\alpha}_t} \nabla \log p(x_t).$$

Thus, the relation between the DDPM model and the score function is

$$s_\theta(x_t, t) = -\frac{1}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t).$$