بسمه تعالی

امتحان پایانی – مدل‌های مولد

نام و نام خانوادگی:

شماره دانشجویی:

صحیح و غلط (با ذکر دلیل دقیق) (۲۰ نمره)

۱. با توجه به اینکه در مدل normalizing flow به شکل دقیق مقدار MLE محاسبه می‌شود، با کیفت ترین تصاویر در میان مدل‌های مولد را می‌تواند تولید کند.

۲. زمان لازم برای تولید نمونه در مدل EBM از سایر مدل‌های مولد بهینه‌تر است.

۳. در انتخاب سطوح نویز متفاوت در Annealed Langevin dynamics، باید واریانس نویزها به اندازه کافی نزدیک باشند تا توزیع‌های مختلف همپوشانی کافی داشته باشند.

پاسخ تشریحی (۴۰ نمره)

۴. چرا در صورت یادگیری به کمک الگوریتم‌های score matching و نمونه‌گیری به کمک Langevin dynamic همچنان نمی‌توان به تصاویر با کیفیتی رسید (۳ دلیل بیاورید)؟ چه راه حلی پیشنهاد می‌کنید؟

۵. در شبیه سازی پیوسته مدل‌های score based به کمک معادلات دیفرانسیل تصادفی می‌توان از روش‌های عددی برای حل معادله برای مثال به شکل زیر استفاده کرد:

$$\mathbf{x} \leftarrow \mathbf{x} - \sigma(t)^2 \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}, t)\Delta t + \sigma(t)\ \mathbf{z} \quad (\mathbf{z} \sim \mathcal{N}(\mathbf{0}, |\Delta t|\ \mathbf{I}))$$
$$t \leftarrow t + \Delta t$$

- آیا این رویکرد برای رسیدن به فرآیند generation مناسب کافی است؟
- الگوریتم predictor-corrector چگونه در فرآیند حل معادله بالا کمک می‌کند و ارتباط آن با Annealed Langevin Dynamic چیست؟

۶. در الگوریتم Denoising score matching تابع score برای فضای نمونه‌های نویزی با کمینه کردن fisher divergence تخمین زده می‌شود. همانطور که می‌دانید این معیار در این روش برابر است با:

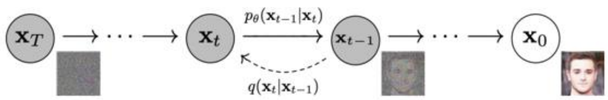
$$\begin{aligned}
&\frac{1}{2} E_{\tilde{\mathbf{x}} \sim p_{\text{data}}}[\|\boldsymbol{s}_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}})\|_2^2] \\
=&\frac{1}{2} E_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x}), \tilde{\mathbf{x}} \sim q_\sigma(\tilde{\mathbf{x}}|\mathbf{x})}[\|\boldsymbol{s}_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}} \mid \mathbf{x})\|_2^2] + \text{const.}
\end{aligned}$$

نشان دهید در صورت در نظر گرفتن نویز نرمال، شبکه تعلیم دیده در این روش در عمل نویز اضافه شده به هر نمونه را تخمین می‌زند.

۷. برای آموزش یک مدل EBM تابع هزینه KL-divergence را میان توزیع مدل و توزیع داده نوشته و به گونه‌ای با ذکر محاسبات آن را بازنویسی کنید که به کمک آن و sampling از شبکه بتوان مدل را آموزش داد. در رابطه با sampling از این مدل نیز در حین training توضیح دهید.

سوال توسعه‌ای (۴۰ نمره)

۸. مدل diffusion یک مدل با چند لایه نهان محسوب می‌شود که در آن توزیع پیشین روی مقدار متغیر نهان نهایی، توزیع نرمال و توزیع شرطی هر لایه نسبت به لایه قبل نیز یک توزیع نرمال است.



The forward process: $q_t(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t I)$.

The reverse process: $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}\left(\mathbf{x}_{t-1}|\mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t)\right)$.

Conditionals have closed form: $q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_{t-1}; \tilde{\mu}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t I\right)$

Where: $\tilde{\mu}(\mathbf{x}_t, \mathbf{x}_0) = \dfrac{\sqrt{\tilde{\alpha}_t}\beta_t}{1-\tilde{\alpha}_t}\mathbf{x}_0 + \dfrac{\sqrt{\alpha_t}(1-\tilde{\alpha}_{t-1})}{1-\tilde{\alpha}_t}\mathbf{x}_t$.

And: $\alpha_t = 1-\beta_t, \;\; \tilde{\alpha}_t = \prod_{s=1}^{t}\alpha_s, \;\; \tilde{\beta}_t = \dfrac{1-\tilde{\alpha}_{t-1}}{1-\tilde{\alpha}_t}\beta_t$.

$$\mathop{\mathbb{E}}_{\substack{\mathbf{x}_0\sim p\\ \mathbf{x}_{1:T}\sim q}}\left[-\sum_{t>1}\log\frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}\right] = \mathop{\mathbb{E}}_{\substack{\mathbf{x}_0\sim p\\ \mathbf{x}_t\sim q_t}}\left[D(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0 \parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))\right]$$
$$= \mathop{\mathbb{E}}_{\substack{\mathbf{x}_0\sim p\\ \mathbf{x}_t\sim q_t}}\left[\frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(\mathbf{x}_t,\mathbf{x}_0) - \mu_\theta(\mathbf{x}_t,t)\|^2\right] + C.$$

$$\mathop{\mathbb{E}}_{\substack{\mathbf{x}_0\sim p\\ \mathbf{x}_{1:T}\sim q}}\left[-\sum_{t>1}\log\frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}\right] = \mathop{\mathbb{E}}_{\substack{\mathbf{x}_0\sim p\\ \varepsilon\sim\mathcal{N}(0,I)}}\left[\frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(\mathbf{x}_t,\mathbf{x}_0) - \mu_\theta(\mathbf{x}_t,t)\|^2\right] + C$$
$$= \mathop{\mathbb{E}}_{\substack{\mathbf{x}_0\sim p\\ \varepsilon\sim\mathcal{N}(0,I)}}\left[\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\tilde{\alpha}_t)}\|\varepsilon - \varepsilon_\theta\left(x_t(\mathbf{x}_0,\varepsilon),t\right)\|^2\right] + C.$$

توزیع نرمال با وجود مزایای محاسباتی در عمل به دلیل ظرفیت پایین گاهی در مدل‌سازی کافی نمی‌باشد که اصطلاحا منجر به مشکل prior hole problem می‌شود که در آن در مناطقی p_{data} زیاد ولی p_{model} کم و یا صفر است. یک دلیل این اتفاق این است که در مدل diffusion با معماری کنونی روابط متغیرهای نهان در لایه‌های مختلف به اندازه کافی مدل می‌شود اما روابط متغیرهای یک لایه به خوبی مدل نمی‌شوند (برای مثال پیکسل‌های یک تصویر در یک لایه نویز مشخص).
به کمک رویکرد EBM و ترکیب مدل diffusion و EBM معماری و تابع هزینه‌ای را پیشنهاد دهید که تا حدی این مشکل را بتواند برطرف کند.