CS 598-DGM: Spring'25
Deep Generative Models

Homework 3

(Due Monday, April 21, 11:59 pm)

- The homework is due at 11:59 pm on the due date. We will be using Gradescope for the homework assignments. **Please do NOT email a copy of your solution.** Contact the TAs (Zhijie, Rohan) if you are having technical difficulties in submitting the assignment. We will NOT accept late submissions.
- Please make sure that each question is clearly marked. You may use as many pages as needed but do not change the order of the questions and answers.
- You are expected to typeset the solutions, i.e., **handwritten solutions will not be graded**. We encourage you to use LaTeX. **When submitting on Gradescope, you are required to assign the correct pages for each sub-problem to the provided outline. If pages are incorrectly assigned or left unassigned on Gradescope, it will result in no credit, and regrade requests regarding this will be declined.** Double-check your submission to ensure accuracy.
- Please use Slack first if you have questions about the homework. You can also come to our (zoom) office hours and/or send us e-mails. If you are sending us emails with questions on the homework, please start subject with "CS 598-DGM: " and send the email to *all course staff*: Arindam, Zhijie, and Rohan.
- The homework consists of written assignments. Please submit your report as a PDF file.

1. (30 points) In the context of Denoising AutoEncoders (DAEs), one uses the smoothed (or noisy) distribution $q_\sigma(\tilde{\mathbf{x}}) = \int_x q_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) q_0(\mathbf{x}) d\mathbf{x}$ for the modeling for a suitable choice of $q_\sigma(\tilde{\mathbf{x}}|\mathbf{x})$. Let $s_\theta(\tilde{\mathbf{x}})$ denote the score function to be estimated.

   (a) (10 points) Let

$$J_1(\theta) = \mathbb{E}_{q_\sigma(\tilde{\mathbf{x}})}\left[\frac{1}{2}\left\|s_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}})\right\|^2\right] , \tag{1}$$

$$J_2(\theta) = \mathbb{E}_{q_\sigma(\mathbf{x},\tilde{\mathbf{x}})}\left[\frac{1}{2}\left\|s_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}}|\mathbf{x})\right\|^2\right] . \tag{2}$$

   Show that $J_1(\theta) = J_2(\theta) + c$ under suitable regularity conditions where c is a constant independent of θ. Please specify the conditions you have used to establish the identity.

   (b) (20 points) For this problem, we assume that the conditional distribution $q_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) = \mathcal{N}(\mathbf{x}, \sigma^2\mathbb{I})$. Consider three univariate smoothed distributions $q_{\sigma_0}(\tilde{\mathbf{x}}, \mathbf{x}), q_{\sigma_1}(\tilde{\mathbf{x}}, \mathbf{x})$, and $q_{\sigma_2}(\tilde{\mathbf{x}}, \mathbf{x})$ with $0 < \sigma_0 < \sigma_1 < \sigma_2$. Note that the $\sigma_i, i = 0, 1, 2$ correspond to the standard deviations of the corresponding Gaussian conditional distributions $q_{\sigma_i}(\tilde{\mathbf{x}}|\mathbf{x})$ and $q_{\sigma_i}(\tilde{\mathbf{x}}, \mathbf{x}) = q_{\sigma_i}(\tilde{\mathbf{x}}|\mathbf{x}) q_0(\mathbf{x})$. Further, we assume $\sigma_0^2 \leq \frac{\sigma_1^2}{d}$ where d is the dimensionality, i.e., $\tilde{\mathbf{x}} \in \mathbb{R}^d$.[1]

   In this setting, with $q_{\sigma_i}, i = 0, 1, 2$, denoting joint distributions, Professor Super Smooth claims that we always have

$$KL(q_{\sigma_1} \| q_{\sigma_0}) \leq KL(q_{\sigma_2} \| q_{\sigma_0}) , \tag{3}$$

   i.e., smoothing with a higher variance ($\sigma_2 > \sigma_1$) Gaussian moves the smoothed joint distribution further away from the joint distribution q_{σ_0}.

   Do you agree/disagree with the Professor? If you agree, you have to prove the claim. If you disagree, you will have to give a counterexample to the claim.

[1] Intuitively, with $\sigma_0 \approx 0$, we have $q_{\sigma_0}(\tilde{\mathbf{x}}) \approx q_0(\mathbf{x})$, so q_{σ_0} can be viewed as an accurate approximation of the true distribution q_0.

2. (18 points) Assume you have trained a generative $p_{\hat{\theta}}(x)$ which accurately models some target distribution $p_*(x)$, i.e., $p_{\hat{\theta}}(x) \approx p_*(x)$. We consider the problem of *likelihood computation*:

(Likelihood Computation)   Given x_{test}, what is the value of $p_{\hat{\theta}}(x_{\text{test}})$?

Consider suitable versions of the following four family of models for $p_{\hat{\theta}}(x)$

(a) Variational Auto-Encoders (VAEs),

(b) Generative Adversarial Networks (GANs),

(c) Diffusion model, specifically Score-SDE [1], which uses isotropic Gaussian as the source distribution,

(d) Flow matching model, specifically Rectified Flow [2], which uses isotropic Gaussian as the source distribution,

(e) Flow matching model, specifically Rectified Flow [2], which uses non-Gaussian $p_0(x)$ as the source distribution, where we cannot compute $p_0(x)$ given any x, but have n samples $\{x_i^{(0)}, i \in [n]\}$ from $p_0(x)$, i.e., $x_i^{(0)} \sim p_0(x), i \in [n]$.

Please answer the following questions:

(i) ($3 \times 5 = 15$ points) For the each of the above five models above, can the model compute $p_{\hat{\theta}}(x_{\text{test}})$ for any given x_{test} exactly?[2] Briefly justify each answer. You can assume ability for 'simulation', e.g., solving ODE/SDE precisely.

(ii) (3 points)[3] For the models which can compute $p_{\hat{\theta}}(x_{\text{test}})$, briefly outline the computation needed, starting from the model for $p_{\hat{\theta}}(x)$ and x_{test}.

References:
[1] (Score-SDE) Y. Song, J. Sohl-Dickstein, D. Kingma, A. Kumar, S. Ermon, B. Poole. Score-based generative modeling through stochastic differential equations. ICLR, 2021.
[2] (Rectified Flow) X. Liu, C. Gong, and Q. Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. ICLR, 2023.

[2] If it is an approximation, upper bound, or lower bound, then that is not acceptable in the context of this question.
[3] We are not giving the breakdown over the five models as that will reveal which ones can actually do the computation correctly.

3. (10 points) The question considers the generative model presented in this paper—we will refer to the model as Score-SDE:

> Y. Song, J. Sohl-Dickstein, D. Kingma, A. Kumar, S. Ermon, B. Poole. Score-based generative modeling through stochastic differential equations ICLR, 2021.

You have been given three things

(a) A trained Score-SDE model $p_{\theta_1}(x)$ over `Dogs`, where the model uses isotropic Gaussians as the source distribution $p_0(x)$,

(b) A trained Score-SDE model $p_{\theta_2}(x)$ over `Huskies`, where the model uses isotropic Gaussians as the source distribution $p_0(x)$,

(c) A set of n samples $\{x_i^{(1)}, i \in [n]\}$ of `Dogs`,

where θ_1, θ_2 are respectively the Score-SDE model parameters for `Dogs` and `Huskies`. Please see the Score-SDE model for details on what these parameters are.

(i) (5 points) Present an inference algorithm which samples a Dog $x_i^{(1)}$ uniformly and then uses suitable reverse SDE to generate a Husky sample.

(ii) (5 points) Argue why such inference may be faster than starting from an isotropic Gaussian sample.

4. (42 points) Let $X_0 \sim p_0(x)$ (not necessarily Gaussian) be the source distribution and $X_1 \sim p_1(x)$ be the target distribution for generative modeling. This question considers conditional flow matching (CFM) and Schrödinger bridge (SB) for generative modeling.

(a) (5 points) What is main difference between training flow models based on maximum likelihood and conditional flow matching? Please explain the difference using suitable mathematical notation.

(b) (5 points) For CFM, given any arbitrary choice of conditional probability path $p_t(x|z)$, can the conditional velocity field (VF) $u_t(x|z)$ be obtained in closed form? Clearly justify your answer.

(c) (12 points) We consider affine conditional flows of the following form for flow matching:

$$\psi_t(x|x_1) = \alpha_t x_1 + \sigma_t x \ , \quad \alpha_0 = 0 = \sigma_1, \alpha_1 = 1, \text{ and } \dot{\alpha}_t, -\dot{\sigma}_t > 0 \ , \ t \in (0,1) \ . \tag{4}$$

i. (4 points) Show that $\alpha_t = t, \sigma_t = (1-t)$ is a valid choice for the parameters.

ii. (4 points) What is the marginal VF for the above choices of the parameters?

iii. (4 points) Will the corresponding transport path be the same as the one obtained by running optimal transport[4] between p_0 and p_1? Clearly justify your answer.

(d) (10 points) Consider a series $\beta(s) > 0, s \in [0,1]$, and construct the parameters

$$\alpha_t = e^{-\frac{1}{2}T(1-t)} \ , \quad \sigma_t = (1-\alpha_t^2) \ , \quad T(t) = \int_0^t \beta(s) ds \tag{5}$$

i. (6 points) Draw plots of α_t, σ_t separately over $t \in [0,1]$ for $\beta(s) = 598, s \in [0,1]$.

ii. (4 points) Does the resulting α_t, σ_t satisfy the conditions in (4)? Clearly justify your answer.

(e) (10 points) This question considers Brownian bridge and Schrödinger bridge in the context of generative modeling.

i. (5 points) What is the main difference between a Brownian bridge and a Schrödinger bridge in the context of generative modeling? Please explain the difference using suitable mathematical notation.

ii. (5 points) Which model would you choose for faster inference time? Clearly justify your answer.

[4] We are considering quadratic cost optimal transport, as discussed in class.