Homework week 7

Problem 1 Explain the concept of score matching and describe how it is used in score-based generative models.

We aim to learn a probability density function (pdf) $p_\theta(x) \geq 0$ with $\int p_\theta(x)\,\mathrm{d}x = 1$. Enforcing the normalization can be intractable for energy-based models, so a standard alternative is to learn the score function, namely,

$$S(x;\theta) = \nabla_x \log p_\theta(x). \tag{0.1}$$

In this case, the loss function (i.e., Explicit score matching (ESM)) is given by

$$\text{Loss}_{ESM}(\theta) = \mathbb{E}_{p(x)}\left[\|S(x;\theta) - \nabla_x \log p_\theta(x)\|^2\right]. \tag{0.2}$$

Since $\nabla_x \log p_\theta(x)$ is unknown, we instead use the implicit score matching (ISM) objective obtained by using integration by parts:

$$\text{Loss}_{ISM}(\theta) = \mathbb{E}_{p(x)}\left[\|S(x;\theta)\|^2 + 2\text{div}S(x;\theta)\right]. \tag{0.3}$$

Note that $\text{Loss}_{ESM}(\theta) = \text{Loss}_{ISM}(\theta) + \mathbb{E}_{p(x)}\left[\|\nabla \log p_\theta(x)\|^2\right]$ and $\mathbb{E}_{p(x)}\left[\|\nabla \log p_\theta(x)\|^2\right]$ doesn't depend on θ, so $\text{Loss}_{ESM}(\theta)$ and $\text{Loss}_{ISM}(\theta)$ share the same minimizer.

To avoid divergence term computations, denoising score matching (DSM) learns scores on noise-corrupted data. Given the original data set $\mathbb{X}_0$ and the data distribution of original data $p_0(x_0)$. We consider the noisy data set $\mathbb{X}$ and the corresponding noisy data distribution $p_\sigma(x)$ as

$$p_\sigma(x) = \int_{\mathbb{R}^d} p(x|x_0)p_0(x_0)\,\mathrm{d}x_0, \qquad x_0 \in \mathbb{X}_0. \tag{0.4}$$

The loss function of DSM can be written as

$$L_{DSM}(\theta) = \mathbb{E}_{p_0(x_0)}\mathbb{E}_{p(x|x_0)}\left[\|S_\sigma(x;\theta) - \nabla \log p(x|x_0)\|^2\right]. \tag{0.5}$$

In particular, if we consider the Gaussian case, then $p_\sigma(x|x_0) = \mathcal{N}(x_0, \sigma^2 I)$ and $\nabla \log p_\sigma(x|x_0) = (x_0 - x)/\sigma^2$.

Problem 2

Question