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

1

Score matching is a method to estimate a probability distribution by matching its score function to that of the data distribution.

The score function of a probability distribution $p(x)$ is the gradient of the log-probability: $\nabla_x log\, p(x)$.
Given data from an unknown true distribution $p_{data}(x)$, and a model $s_\theta(x) \approx \nabla_x log\, p_\theta(x)$.

$$\mathcal{L}_{\text{SM}}(\theta) = \frac{1}{2}\,\mathbb{E}_{p_{\text{data}}}\left[\left\|s_\theta(x) - \nabla_x \log p_{\text{data}}(x)\right\|^2\right]$$

- Forward (diffusion) process: A known process $q(x_t|x_0)$corrupts the data $x_0\ p_{data}$ with noise over time $t \in [0, T]$, typically by adding Gaussian noise.
- Reverse process: The goal is to learn a model that can reverse this diffusion — turning noise back into data.

Denoising Score Matching:
To train the score network $s_\theta(x_t, t)$, we use denoising score matching, which is more suitable when the noise is added in a known fashion.

DSM loss:

$$\mathcal{L}_{\text{DSM}} = \mathbb{E}_{x_0 \sim p_{\text{data}},\, t,\, x_t \sim q(x_t|x_0)}\left[\lambda(t)\left\|s_\theta(x_t, t) + \frac{1}{\sigma_t^2}(x_t - x_0)\right\|^2\right]$$

The reverse-time stochastic differential equation can be approximated using Langevin dynamics, which iteratively samples:

$$x_{t-\Delta t} = x_t + \epsilon\, s_\theta(x_t, t) + \sqrt{2\epsilon}\, z, \quad z \sim \mathcal{N}(0, I)$$