Score Matching and Score-Based (Diffusion) Generative Models

1. Basic Concept of Score Matching

Definition:

Score matching is an unsupervised learning method designed to estimate the **score function** of a data distribution, i.e., the gradient of the log-density:

$$s(x) = \nabla_x \log \, p_{\text{data}}(x)$$

Challenge:

Directly estimating the log-density gradient is often infeasible because the exact likelihood of the data is unknown. Score matching addresses this by minimizing the following objective:

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

Implementation:

To avoid directly computing $\nabla_x \log \, p_{\text{data}}(x)$, Hyvärinen introduced a technique using a contrastive loss, making score matching practical for unsupervised learning.

---

2. Score-Based (Diffusion) Generative Models

Core Idea:

Score-based generative models use score matching to gradually transform data from its true distribution to a simple Gaussian noise distribution, while learning how to reconstruct data from noise. This process can be viewed as **denoising**.

Process Overview:

1. **Forward Process:**
   Start with the data distribution $p_{\text{data}}(x)$and progressively add noise, ultimately reaching a simple Gaussian distribution $p_T(x)$.
2. **Reverse Process:**
   Learn how to generate data from the noisy distribution $p_T(x)$back to

$p_{\text{data}}(x)$using the learned score function.

Training Objective:

Using score matching, the model minimizes:

$$\mathcal{L}(\theta) = \mathbb{E}_{x \sim p_{\text{data}}}[\| s_\theta(x) - \nabla_x \log p_{\text{data}}(x) \|^2]$$

This allows the model to learn the gradient of the data distribution and enables generation from noise.

Generation Process:

Start with a sample from a Gaussian noise distribution, then iteratively denoise using the learned score function until the final sample matches the target data distribution.

---

3. Applications of Score Matching in Score-Based Generative Models

Advantages:

- **No explicit likelihood required:**
  Score matching does not rely on the explicit log-likelihood of the data distribution, allowing flexible training.
- **High-quality sample generation:**
  Learning the data distribution's gradient enables high-quality and diverse sample generation comparable to GANs.
- **Precise likelihood estimation:**
  With the learned score function, exact likelihoods can be computed for model evaluation.

Use Cases:

- **Image generation:**
  Generate high-resolution, diverse images.
- **Conditional generation:**
  Generate samples conditioned on inputs, such as image-to-image

translation.

- **Inverse problem solving:**
  Recover true data from incomplete or noisy observations using the learned score function.