ML Programming assignment VII

Name : 林欣妤
Student number : 314652023

October 21, 2025

1 Introduction

Score matching is a technique for training probabilistic models when the normalization constant of the density is intractable. Instead of learning the density directly, score matching focuses on learning the gradient of the log-density, known as the *score function*. This method has become particularly important in the context of score-based generative models, also called diffusion models.

2 Score Matching

2.1 Definition

Let $x \sim p_{\text{data}}(x)$ be the data distribution, and $p_\theta(x)$ be a parametric model:

$$p_\theta(x) = \frac{\tilde{p}_\theta(x)}{Z_\theta}, \quad Z_\theta = \int \tilde{p}_\theta(x) dx$$

where Z_θ may be intractable. The **score function** of a density $p(x)$ is defined as

$$s_p(x) = \nabla_x \log p(x).$$

Instead of matching densities directly, score matching aims to match the model score $s_\theta(x)$ with the data score $s_{\text{data}}(x)$:

$$s_\theta(x) = \nabla_x \log p_\theta(x) \approx s_{\text{data}}(x) = \nabla_x \log p_{\text{data}}(x).$$

2.2 Score Matching Objective

The original Hyvärinen score matching loss is

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

Using integration by parts, this can be rewritten without requiring $s_{\text{data}}(x)$:

$$\mathcal{L}_{\text{SM}}(\theta) = \mathbb{E}_{x\sim p_{\text{data}}}\Big[\text{Tr}(\nabla_x s_\theta(x)) + \frac{1}{2}\|s_\theta(x)\|^2\Big],$$

where $\text{Tr}(\nabla_x s_\theta(x))$ is the divergence of the score function. This makes it practical for training unnormalized models.

3 Score-Based (Diffusion) Generative Models

3.1 Forward Diffusion Process

Diffusion models define a forward process that gradually adds noise to data:

$$x_t = \sqrt{\alpha_t} x_0 + \sqrt{1-\alpha_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, I),$$

where t indicates the noise level. At the final time step T, x_T is nearly Gaussian.

3.2 Reverse Process and Score Function

The reverse process aims to generate samples from the data distribution by using the score function of the noisy data:

$$dx = \left[f_\theta(x,t) - \frac{1}{2} g_t^2 \nabla_x \log p_t(x) \right] dt + g_t d\bar{W}_t,$$

where $\nabla_x \log p_t(x)$ is the score at time t.

By learning a neural network $s_\theta(x,t) \approx \nabla_x \log p_t(x)$, we can simulate this reverse process to generate realistic samples.

3.3 Training via Denoising Score Matching

In practice, the score network is trained using denoising score matching:

$$\mathcal{L}(\theta) = \mathbb{E}_{x_0 \sim p_{\text{data}}, \epsilon \sim \mathcal{N}(0,I), t} \left[\lambda(t) \| s_\theta(x_t, t) - \nabla_{x_t} \log p(x_t | x_0) \|^2 \right],$$

where $p(x_t|x_0)$ is known from the forward diffusion process, allowing computation of the target score.

4 Summary

- Score matching enables learning unnormalized densities by matching gradients of log-densities rather than the densities themselves.
- Diffusion models gradually add noise to data and learn the score function of the noisy data at each noise level.
- Once the score is learned, the reverse process guided by the score function can generate realistic samples from the learned data distribution.

References

[1] Chat-GPT (Apply GPT to revise and correct the English content of the report, and ask about some programming techniques)

[2] Y. Song, S. Garg J. X. Shi and S. Ermon, *Sliced Score Matching: A Scalable Approach to Density and Score Estimation*, Machine Learning, arXiv (2019)