WrittenAssignment_111652026劉馥瑞

1.

In score matching, we want to estimate a model score function $S(x;\theta) = \nabla_x \log \tilde{p}(x;\theta)$ that approximates the data score $\nabla_x \log p(x)$ without needing the normalization constant of p(x). The denoising or Hyvärinen score matching objective is:

$L\_SM = E_{x \sim p(x)} [\frac{1}{2}\|S(x;\theta)\|^2 + \nabla_x \cdot S(x;\theta)]$.

But in high dimensions, computing the divergence term $\nabla_x \cdot S(x;\theta) = \Sigma_i \partial S_i/\partial x_i$ can be expensive. To reduce this cost, Sliced Score Matching (Song et al., 2020) randomizes the direction of differentiation.

We randomly draw a direction $v \in \mathbb{R}^d$ from an isotropic distribution p(v), typically $v \sim N(0, I)$. The Sliced Score Matching loss is defined as:

$L\_SSM = E_{x \sim p(x)} E_{v \sim p(v)} [\frac{1}{2} (v^T S(x;\theta))^2 + v^T \nabla_x (v^T S(x;\theta))]$.

Multiplying by 2 (or equivalently removing ½ and adjusting constants) gives:

$L\_SSM = E_{x \sim p(x)} E_{v \sim p(v)} [\|v^T S(x;\theta)\|^2 + 2v^T \nabla_x (v^T S(x;\theta))]$.

This matches the given expression. The difference is just a scaling convention.

- $v^T S(x;\theta)$: projection of the score vector onto direction v.
- $v^T \nabla_x (v^T S(x;\theta))$: directional derivative of that projection along v.

Averaging over random v gives an unbiased stochastic estimate of the full divergence term.

2.

An SDE describes how a random variable evolves over continuous time with both deterministic and stochastic effects:

$dx = f(x, t)\, dt + g(x, t)\, dW\_t$

where f(x, t) is the drift (deterministic) term, g(x, t) is the diffusion (noise) term, and dW_t is the Wiener process (Gaussian noise increment). SDEs are key to diffusion models, describing how data is gradually noised forward and denoised backward using learned scores.

3.

How does the Euler–Maruyama method extend the standard Euler method used for ordinary differential equations (ODEs)?