ML_AS8 111652042 顏友君

1. Show that the sliced score matching (SSM) loss can also be written as

$$L_{SSM} = \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}\left[\|v^T S(x;\theta)\|^2 + 2v^T\nabla_x(v^T S(x;\theta))\right].$$

From week 7-2 notes, we have

$$L_{SSM}(\theta) = \mathbb{E}_{x\sim p(x)}\|S(x;\theta)\|^2 + \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}\left[2v^T\nabla_x(v^T S(x;\theta))\right] \quad \text{— ①}$$

We want to show that SSM loss can also be written as

$$L_{SSM}(\theta) = \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}\left[\|v^T S(x;\theta)\|^2 + 2v^T\nabla_x(v^T S(x;\theta))\right]$$

where $\mathbb{E}_{x\sim p(x)}$ is the expectation over data sample x from $p(x)$,

$\mathbb{E}_{v\sim p(v)}$ is the expectation over random projection directions v, $S(x;\theta)$ is the score function.

pf. Let $v\sim N(0,I)$. We have $\mathbb{E}[vv^T] = I$. (Proof also given in week 7-2 notes)

Then for any vector $s\in\mathbb{R}^D$, we have

$$\mathbb{E}[(v^Ts)^2] = \mathbb{E}[s^Tvv^Ts] = s^T\mathbb{E}[vv^T]s = s^TIs = \|s\|^2$$

Substituting $\mathbb{E}_{v\sim p(v)}[(v^TS(x;\theta))^2] = \|S(x;\theta)\|^2$ into ①,

$$\begin{aligned}
L_{SSM}(\theta) &= \mathbb{E}_{x\sim p(x)}\|S(x;\theta)\|^2 + \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}\left[2v^T\nabla_x(v^TS(x;\theta))\right]\\
&= \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}\left[(v^TS(x;\theta))^2\right] + \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}\left[2v^T\nabla_x(v^TS(x;\theta))\right]\\
&= \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}\left[(v^TS(x;\theta))^2 + 2v^T\nabla_x(v^TS(x;\theta))\right]\\
&= \mathbb{E}_{x\sim p(x)}\mathbb{E}_{v\sim p(v)}\left[\|v^TS(x;\theta)\|^2 + 2v^T\nabla_x(v^TS(x;\theta))\right]. \quad \square
\end{aligned}$$

2. Briefly explain SDE.

SDE (Stochastic Differential Equation [隨機微分方程]) is used to describe how data gradually turn into noise (Forward) and vice versa (Reverse).

The forward SDE adds noise to data over time, defined as

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

where $f(x,t)$ is drift coefficient, $g(t)$ is diffusion coefficient, w is standard Wiener process.

The reverse SDE is to get the reverse-time SDE.

$$dx = [f(x,t) - g^2(t)\nabla_x \log p_t(x)]\,dt + g(t)\,d\bar{w},$$ where $p_t(x)$ is the probability density of $x(t)$,

which removes noise step by step to generate data from random noise.

The key idea is that if we can estimate the score function $\nabla_x \log p_t(x)$ using a neural network, we can simulate this reverse SDE to generate realistic samples.

In short, SDE provide a continuous time framework that unifies previous score-based and diffusion generative models.

3. Unanswered question.

如果在 SDE 中的 drift 及 diffusion 常數改成由模型學習而不是事先固定，

會有更好的表現嗎？還是說反而會破壞掉現有 SDE 的理論架構？