Machine Learning

313704071_HW8

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

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

We know that $\nabla \cdot S(x;\theta) = trace(\nabla S)$.

Hutchinson's trace estimator:

Let $v \in \mathbb{R}^d$ is a random vector, such that $E[vv^T] = 1$, then

$$tr(A) = \mathbb{E}_{v\sim p(v)}(v^T A v)$$

We also know that

$L_{ISM}(\theta) = \mathbb{E}_{x\sim p(x)}[\|S(x;\theta)\|^2 + 2\nabla_x \cdot S(x;\theta)]$

Hence, the $\nabla_x \cdot S(x;\theta)$ in ISM loss can be rewritten as

$$\nabla_x \cdot S(x;\theta) = trace(\nabla S) = \mathbb{E}_{v\sim p(v)}(v^T(\nabla S)v) = \mathbb{E}_{v\sim p(v)}(v^T\nabla(vS))$$

Then, combine it with ISM loss and become SSM loss.

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

2. Briefly explain SDE

The Stochastic Differential Equation:

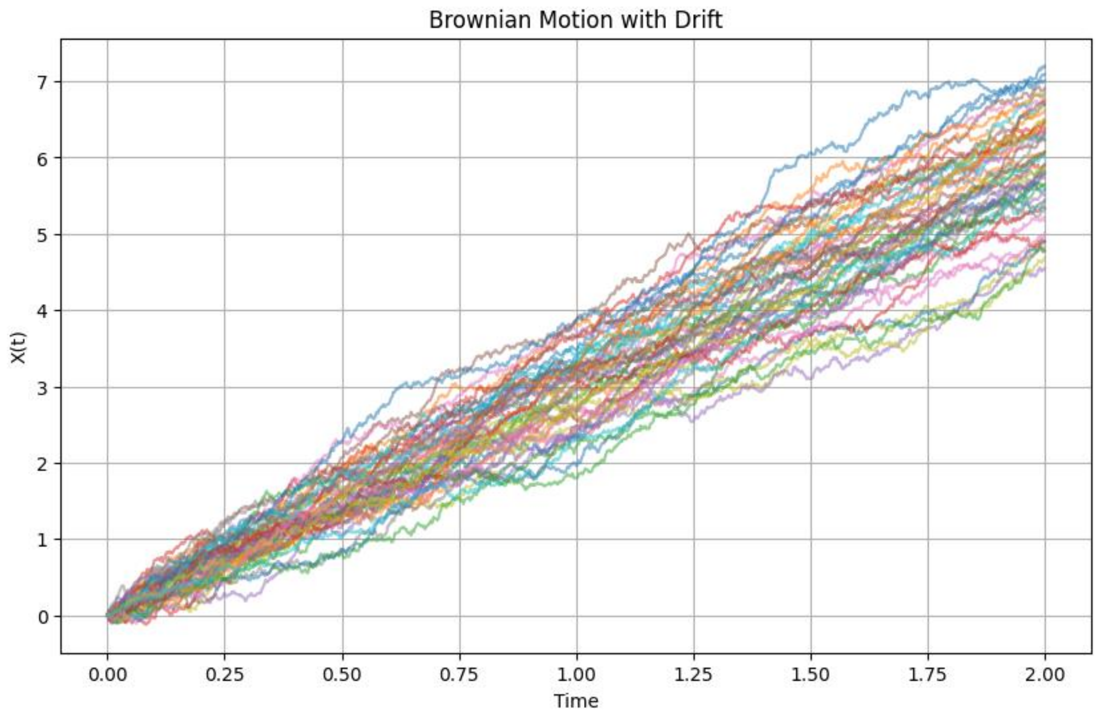
$$dX_t = f(X_t,t)dt + G(X_t,t)dW_t$$
$$X(0) = X_0$$

where X_t is stochastic process, W_t is standard Brownian motion.

The forward SDE is a process of adding noise, and drift part $f(X_t,t)$ control the average trend, on the other hand, diffusion part $G(X_t,t)$ control the strength of noise.

The figure shows sample paths of a Brownian motion with drift. For any fixed time t, the random variable $X(t)$ is normally distributed with mean μt and variance $\sigma^2 t$.

The empirical histogram across paths at time t approximates this PDF.



3. Unanswered Questions

(1)為什麼需要 $f(X_t,t)$?

GPT: 沒有 f 就沒有系統性的趨勢；只會靠雜訊擴散、均值不動、難以把系統帶往指定區域。

(2)加噪的起點可以不同?

GPT: 影像加噪（同一張圖內）：每個像素/通道的起點就是它自己的像素值 $x_0[i]$（彼此通常不同），然後各自加上獨立高斯噪聲沿時間演化。所以起點不同是正常的。