Machine Learning

Assignment 8

CHANG Yung-Hsuan (張永璿)

111652004

eiken.sc11@nycu.edu.tw

October 27, 2025

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

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

Proof. By definition given by the instructor,

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

It suffices to show

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

For any fixed x (hence fixed $S := S(x; \theta) \in \mathbb{R}^d$),

$$\mathbb{E}_{v \sim p(v)} \left\| v^T S \right\|^2 = \mathbb{E}_{v \sim p(v)} \left(\left(v^T S \right)^T \left(v^T S \right) \right)$$

$$= \mathbb{E}_{v \sim p(v)} \left(S^T \left(v v^T \right) S \right)$$

$$= S^T \mathbb{E}_{v \sim p(v)} \left(v v^T \right) S$$

$$= S^T I S$$

$$= \| S \|^2.$$

Note that v is a random vector satisfying $\mathbb{E}(vv^t) = I$, so the fourth equation stands. Thus, taking

$\mathbb{E}_{x \sim p(x)}$ on both sides yields

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

Substituting back to the definition, in order to replace $\mathbb{E}_{x \sim p(x)} \| S(x; \theta) \|^2$, yields

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

2. Briefly explain what a stochastic differential equation is.

   Explanation. A stochastic differential equation extends an ordinary differential equation (deterministic) by introducing noise (random) through a stochastic term. Formally, it is written as

   $$dX_t = f(X_t, t) \, dt + G(X_t, t) \, dW_t,$$

   where $f(X_t, t) \, dt$ is the drift (deterministic **trend**) and $G(X_t, t) \, dW_t$ is the diffusion (random **fluctuation**) driven by Brownian motion W_t.

   Intuitively, a stochastic differential equation describes how a system evolves under both deterministic dynamics and random noise, making it the stochastic counterpart of an ordinary differential equation.

3. (*Old question 2.*) Suppose $f(t)$ is a continuous function and W_t is the Brownian motion. Show that $\int_0^t f(s) \, dW_s$ is a Gaussian process with zero mean and variance $\int_0^t f(s)^2 \, ds$.

   Proof. Define $I_t := \int_0^t f(s) \, dW_s$ for convenience in shorthand. Let $0 = t_0 < t_1 < \ldots < t_n = t$ be a partition. From the notes, the stochastic integral is the limit (as $n \to \infty$) of

   $$I_{t,n} := \sum_{k=0}^{n-1} f(t_k) \left(W_{t_{k+1}} - W_{t_k} \right).$$

   Note that $W_{t_{k+1}} - W_{t_k}$ are independent and

   $$W_{t_{k+1}} - W_{t_k} \sim \mathcal{N}(0, t_{k+1} - t_k).$$

   Each partial sum is Gaussian with a known mean and a known variance; thus, $I_{t,n}$ is a deterministic linear combination of independent Gaussian variables, hence Gaussian. Moreover,

   $$\mathbb{E}(I_{t,n}) = \sum_{k=0}^{n-1} f(t_k) \mathbb{E} \left[W_{t_{k+1}} - W_{t_k} \right]$$

$$= \sum_{k=0}^{n-1} f(t_k)0$$

$$= 0,$$

and

$$\mathbb{V}(I_{t,n}) = \sum_{k=0}^{n-1} f(t_k)^2 \mathbb{V}\left(W_{t_{k+1}} - W_{t_k}\right)$$

$$= \sum_{k=0}^{n-1} f(t_k)^2 (t_{k+1} - t_k).$$

By continuity of f, the Riemann sums converge

$$\sum_{k=0}^{n-1} f(t_k)^2 (t_{k+1} - t_k) = \int_0^t f(s)^2 \, ds.$$

Therefore,

$$\mathbb{E}(I_t) = \lim_{n\to\infty} \mathbb{E}(I_{t,n}) = 0, \quad \text{and} \quad \mathbb{V}(I_t) = \lim_{n\to\infty} \mathbb{V}(I_{t,n}) = \int_0^t f(s)^2 \, ds.$$

To show a process is Gaussian, it suffices to show that each X_t is a linear combination (or the limit of a linear combinations) of independent Gaussian variables, since Gaussianity is preserved under linear transformations and limits.

In our case, $I_t = \int_0^t f(s) \, dW_s$ is the limit of $\sum_{k=0}^{n-1} f(t_k)\left(W_{t_{k+1}} - W_{t_k}\right)$, a linear combination of independent Gaussian increments of Brownian motion. Hence $\{I_t\}$ is a Gaussian process.

4. There are unanswered questions from the lecture, and there are likely more questions we haven't covered. Take a moment to think about these questions. Write down the ones you find important, confusing, or interesting.

   Answer. In class, it was mentioned that sampling x_0 from p would satisfy the Fokker–Planck equation. I hope to gain a clearer understanding next week when we see the proof. For now, I wonder what kinds of probability densities actually satisfy a given Fokker–Planck equation.

   To be more precise, for which drift f and diffusion G will the solution $p(x,t)$ of

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{\sigma^2}{2} \cdot p_{xx}$$

converge to a Gaussian distribution as $t \to \infty$?

In some special cases, such as pure diffusion or the Ornstein–Uhlenbeck process, the solution indeed converges to a normal distribution, as mentioned in class. I am curious whether we can characterize all families of solutions that remain invariant or converge to a stationary Gaussian distribution.

5. Some part of conversation with ChatGPT for me to get better understanding about the lecture.

   a. *Me*: What does it mean by saying that "the p of the sampling of x_0 satisfies the Fokker-Planck equation?"

   ChatGPT: It does not refer to an individual sample x_0 but to its **distribution** $p(x, 0)$, which serves as the initial condition of the Fokker-Planck equation. When we say "sampling x_0 from p," it means that the initial random variable X_0 is drawn from the (assumed) density $p(x, 0)$. Once the process evolves according to the stochastic differential equation, the time-dependent density $p(x, t)$ of X_t will satisfy the Fokker-Planck equation. Thus, it is the **distribution** $p(x, t)$, not a single sample, that satisfies the equation.

   b. *Me*: So it's the $p(x, t)$ satisfies the Fokker–Planck equation; but where is it in the stochastic differential equation?

   ChatGPT: The stochastic differential equation itself only describes the dynamics of **samples** X_t, not the explicit form of their distribution. However, for each time t, the random variable X_t has a density $p(x, t) = P(X_t = x)$. This density is implicitly defined by the stochastic differential equation through the way X_t evolves. By applying Itô's lemma to a smooth test function $\varphi(x)$ and taking expectations, we can derive how $p(x, t)$ must evolve over time; this yields the Fokker-Planck equation. In short, $p(x, t)$ is "hidden" inside the stochastic differential equation as the probability distribution of the process generated by that stochastic differential equation.