Machine Learning
Assignment 8

CHANG Yung-Hsuan (張永璿)
111652004
eiken.sc11@nycu.edu.tw

October 22, 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,

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

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 $S \coloneqq S(x;\theta) \in \mathbb{R}^d$),

$$\begin{aligned}
\mathbb{E}_{v\sim p(v)}\left\|v^T S\right\|^2 &= \mathbb{E}_{v\sim p(v)}\left(\left(v^T S\right)^T\left(v^T S\right)\right) \\
&= \mathbb{E}_{v\sim p(v)}\left(S^T\left(vv^T\right)S\right) \\
&= S^T\mathbb{E}_{v\sim p(v)}\left(vv^T\right)S \\
&= S^T I_d S \\
&= \|S\|^2.
\end{aligned}$$

Note that v is a random vector satisfying $\mathbb{E}\left(vv^t\right) = I$, 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

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

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

   $$\mathrm{d}X_t = f(X_t,t)\,\mathrm{d}t + G(X_t,t)\,\mathrm{d}W_t,$$

   where $f(X_t,t)\,\mathrm{d}t$ is the drift (deterministic **trend**) and $G(X_t,t)\,\mathrm{d}W_t$ 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 $\int_0^t f(s)\,\mathrm{d}W_s$ is a Gaussian process with zero mean and variance $\int_0^t f(s)^2\,\mathrm{d}s$.

   Proof. Define $I_t := \int_0^t f(s)\,\mathrm{d}W_s$ 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,

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

$$
\begin{aligned}
&= \sum_{k=0}^{n-1} f(t_k)0 \\
&= 0,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{V}\big(I_{t,n}\big) &= \sum_{k=0}^{n-1} f(t_k)^2 \mathbb{V}\Big(W_{t_{k+1}} - W_{t_k}\Big) \\
&= \sum_{k=0}^{n-1} f(t_k)^2 (t_{k+1} - t_k).
\end{aligned}
$$

By continuity of f, the Riemann sums converge

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

Therefore,

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

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, $I_t = \displaystyle\int_0^t f(s) \, \mathrm{d}W_s$ is the limit of $\displaystyle\sum_{k=0}^{n-1} f(t_k)\Big(W_{t_{k+1}} - W_{t_k}\Big)$, 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.