Machine Learning Week 8 Assignment

Question 1

> Show that the sliced score matching (SSM) loss can also be written as
>
> $$L_{SSM} = E_{x \sim p(x)} E_{v \sim p(v)} \left[\|v^T S(x;\theta)\|^2 + 2v^T \nabla_x (v^T S(x;\theta)) \right]$$

We have already have the form as following from the class

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

Rewrite the term $\|S(x;\theta)\|^2$ as following:

$$\begin{aligned} \|S(x;\theta)\|^2 &= S^T S \\ &= S^T I S \\ &= S^T E_{v \sim p(v)}[vv^T] S \\ &= E_{v \sim p(v)}[S^T v v^T S] \\ &= E_{v \sim p(v)} \|v^T S(x;\theta)\|^2 \end{aligned}$$

Plug in then we have the desired result.

Question 2

> Briefly explain SDE.

我們都知道一個簡單的微分方程的形式：

$$\frac{dy}{dt} = f(t, y)$$

也可以被寫成

$$dx_t = f(x_t, t) dt$$

這樣寫更強調了 x（位置）與 t（時間）的關係。
當初始值被給定之後，這個微分方程的解給我們一個函數，也就是我們從微分後的形式就能推敲出原函數的走向。

而 SDE 的想法是在原本確定的函數走向上加一個常態分布的擾動項，寫成

$$\begin{cases} dx_t = f(x_t, t)dt + G(x_t, t)dW_t \\ x(0) = x_0 \end{cases}$$

其中 G 可視為擴散項，控制這個擾動的強度。而 dW_t 則是這個擾動的來源，就是 Wiener process，它滿足 $E[W_t] = 0$ 及 $\text{Var}[W_t] = t$。

也就是說，仍然可以想像 x 基本上會照著 f 提供的方向去走，但在其中再加上由 G 控制強度的擾動項 W_t。

並且我們可以將上式寫成積分的形式如下：

$$x_t = x_0 + \int_0^t f(x_s, s)ds + \int_0^t G(x_s, s)dW_s$$

稱為 Itô's ingegral form.

以下看一個例子：

> Solve the SDE:
>
> $$\begin{cases} dX_t = \mu dt + \sigma dW_t \\ X(0) = X_0 \end{cases}$$

重新寫成 Itô's integral form，我們可以得到

$$X_t = X_0 + \int_0^t \mu ds + \int_0^t \sigma dW_s = X_0 + \mu t + \sigma W(t)$$

也就是說，X_t 大致上依照 $X_0 + \mu t$ 的軌跡走（也就是一條斜率為正的直線，若 $\mu > 0$），但實際是像粉色的軌跡走，因為有 σ 強度的常態分布擾動。