1. Explain the concept of score matching and describe how it is used in score-based (diffusion) generative models.

給定 data $x \in \mathbb{R}^n$，我們想知道 x 是從哪個 pdf $p(x)$ 的 distribution 抽出的

定義 score function 為 $S(x) := \nabla_x \log p(x)$，$(S: \mathbb{R}^n \to \mathbb{R}^n)$

向量 $S(x)$ 表示在 x 點，往哪個方向前進，pdf $p(x)$ 會上升得最快，

i.e., 描述往分佈中更合理、常見的樣本前進

而 score matching 是描述學習 score function 的方法，至少有以下3種：

① Explicit score matching (ESM)，其 loss function 為

$$L_{ESM}(\theta) := E_{x\sim p(x)} \| S(x;\theta) - \nabla_x \log p(x) \|^2$$

② Implicit score matching (ISM)，其 loss function 為

$$L_{ISM}(\theta) := E_{x\sim p(x)} \left[\| S(x;\theta) \|^2 + 2 \nabla_x \cdot S(x;\theta) \right]$$

③ Denoising score matching (DSM)，其 loss function 為

$$L_{DSM}(\theta) := E_{x_0\sim p(x_0)} E_{(x|x_0)\sim p(x|x_0)} \| S_\sigma(x;\theta) - \nabla_x \log P(x|x_0) \|^2$$

where x_0 是 original data

$P_0(x_0)$ 是 original pdf

x 是 perturbed data，and $x = x_0 + \varepsilon_\sigma$，$\varepsilon_\sigma \sim N(0, \sigma^2 I)$

$P_\sigma(x)$ 是 pdf of perturbed pdf

在 score-based (diffusion) generative model 中，是利用 Denoising score matching 來學習 $P_\sigma(x)$ 的 score function，

$\because$ $x = x_0 + \varepsilon_\sigma$，$\varepsilon_\sigma \sim N(0, \sigma^2 I)$

$\therefore$ $P(x|x_0) = \dfrac{1}{(2\pi)^{\frac{n}{2}} \sigma^d} e^{-\frac{\|x-x_0\|^2}{2\sigma^2}}$

$\Rightarrow \log P(x|x_0) = \log \dfrac{1}{(2\pi)^{\frac{n}{2}} \sigma^d} - \dfrac{\|x-x_0\|^2}{2\sigma^2}$

$\Rightarrow \nabla_x \log P(x|x_0) = -\dfrac{1}{\sigma^2}(x - x_0)$

So,

$$
\begin{aligned}
L_{DSM}(\theta) &:= E_{x_0 \sim p(x_0)} E_{(x|x_0) \sim p(x|x_0)} \| S_\sigma(x;\theta) - \nabla_x \log p(x|x_0) \|^2 \\
&= E_{x_0 \sim p(x_0)} E_{(x|x_0) \sim p(x|x_0)} \| S_\sigma(x;\theta) + \frac{1}{\sigma^2}(x - x_0) \|^2 \\
&= E_{x_0 \sim p(x_0)} E_{(x|x_0) \sim p(x|x_0)} \frac{1}{\sigma^4} \| \sigma^2 S_\sigma(x;\theta) + x - x_0 \|^2
\end{aligned}
$$

$x = x_0 + \varepsilon_\sigma$, $\varepsilon_\sigma \sim N(0, \sigma^2 I)$

$\Updownarrow$

On the other hand, 若給定 data x_0, 加噪得到 $x = x_0 + \sigma\varepsilon$, $\varepsilon \sim N(0, I)$ 想學習一個 function h_θ 嘗試復原乾淨的影像，則可利用 Denoising auto encoder，其 loss function 為

$$
L_{DAE} := E_{x_0 \sim p(x_0)} E_{(x|x_0) \sim p(x|x_0)} \| h_\theta(x) - x_0 \|^2
$$

i.e.,

$$
\overset{x}{\boxed{\text{noising}}} \xrightarrow{h_\theta(x)} \overset{x_0}{\boxed{\text{clean}}}
$$

$h_\theta(x)$ 稱為去噪器

在 Vincent 2011年的 paper (A connection between score matching and denoising autoencoders) 中證明了若在 Denoising auto encoder 的去噪器是最佳的（在給定噪音 σ 下），即 $h_\theta(x)$ 預測 $E[x_0|x]$，則此去噪器 $h_\theta(x)$ 與 score function $S(x)$ 的關係為

$$
h_\theta(x) = x + \sigma^2 S_\sigma(x;\theta)
$$

$\therefore$ $L_{DSM} = C L_{DAE}$, where C is a constant

也就是說 DSM 是用來學髒的圖的 score function，DAE 是在學將髒的圖復原成乾淨的圖中的去噪 (mapping)，而這二件事是等價的！

2. Unanswered Questions

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.
- You do **not** need to answer them—just state them clearly.

在上課中有推導了 $\arg\min_\theta L_{ESM} = \arg\min_\theta L_{ISM} = \arg\min_\theta L_{DSM}$

Q：這些推導要如何嚴格、清楚的寫下？

其中會有什麼假設或是用到什麼分析上的理論？