

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 $S(x) := \nabla_x \log p(x)$, ($S: \mathbb{R}^n \rightarrow \mathbb{R}^n$)

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

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

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

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

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

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

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,

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

$$\because p(x|x_0) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{\|x-x_0\|^2}{2\sigma^2}}$$

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

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

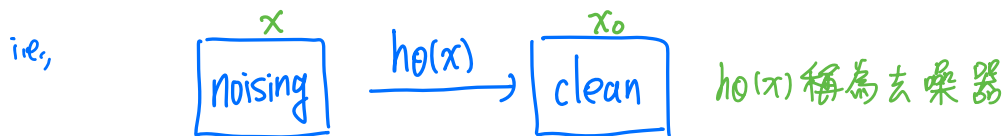
$$\begin{aligned}
S_0, \quad L_{DSM}(\theta) &:= E_{x_0 \sim p(x_0)} E_{(x|x_0) \sim p(x|x_0)} \| S_\sigma(x; \theta) - \nabla_x \log P(x|x_0) \|^2 \\
&= E_{x_0 \sim p(x_0)} E_{(x|x_0) \sim p(x|x_0)} \| S_\sigma(x; \theta) + \frac{1}{\sigma^2} (x - x_0) \|^2 \\
&= E_{x_0 \sim p(x_0)} E_{(x|x_0) \sim p(x|x_0)} \frac{1}{\sigma^4} \| \sigma^2 S_\sigma(x; \theta) + x - x_0 \|^2
\end{aligned}$$

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



On the other hand, 若給定 data x_0 , 加噪得到 $x = x_0 + \sigma \Sigma$, $\Sigma \sim N(0, I)$ 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

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



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

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

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

也就是說 DSM 是用來學習圖片的 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.

在上课中有 推导了 $\operatorname{argmin}_{\theta} L_{ESM} = \operatorname{argmin}_{\theta} L_{ISM} = \operatorname{argmin}_{\theta} L_{DSM}$

Q: 这些推导要如何严格、清楚的写下?

其中会有什麼假设或是用到什麼分析上的理论?