1

Score matching: to learn data's probability density function

But $p(x;\theta)=\frac{e^{q_\theta(x;\theta)}}{Z(\theta)}$ is difficult.

to learn

Instead, we try to learn $S(x)=\nabla_x \ln p(x;\theta)=\nabla_x \ln q_\theta(x;\theta)$

methods:

1. Explict Score ~~matt~~ matching (ESM)

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

we usually unknown $\nabla_x \ln p(x)$, so actually we can't use it.

2. Implicit Score matching (ISM)

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

where $\nabla_x \cdot S(x;\theta)=tr(\nabla_x S(x;\theta))$

but $tr(\nabla_x S(x;\theta))$ is expensive

compute

3. Denoising Score matching (DSM)

$L_{DSM}(\theta)=E_{x_0\sim p_0(x_0)}E_{x_1|x_0\sim p(x_1|x_0)}[||S_\theta(x;\theta)-\nabla_x \ln p(x_1|x_0)||^2$

$x = x_0 + \epsilon_\sigma \Rightarrow \nabla_x \ln p(x|x_0)$ has explicit solution $-\frac{1}{\sigma^2}(x - x_0)$

but score is noising, but we can training.

4. Sliced Score matching (SSM)

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

any A, it $tr(A) \sim \mathbb{E}_{v\sim p(v)}[v^T A v]$ estimated

$\mathbb{E}[vv^T] = I \Rightarrow O(d^2) \to O(d)$

but variance is more

2.

~~I still can't~~ ⓐ

I like to see some simple example, like 代數好看看，像數學課本那樣！