Scientific Computing in Deep Generative Models

专题 1: Score-based 模型与 Flow 模型简介

主要内容

前向 SDE:

$$dX_t = \mathbf{b}(X_t, t)dt + \sigma(X_t, t)d\mathbf{W}_t$$

逆向 SDE（当 σ 为与 x 无关的标量 $\sigma(t)$ 时）:

$$dX_t = \left[\mathbf{b}(X_t, t) - \sigma(t)^2 \nabla \log p_t(X_t)\right] dt + \sigma(t)d\bar{\mathbf{W}}_t$$

概率 Flow ODE（当 σ 为与 x 无关的标量 $\sigma(t)$ 时）:

$$dX_t = \left[\mathbf{b}(X_t, t) - \frac{1}{2}\sigma(t)^2 \nabla \log p_t(X_t)\right] dt \qquad X_T \sim \mathcal{N}(0, I)$$

Ito 公式:

$$df(X_t, t) = \left(\frac{\partial f}{\partial t} + \nabla f \cdot \mathbf{b} + \frac{1}{2}\mathrm{Tr}\left(\sigma\sigma^\top \nabla^2 f\right)\right) dt + (\nabla f)^\top \sigma d\mathbf{W}_t$$

无穷小生成元 $\mathcal{L}_t$:

$$\mathcal{L}_t f = \sum_i b_i(x, t)\partial_i f + \frac{1}{2}\sum_{i,j}(\sigma\sigma^\top)_{ij}\partial^2_{i,j} f$$

伴随算子 $\mathcal{L}_t^*$:

$$\mathcal{L}_t^* p = -\sum_i \partial_i(b_i p) + \frac{1}{2}\sum_{i,j}\partial^2_{i,j}\left((\sigma\sigma^\top)_{ij} p\right)$$

Hamilton 量:

$$H(p, q) = \frac{1}{2}|p|^2 + V(q)$$

Gibbs 分布:

$$\rho_\beta(p, q) = \frac{1}{Z}e^{-\beta H(p,q)}$$

欠阻尼 Langevin 动力学:

$$\begin{cases} dq_t = p_t dt \\ dp_t = -\nabla V(q_t)dt - \gamma p_t dt + \sqrt{2\gamma\beta^{-1}}dW_t \end{cases}$$

Score Matching:

$$\min_\theta \mathbb{E}\|s_\theta(x) - \nabla \log \pi(x)\|^2$$

Flow Matching:

$$\min_\theta \mathbb{E}_{t\sim U, z\sim p_{data}, x\sim p_t(\cdot|z)}\|u_\theta(x, t) - v(x, t)\|^2$$

Conditional Score Matching:

$$\min_\theta \mathbb{E}_{z\sim p_{data},x=z+\sigma\epsilon}\|s_\theta(x)-\nabla\log\pi(x)\|^2=\min_\theta \mathbb{E}_{z\sim p_{data},x=z+\sigma\epsilon}\|s_\theta(x)+\frac{\epsilon}{\sigma}\|^2$$

DDPM:

$$x_{t+1}=\alpha_t x_t+\beta_t\epsilon$$

$$\min_\theta \mathbb{E}_{t\sim U,z\sim p_{data},\epsilon\sim\mathcal{N}(0,I)}\frac{1}{\beta_t^2}\|\beta_t s_\theta(\alpha_t z+\beta_t\epsilon)+\epsilon\|^2$$

Conditional Flow Matching:

$$\min_\theta \mathbb{E}_{t\sim U,z\sim p_{data},x\sim p_t(\cdot|z)}\|u_\theta(x,t)-v(x,t|z)\|^2$$

条件优化与全局优化的等价性:

$$Loss_{FM}(\theta)=Loss_{CFM}(\theta)+C$$

$$Loss_{SM}(\theta)=Loss_{CSM}(\theta)+C$$

Mode 覆盖观察:

从原点出发的粒子在采样初期倾向于最先覆盖欧氏距离最小的模态

自适应修正分布:

$$dx=-g(x)\nabla V(x)dt+\beta^{-1}\nabla g(x)dt+\sqrt{2\beta^{-1}g(x)}dW$$

ABO 格式:

$$d\begin{pmatrix}x\\p\end{pmatrix}=\underbrace{\begin{pmatrix}p\\0\end{pmatrix}dt}_{\text{A: 位置漂移}}+\underbrace{\begin{pmatrix}0\\-\nabla V(x)\end{pmatrix}dt}_{\text{B: 动量漂移}}+\underbrace{\begin{pmatrix}0\\-\gamma p dt+\sqrt{2\beta^{-1}}dW_t\end{pmatrix}}_{\text{O: Ornstein-Uhlenbeck 过程}}+\underbrace{\begin{pmatrix}0\\\beta^{-1}\nabla\log g(x)\end{pmatrix}dt}_{g(x)\text{ 修正}}$$

根据 $g(x)$ 修正归属不同，以下 B 与 O 步需分别调整：

- **A 步** (Exact Drift): $\Psi_A^h(x,p)=(x+hp,p)$(隐式中点保辛)
- **B 步** (Exact Gradient): $\Psi_B^h(x,p)=(x,p-h\nabla V(x))$
- **O 步** (OU 过程精确解):

$$\Psi_O^h(x,p)=\left(x,e^{-h\gamma}p+\sqrt{\beta^{-1}(1-e^{-2\gamma h})}\mathcal{Z}\right),\quad \mathcal{Z}\sim\mathcal{N}(0,I)$$

Baker-Campbell-Hausdorff (BCH) 公式:

$$e^Xe^Y=e^Z,\quad Z=X+Y+\frac{1}{2}[X,Y]+\dots$$

$$\hat{\mathcal{L}}=\mathcal{L}_{LD}+h\hat{\mathcal{L}}_1+h^2\hat{\mathcal{L}}_2+\mathcal{O}(h^4)$$

积分形式解：

$$\boldsymbol{x}_t = e^{\int_s^t f(\tau)d\tau}\boldsymbol{x}_s + \int_s^t \left(e^{\int_\tau^t f(r)dr} \frac{g^2(\tau)}{2\sigma_\tau} \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_\tau, \tau) \right) d\tau$$

$$\boldsymbol{x}_t = \frac{\alpha_t}{\alpha_s}\boldsymbol{x}_s - \alpha_t \int_{\lambda_s}^{\lambda_t} e^{-\lambda} \hat{\boldsymbol{\epsilon}}_\theta(\hat{\boldsymbol{x}}_\lambda, \lambda) d\lambda$$

k **阶格式：**

$$\boldsymbol{x}_{t_{i-1}\to t_i} = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}\tilde{\boldsymbol{x}}_{t_{i-1}} - \alpha_{t_i} \sum_{n=0}^{k-1} \hat{\boldsymbol{\epsilon}}_\theta^{(n)}(\hat{\boldsymbol{x}}_{\lambda_{t_{i-1}}}, \lambda_{t_{i-1}}) \int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda} \frac{(\lambda - \lambda_{t_{i-1}})^n}{n!} d\lambda + \mathcal{O}(h_i^{k+1})$$

参考文献

[1] Peter Holderrieth and Ezra Erives. *An Introduction to Flow Matching and Diffusion Models*. 2025. https://arxiv.org/abs/2506.02070.

[2] Omar Chehab and Anna Korba. “A Practical Diffusion Path for Sampling”. In: *ICML 2024 Workshop on Structured Probabilistic Inference & Generative Modeling*. 2024. https://openreview.net/forum?id=1pjDEW2AsZ.

[3] A. Leroy et al. “Adaptive Stepsize Algorithms for Langevin Dynamics”. In: *SIAM Journal on Scientific Computing* 46.6 (2024), A3574–A3598. https://doi.org/10.1137/24M1658590.

[4] Cheng Lu et al. “DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps”. In: *Advances in Neural Information Processing Systems*. Ed. by Alice H. Oh et al. 2022. https://openreview.net/forum?id=2uAaGwlP_V.

[5] Yang Song et al. “Score-Based Generative Modeling through Stochastic Differential Equations”. In: *International Conference on Learning Representations*. 2021. https://openreview.net/forum?id=PxTIG12RRHS.

[6] Grigorios A. Pavliotis. *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Vol. 60. Texts in Applied Mathematics. New York, NY: Springer New York, 2014.