生成式AI与随机微分方程 (Lec 2)

Constructing a Training Target for Flow and Diffusion Models

笔记整理：Gemini

`MIT IAP 2025 | Jan 22, 2025`

2025 年 7 月 25 日

目录

1 训练问题：如何为生成模型找到目标？

在 Lecture 1 中，我们学习了流模型和扩散模型是如何通过求解一个由神经网络 u_t^θ 参数化的ODE或SDE来生成样本的。

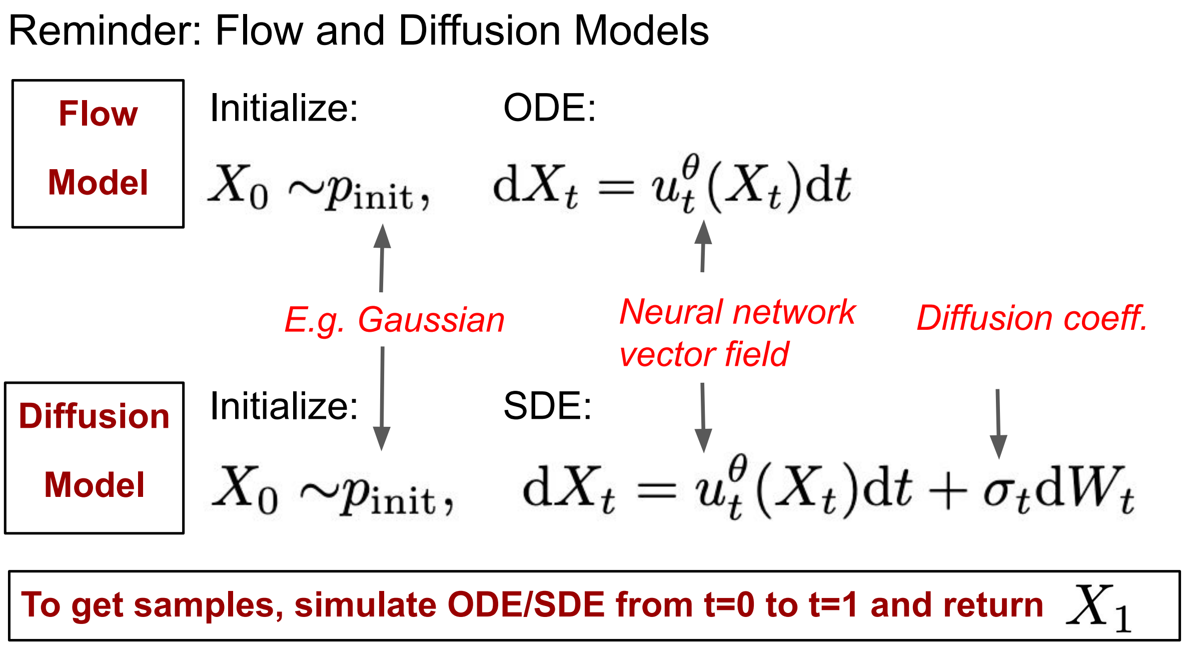

图 1: 流模型与扩散模型回顾 (源自幻灯片 Lec2, Page 2)。

然而，一个未经训练的神经网络 u_t^θ 只会产生无意义的噪声。为了让它能生成与真实数据分布 p_{data} 一致的样本，我们必须对其进行训练。

1.1 训练的核心挑战

在标准的监督学习中，我们有明确的标签 (label) 作为训练目标。但在生成任务中，我们并没有这样的标签。我们的目标是让模型产生的最终样本分布 p_1^θ 与真实数据分布 p_{data} 相匹配：

$$p_1^\theta(x) \approx p_{\text{data}}(x)$$

这个目标是关于整个分布的，我们无法直接用它来指导每一步的向量场 $u_t^\theta(x)$ 的更新。因此，本节课的核心问题是：

如何为向量场 $u_t^\theta(x)$ 找到一个可以直接优化的、逐点的训练目标 $u_t^{target}(x)$？

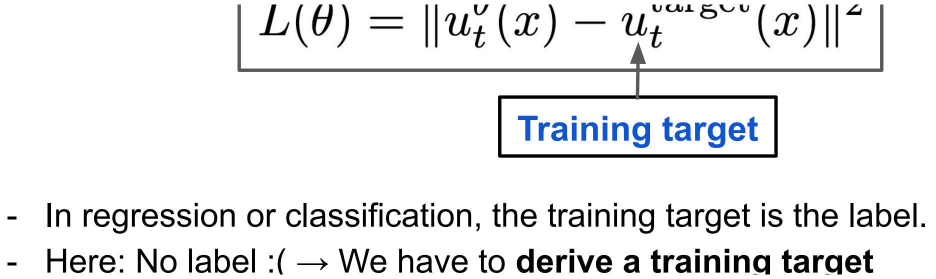

图 2: 训练目标是找到一个可以用于最小化均方误差的 $u_t^{\text{target}}(x)$ (源自幻灯片 Lec2, Page 4)。

2 核心工具：概率路径 (Probability Path)

为了找到目标向量场，我们首先要定义一个从初始噪声分布 p_{init} 到目标数据分布 p_{data} 的"桥梁"。这个桥梁就是**概率路径** $p_t(x)$，一个随时间 $t \in [0, 1]$ 连续变化的概率分布族。

一个概率密度函数族，它满足：

- $p_0(x) = p_{\text{init}}(x)$ （在 $t = 0$ 时是初始噪声分布）
- $p_1(x) = p_{\text{data}}(x)$ （在 $t = 1$ 时是真实数据分布）

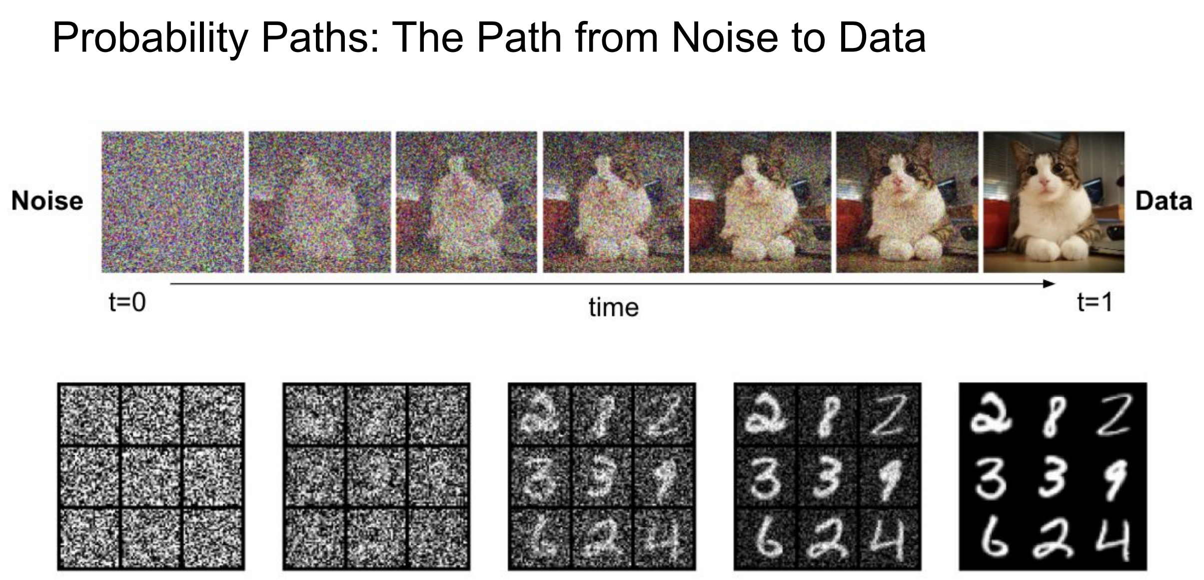

图 3: 概率路径可视化：从左侧的纯噪声 $(t = 0)$ 平滑地过渡到右侧的真实数据 $(t = 1)$ (源自幻灯片 Lec2, Page 9)。

2.1 条件路径 vs. 边际路径

理解“条件”与“边际”的区别至关重要，这是本节课的重点。

- **条件 (Conditional)**: 指的是**针对单个数据点** $z \in p_{\text{data}}$ 的演化过程。它描述了如何从一个随机噪声点演化到**某一个特定**的真实数据点。
- **边际 (Marginal)**: 指的是**在整个数据分布上**进行平均或积分后的整体演化过程。它描述了整个噪声分布是如何演化成整个数据分布的。

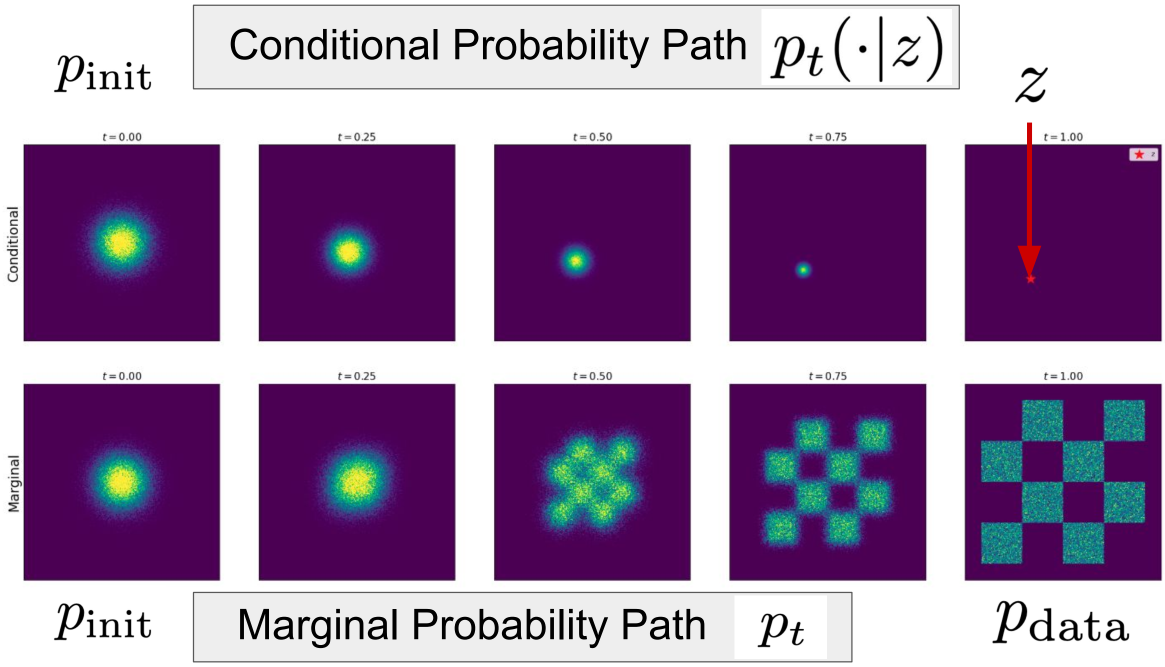

图 4: 上排为“条件概率路径”，展示了从一个高斯噪声分布收敛到一个特定数据点（星号）的过程。下排为“边际概率路径”，是所有条件路径的叠加，展示了整个分布的演化 (源自幻灯片 Lec2, Page 12)。

3 推导训练目标：向量场与分数函数

3.1 核心策略：从简单条件到复杂边际

我们的核心策略是“分而治之”，然后“整合归一”。

1. **定义简单的“条件”路径和场**: 我们可以很容易地定义一个**条件概率路径** $p_t(x|z)$，它描述了如何从一个噪声点演化到一个**特定**的数据点 z。例如，一个高斯模糊路径。对于这样简单的路径，其对应的**条件向量场** $u_t(x|z)$ 通常有简单的解析形式，是可计算的。
2. **挑战**: 在生成时，模型从一个随机噪声 x_0 出发，它并**不知道**自己最终应该生成哪一个具体的数据点 z。因此，模型无法直接使用任何一个特定的条件向量场 $u_t(x|z)$。
3. **解决方案**: 模型必须学习一个“平均”的、总体的行为。这个平均行为就是**边际向量场** $u_t(x)$。它是在给定当前位置 x 的条件下，对所有可能的条件向量场 $u_t(x|z)$ 进行加权平均（期望）的结果。

这个边际向量场 $u_t(x)$ 不再依赖于任何特定的终点 z，因此可以被我们的神经网络学习。它就是我们梦寐以求的训练目标。

3.2 流模型 (ODE) 的训练目标：边际向量场

3.2.1 连续性方程 (Continuity Equation)

这是连接概率路径 $p_t(x)$ 和其对应的向量场 $u_t(x)$ 的基本物理定律，它本质上是概率质量的守恒定律。

如果一个概率分布 $p_t(x)$ 是由一个向量场 $u_t(x)$ 驱动的ODE所产生的，那么它们必须满足以下偏微分方程：

$$\frac{\partial}{\partial t}p_t(x) = -\nabla \cdot (p_t(x)u_t(x))$$

其中 $\nabla\cdot$ 是散度算子。这个方程的直观解释是：某个区域内概率密度的变化率，等于流入该区域的概率通量减去流出的概率通量。

3.2.2 从条件到边际

有了连续性方程，我们就可以推导出边际向量场。

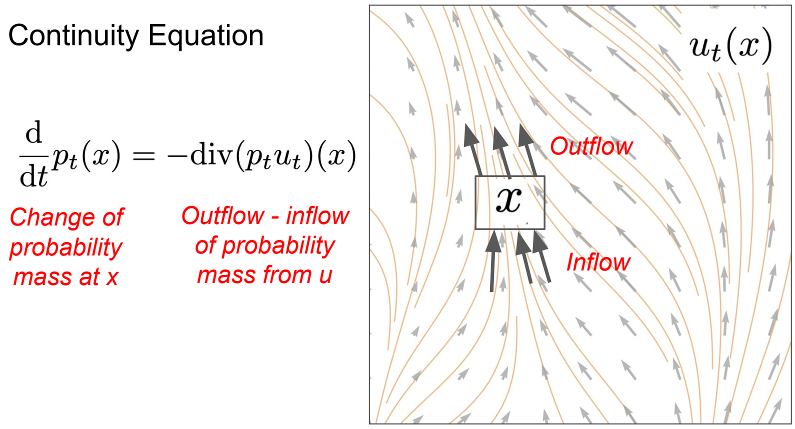

图 5: 连续性方程的直观解释：区域 $\mathcal{X}$ 内概率密度的变化取决于流入和流出的概率质量 (源自幻灯片 Lec2, Page 18)。

边际向量场是条件向量场 $u_t^{\text{target}}(x|z)$ 在后验分布 $p_t(z|x)$ 下的期望：

$$u_t^{\text{target}}(x) = \mathbb{E}_{z\sim p_t(z|x)}[u_t^{\text{target}}(x|z)] = \int u_t^{\text{target}}(x|z)p_t(z|x)dz$$

其中 $p_t(z|x) = \frac{p_t(x|z)p_{\text{data}}(z)}{p_t(x)}$ 是在时刻 t 观察到粒子在位置 x 的条件下，其最终目标是 z 的概率（通过贝叶斯定理得到）。

这个公式是**流匹配 (Flow Matching)**的核心。它告诉我们，理想的边际向量场（我们的训练目标），可以通过对简单的条件向量场进行加权平均得到。

3.3 扩散模型 (SDE) 的训练目标：边际分数函数

对于SDE，情况类似，但方程和目标有所不同。

3.3.1 福克-普朗克方程 (Fokker-Planck Equation)

这是SDE版本的连续性方程。它在连续性方程的基础上增加了一个描述随机扩散的项。

$$\frac{\partial}{\partial t}p_t(x) = -\nabla\cdot(p_t(x)u_t(x)) + \frac{\sigma_t^2}{2}\Delta p_t(x)$$

其中 Δ 是拉普拉斯算子。新增的 $\frac{\sigma_t^2}{2}\Delta p_t(x)$ 项被称为**热扩散项**，它描述了概率密度由于随机运动而向周围扩散的趋势。

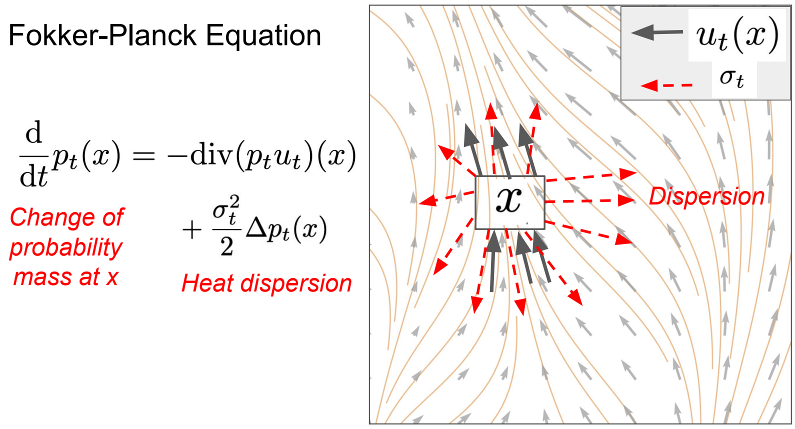

图 6: 福克-普朗克方程的直观解释：除了向量场驱动的流动外，还存在一个由噪声引起的向四周的扩散 (源自幻灯片 Lec2, Page 29)。

3.3.2 分数函数 (Score Function)

在扩散模型中，我们通常不直接学习向量场，而是学习一个与之等价但更容易处理的量——**分数函数**。

分数函数被定义为概率密度对数对 x 的梯度：

$$s_t(x) = \nabla_x \log p_t(x)$$

它指向概率密度增长最快的方向。

与向量场类似，边际分数函数是条件分数函数的期望：

$$s_t^{\text{target}}(x) = \nabla_x \log p_t(x) = \mathbb{E}_{z\sim p_t(z|x)}[\nabla_x \log p_t(x|z)]$$

这个公式是**分数匹配 (Score Matching)**的核心，它为扩散模型的训练提供了目标。

3.4 总结：六个核心公式

本节课的所有推导最终归结为六个核心概念及其公式，这是下一节课构建训练算法的基础。

	Notation	Key property	Gaussian example
Conditional Probability Path	$p_t(\cdot\|z)$	Interpolates p_{init} and a data point z	$\mathcal{N}(\alpha_t z, \beta_t^2 I_d)$
Conditional Vector Field	$u_t^{\text{target}}(x\|z)$	ODE follows conditional path	$\left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right) z + \frac{\dot{\beta}_t}{\beta_t} x$
Conditional Score	$\nabla \log p_t(x\|z)$	Gradient of log likelihood	$-\frac{x - \alpha_t z}{\beta_t^2}$

图 7: 条件路径、向量场和分数的定义与示例 (源自幻灯片 Lec2, Page 33)。

	Notation	Key property	Formula
Marginal Probability Path	p_t	Interpolates p_{init} and p_{data}	$\int p_t(x\|z) p_{\text{data}}(z) \mathrm{d}z$
Marginal Vector Field	$u_t^{\text{target}}(x)$	ODE follows marginal path	$\int u_t^{\text{target}}(x\|z) \frac{p_t(x\|z) p_{\text{data}}(z)}{p_t(x)} \mathrm{d}z$
Marginal Score Function	$\nabla \log p_t(x)$	Can be used to convert	$\int \nabla \log p_t(x\|z) \frac{p_t(x\|z) p_{\text{data}}(z)}{p_t(x)} \mathrm{d}z$

图 8: 边际路径、向量场和分数的定义 (源自幻灯片 Lec2, Page 34)。