

35 流匹配模型

流匹配模型 (flow matching model) 是一种概率生成模型。考虑概率分布从简单的先验分布 (如标准高斯分布) 向复杂的分布 (如数据分布) 随时间的连续演化过程。演化的路径称为概率路径 (probability path) 或概率流 (probability flow)。概率路径由依赖于时间的速度场 (velocity field) 以及先验分布决定。整个过程由基于速度场的常微分方程 (ODE) 描述。

流匹配学习的目标是利用训练数据拟合速度场。其核心思想是不直接拟合难于处理的速度场本身, 而是构建一个等价的学习目标, 拟合易于处理的条件速度场, 从而实现从标准高斯分布到数据分布的连续演化。

在数据生成过程中, 首先从标准高斯分布中随机采样一个样本, 然后利用学习得到的速度场, 通过数值积分方法求解基于速度场的常微分方程, 逐步计算出该样本的演化轨迹, 最终得到真实数据分布中的一个采样样本。

目前常用的流匹配方法是由 Lipman 等人于 2022 年系统地提出并加以完善的。

35.1 节概述流匹配模型的基本原理。35.2 节讲述流匹配模型的学习和生成算法, 包括理论推导。35.3 讲解流匹配模型与扩散模型的关系。

35.1 基本原理

在物理学中, “流 (flow)” 通常用于描述物质在确定性动力学作用下随时间的连续演化过程, 例如流体的整体流动与形变。流匹配中的 “流” 借鉴了这一概念, 用来刻画概率分布的连续演化。

可以将概率分布随时间的演化直观地理解为样本空间中的概率质量的 “流”。在初始时刻 $t = 0$, 概率质量呈现为简单的先验分布 $p_0(\mathbf{x})$; 随着时间的推移, 概率质量连续平滑地流动与变形, 在时刻 t 变成分布 $p_t(\mathbf{x})$; 在终止时刻 $t = 1$, 概率质量恰好演化为复杂的数据分布 $p_1(\mathbf{x})$ 。

图 35.1 示意了流匹配问题的基本原理。假设先验分布 $p_0(\mathbf{x})$ 为标准高斯分布, 其样本为随机噪声图像; 而数据分布 $p_1(\mathbf{x})$ 是要学习的目标分布, 其样本为真实图像。流匹配方法将数据分布的学习与采样问题形式化为上述概率流的建模问题: 通过从图像数据的学习, 掌握分布随时间演化的规律, 并据此生成新的图像。

概率流的演化由一个依赖于时间的速度场 $\mathbf{v}_t(\mathbf{x})$ 以及先验分布决定。速度场表示样本在空间中的随时间演化的方向和速率。基于速度场的常微分

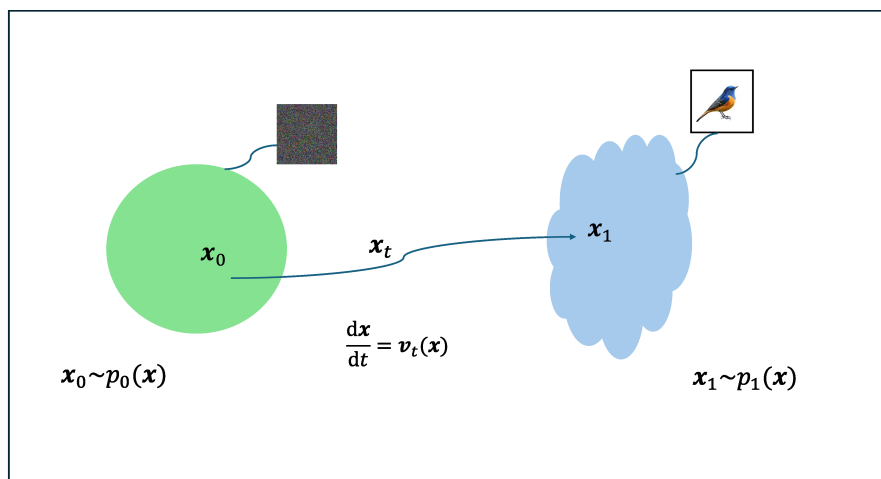


图 35.1: 流匹配: 从先验分布到数据分布的确定性的连续演化

方程的解是样本在空间中的演化轨迹。

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}_t(\mathbf{x})$$

流匹配的核心技术在于：构造一个与速度场学习等价的条件速度场学习目标，从而学习要学习的速度场。实际使用神经网络 $\mathbf{u}_\theta(\mathbf{x}, t)$ 对该速度场进行表示。学习完成后，将速度场 $\mathbf{u}_\theta(\mathbf{x}, t)$ 作为数据生成的模型。在初始时刻给定一个来自先验分布的样本，通过数值积分方法求解由该速度场定义的常微分方程，即可将样本逐步演化至终止时刻，从而获得服从数据分布的样本。

流匹配模型属于流模型（flow-based models）的一种，其形式与归一化流（normalizing flow）和连续归一化流（continuous normalizing flow, CNF）相近，均通过刻画概率分布的确定性的连续演化来实现生成模型。在实际应用中，流匹配模型通常表现出更高的采样效率和更好的生成质量。

流匹配模型与扩散模型有本质的不同，前者基于确定性动力学原理，后者基于随机性动力学原理；但两者之间也存在着紧密的联系。在概率流建模方面，可以建立起两者的对应关系。扩散模型中的随机扩散过程可以通过概率流常微分方程表示为确定性的连续演化过程，而流匹配模型正是针对概率流的确定性连续演化进行建模的。流匹配模型通常比扩散模型有更高的采样效率。扩散模型习惯将数据定义在时刻 0，噪声定义在时刻 T ；而流匹配模型习惯将噪声定义在时刻 0，数据定义在时刻 1。

35.2 流匹配问题

本节叙述概率路径、速度场、连续性方程等概念，给出流匹配学习问题的定义。

35.2.1 概率路径和速度场

考虑定义在连续时间区间 $[0, 1]$ 上的一个概率分布的序列。在初始时刻 $t = 0$ ，概率分布的概率密度函数为 $p_0(\mathbf{x})$ （简记为 p_0 ）；在终止时刻 $t = 1$ ，密度函数为 $p_1(\mathbf{x})$ （简记为 p_1 ）；在任意中间时刻 $t \in (0, 1)$ ，密度函数为 $p_t(\mathbf{x})$ （简记为 p_t ）。这里， $\mathbf{x} \in \mathbb{R}^d$ 表示样本空间中的样本，也是样本空间中的位置， d 是样本空间的维度。

概率分布在时刻 $t = 0$ 是先验分布 $p_0(\mathbf{x})$ ，随时间逐步演化：在时刻 t 变为中间分布 $p_t(\mathbf{x})$ ，在时刻 $t = 1$ 变为数据分布 $p_1(\mathbf{x})$ 。这个概率分布演化过程 $\{p_t(\mathbf{x})\}_{t \in [0, 1]}$ 被称为概率路径 (probability path) 或者概率流 (probability flow)。有时为了方便仅用时刻 t 的密度函数 $p_t(\mathbf{x})$ 代表整个概率路径。通常假设先验分布 $p_0(\mathbf{x})$ 是标准高斯分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ ，数据分布 $p_1(\mathbf{x}) = q(\mathbf{x})$ 是要学习的目标分布。

设 $\mathbf{x} \in \mathbb{R}^d$ 表示样本在时刻 t 的位置，满足 $\mathbf{x} \sim p_t(\cdot)$ 。样本 \mathbf{x} 随时间的演化由一个常微分方程 (ordinary differential equation, ODE) 描述：

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}(\mathbf{x}, t) \quad (35.1)$$

其中 $\mathbf{v}(\mathbf{x}, t) : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d$ 是一个向量场，它表示在给定时刻 t 和位置 \mathbf{x} 时，样本演化的方向与速率。称 $\mathbf{v}(\mathbf{x}, t)$ 为速度场 (velocity field)，将其简写为 $\mathbf{v}_t(\mathbf{x})$ 。假设 \mathbf{x}_0 是初始位置。

若速度场 $\mathbf{v}(\mathbf{x}, t)$ 对空间变量 \mathbf{x} 局部 Lipschitz 连续，对时间变量 t 连续，并满足线性增长条件，则对于任意初始位置 \mathbf{x}_0 ，常微分方程 (35.1) 在 $t \in [0, 1]$ 上存在唯一解。

由于速度场是依赖于时间的，即每一个时刻 t 都有一个速度场。当初始位置 \mathbf{x}_0 和速度场 $\mathbf{v}_t(\mathbf{x})$ 确定时，样本在所有时刻的位置，即其演化轨迹 $\mathbf{x}_0 \rightarrow \cdots \rightarrow \mathbf{x} \rightarrow \cdots \rightarrow \mathbf{x}_1$ 由该 ODE 唯一确定。其解可写为积分形式

$$\mathbf{x} = \mathbf{x}_0 + \int_0^t \mathbf{v}(\mathbf{x}_\tau, \tau) d\tau \quad (35.2)$$

也可以用数值积分的方法求解该 ODE。所以可以说基于速度场的 ODE 的解表示的是样本在空间中的演化轨迹。

这时，称速度场 $\mathbf{v}(\mathbf{x}, t)$ 诱导 (induce) 了概率路径 $p_t(\mathbf{x})$ ，用以下形式表示。

$$\mathbf{x}_0 \sim p_0(\cdot), \frac{d\mathbf{x}}{dt} = \mathbf{v}_t(\mathbf{x}) \Rightarrow \mathbf{x} \sim p_t(\cdot) \quad (35.3)$$

需要注意的是，能够诱导同一概率路径的速度场并不一定唯一。

ODE 的初始位置 \mathbf{x}_0 和每一个时刻 t 的解 \mathbf{x} 之间存在一个函数 ψ ，称之为流 (flow)。

$$\mathbf{x} = \psi(\mathbf{x}_0, t) \quad (35.4)$$

ODE 也可以写作

$$\begin{aligned} \frac{d\psi(\mathbf{x}_0, t)}{dt} &= \mathbf{v}(\psi(\mathbf{x}_0, t), t) \\ \psi(\mathbf{x}_0, 0) &= \mathbf{x}_0 \end{aligned} \quad (35.5)$$

这里， $\psi: \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d$ 表示样本从 \mathbf{x}_0 到 \mathbf{x} 的演化轨迹。将 $\psi(\mathbf{x}_0, t)$ 简写为 $\psi_t(\mathbf{x}_0)$ 。

通常假设流 $\psi_t(\mathbf{x}_0)$ 是一个微分同胚 (diffeomorphism)，即在适当的光滑性条件下，它是光滑、可逆且逆也是光滑的变换。流是在连续归一化流中也使用的概念。

可以看出，ODE 由速度场 $\mathbf{v}_t(\mathbf{x})$ 定义，其解 \mathbf{x} 可以表示为流 $\psi_t(\mathbf{x}_0)$ 。速度场 $\mathbf{v}_t(\mathbf{x})$ 与先验分布 $p_0(\mathbf{x})$ 决定一条概率路径。若能找到一个“正确”的速度场 $\mathbf{v}_t(\mathbf{x})$ ，则通过求解上述 ODE，可以让先验分布 $p_0(\mathbf{x})$ 的样本演化成数据分布 $p_1(\mathbf{x})$ 的样本。

35.2.2 连续性方程

在每一个时刻 t ，概率质量都等于 1，是恒等的。

$$\int p_t(\mathbf{x}) d\mathbf{x} = 1$$

这就意味着概率路径在演化时，必须满足这个约束条件。

在物理学中，物质的运动必须遵守质量守恒定律，其数学表达是连续性方程 (continuity equation)。对应着，概率路径的演化也必须满足守恒定律，也有相应的连续性方程。具体地，概率密度 $p_t(\mathbf{x})$ 和速度场 $\mathbf{v}_t(\mathbf{x})$ 的关系由概率连续性方程描述：

$$\frac{\partial p_t(\mathbf{x})}{\partial t} + \nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x}) \mathbf{v}_t(\mathbf{x})) = 0 \quad (35.6)$$

其中 $\nabla_{\mathbf{x}} \cdot$ 是散度算子。

连续性方程中, $\frac{\partial p_t(\mathbf{x})}{\partial t}$ 表示在位置 \mathbf{x} 的概率密度随时间的变化率, $\nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x})\mathbf{v}_t(\mathbf{x}))$ 表示从周围流入到该位置或从该位置流出到周围的概率密度。前者大于 0, 而后者小于 0 时, 概率质量从周围流入 (汇聚); 前者小于 0, 而后者大于 0 时, 概率质量向周围流出 (发散)。

关于连续性方程有以下定理成立。

定理 35.1 (连续性方程) 对于由速度场 $\mathbf{v}_t(\mathbf{x})$ 诱导的概率路径 $p_t(\mathbf{x})$, 若该速度场 $\mathbf{v}_t(\mathbf{x})$ 满足一定的光滑性条件 (对空间变量 \mathbf{x} 局部 Lipschitz 连续, 对时间变量 t 连续, 并满足线性增长条件), 则概率路径 $p_t(\mathbf{x})$ 满足如下连续性方程:

$$\frac{\partial}{\partial t} p_t(\mathbf{x}) = -\nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x})\mathbf{v}_t(\mathbf{x})) \quad (35.7)$$

其中, $\nabla_{\mathbf{x}} \cdot$ 是散度算子。反之, 若概率路径 $p_t(\mathbf{x})$ 满足此方程及初始条件 $p_0(\mathbf{x})$, 且方程中的速度场 $\mathbf{v}_t(\mathbf{x})$ 满足上述光滑性条件, 则概率路径 $p_t(\mathbf{x})$ 是由速度场 $\mathbf{v}_t(\mathbf{x})$ 诱导的概率路径。

35.2.3 流匹配

假设有训练数据集, 其中的样本 \mathbf{x} 来自一个未知的数据分布 $q(\mathbf{x})$ 。目标是要学习这个分布。假设存在一个从先验分布 $p_0(\mathbf{x})$ (标准高斯分布) 到数据分布 $p_1(\mathbf{x}) = q(\mathbf{x})$ 演化的概率路径, 产生了训练集的这些样本。流匹配 (flow matching) 旨在学习一个神经网络 $\mathbf{u}_{\theta}(\mathbf{x}, t)$, 能够拟合决定这个概率路径的速度场 $\mathbf{v}_t(\mathbf{x})$ (也就是上述正确的速度场)。损失函数定义为

$$L_{\text{FM}}(\theta) = \mathbb{E}_{t, \mathbf{x}} \|\mathbf{u}_{\theta}(\mathbf{x}, t) - \mathbf{v}_t(\mathbf{x})\|_2^2, \quad t \sim U[0, 1], \mathbf{x} \sim p_t(\cdot) \quad (35.8)$$

其中神经网络 $\mathbf{u}_{\theta}(\mathbf{x}, t) : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ 表示在位置 \mathbf{x} 和时刻 t 的速度向量, θ 是神经网络的参数, $\mathbf{v}_t(\mathbf{x}) \in \mathbb{R}^d$ 表示目标的速度场, $\|\cdot\|$ 是 L_2 范数, $U[0, 1]$ 是区间 $[0, 1]$ 上的均匀分布。

学到神经网络速度场之后, 就可以从标准高斯分布出发, 生成数据分布的样本。这里面临的问题是, 损失函数 $L_{\text{FM}}(\theta)$ 是无法直接计算的, 因为它依赖于真实的未知的速度场 $\mathbf{v}_t(\mathbf{x})$ 和难以直接采样的中间分布 $p_t(\mathbf{x})$ 。流匹配方法通过实现一个等价的条件速度场的学习解决这个问题。

35.3 流匹配方法

本节将系统阐述流匹配算法及其理论基础。首先，定义条件概率路径和条件速度场。然后，推导它们与要学习的目标概率路径和目标速度场（即边缘概率路径和边缘速度场）之间的关系。接着，证明学习目标边缘速度场等价于学习条件速度场。最后，推导出条件概率路径为高斯分布时，流匹配的常用学习算法。

35.3.1 条件概率路径与条件速度场

首先，构造条件速度场和条件概率路径。定义在连续时间区间 $[0, 1]$ 上的条件概率分布的序列，称之为条件概率路径。给定样本（或位置） \mathbf{x}_1 ，条件概率分布为 $p_t(\mathbf{x}|\mathbf{x}_1)$ ，其中 \mathbf{x} 表示在时刻 $t \in (0, 1)$ 的样本空间中的位置。假设初始条件概率分布和终止条件概率分布分别是

$$p_0(\mathbf{x}|\mathbf{x}_1) = \pi(\mathbf{x}), \quad p_1(\mathbf{x}|\mathbf{x}_1) = \delta_{\mathbf{x}_1}(\mathbf{x}) \quad (35.9)$$

其中 $\pi(\mathbf{x})$ 是任意概率分布，通常是标准高斯分布； $\delta_{\mathbf{x}_1}(\mathbf{x})$ 表示狄拉克 δ 分布满足

$$\delta_{\mathbf{x}'}(\mathbf{x}) = \begin{cases} +\infty, & \mathbf{x} = \mathbf{x}', \\ 0, & \mathbf{x} \neq \mathbf{x}', \end{cases} \quad \text{且} \quad \int_{\mathbb{R}^d} \delta_{\mathbf{x}'}(\mathbf{x}) d^d \mathbf{x} = 1.$$

对应地，定义条件速度场。给定样本 \mathbf{x}_1 ，条件速度场为 $\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)$ ，其中 \mathbf{x} 表示在时刻 $t \in (0, 1)$ 的样本的位置。

这样，可以定义在每一个时刻的样本演化的常微分方程（ODE）

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) \quad (35.10)$$

在样本初始位置 \mathbf{x}_0 和条件速度场 $\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)$ 确定的条件下，条件概率路径 $p_t(\mathbf{x}|\mathbf{x}_1)$ 就确定。也就是说，条件速度场诱导条件概率路径。

$$\mathbf{x}_0 \sim \pi(\cdot), \mathbf{x}_1 \sim q(\cdot), \frac{d\mathbf{x}}{dt} = \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) \Rightarrow \mathbf{x} \sim p_t(\cdot|\mathbf{x}_1) \quad (35.11)$$

根据定理 35.1，条件概率路径满足对应的连续性方程。

$$\frac{\partial}{\partial t} p_t(\mathbf{x}|\mathbf{x}_1) + \nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x}|\mathbf{x}_1) \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)) = 0 \quad (35.12)$$

35.3.2 边缘概率路径与边缘速度场

接着，推导出条件概率路径对应的边缘概率路径就是学习目标的概率路径，条件速度场对应的边缘速度场就是学习目标的速度场。

定义边缘概率路径为条件概率路径 $p_t(\mathbf{x}|\mathbf{x}_1)$ 关于数据分布 $q(\mathbf{x}_1)$ 的期望。

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x}|\mathbf{x}_1)q(\mathbf{x}_1)d\mathbf{x}_1 \quad (35.13)$$

其中 $q(\mathbf{x}_1)$ 是数据分布。容易推出，初始边缘分布和终止边缘分布分别满足

$$p_0(\mathbf{x}) = \pi(\mathbf{x}), \quad p_1(\mathbf{x}) = q(\mathbf{x}) \quad (35.14)$$

其中 \mathbf{x} 表示样本空间中的位置。

定义边缘速度场为条件速度场 $\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)$ 关于后验概率分布 $p_t(\mathbf{x}_1|\mathbf{x})$ 的期望。根据贝叶斯定理有

$$\mathbf{v}_t(\mathbf{x}) = \int \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) \frac{p_t(\mathbf{x}|\mathbf{x}_1)q(\mathbf{x}_1)}{p_t(\mathbf{x})} d\mathbf{x}_1 \quad (35.15)$$

以下定理 35.2 保证，上述边缘速度场可以诱导边缘概率路径，也就是学习目标的速度场和概率路径。

定理 35.2 若条件速度场 $\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)$ 诱导条件概率路径 $p_t(\mathbf{x}|\mathbf{x}_1)$ ，则对应的边缘速度场 $\mathbf{v}_t(\mathbf{x})$ 诱导对应的边缘概率路径 $p_t(\mathbf{x})$ ，即边缘速度场和边缘概率路径满足连续性方程 (35.6)。

$$\mathbf{x}_0 \sim \pi(\cdot), \frac{d\mathbf{x}}{dt} = \mathbf{v}_t(\mathbf{x}) \Rightarrow \mathbf{x} \sim p_t(\cdot) \quad (35.16)$$

证明

$$\begin{aligned} \frac{\partial p_t(\mathbf{x})}{\partial t} &= \frac{\partial}{\partial t} \int p_t(\mathbf{x}|\mathbf{x}_1)q(\mathbf{x}_1)d\mathbf{x}_1 \\ &= \int \frac{\partial}{\partial t} p_t(\mathbf{x}|\mathbf{x}_1)q(\mathbf{x}_1)d\mathbf{x}_1 \\ &= \int -\nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x}|\mathbf{x}_1)\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)) q(\mathbf{x}_1)d\mathbf{x}_1 \\ &= -\nabla_{\mathbf{x}} \cdot \int p_t(\mathbf{x}|\mathbf{x}_1)\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)q(\mathbf{x}_1)d\mathbf{x}_1 \\ &= -\nabla_{\mathbf{x}} \cdot \left(p_t(\mathbf{x}) \int \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) \frac{p_t(\mathbf{x}|\mathbf{x}_1)q(\mathbf{x}_1)}{p_t(\mathbf{x})} d\mathbf{x}_1 \right) \\ &= -\nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x})\mathbf{v}_t(\mathbf{x})) \end{aligned}$$

第一步利用边缘概率路径的定义，第二步使用莱布尼茨积分法则，第三步基于连续性方程，第四步使用莱布尼茨积分法则，第五步根据积分的性质，第六步利用边缘速度场的定义。这里假设相关函数满足可积性与光滑性条件，使得积分与微分可交换。

■

35.3.3 流匹配的目标函数

考虑学习一个神经网络 $\mathbf{u}_\theta(\mathbf{x}, t)$ ，能够拟合决定条件概率路径的条件速度场 $\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)$ 。损失函数定义为

$$L_{\text{CFM}}(\theta) = \mathbb{E}_{t, \mathbf{x}_1, \mathbf{x}} \|\mathbf{u}_\theta(\mathbf{x}, t) - \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)\|_2^2 \quad (35.17)$$

$$t \sim U[0, 1], \mathbf{x}_1 \sim q(\cdot), \mathbf{x} \sim p_t(\cdot|\mathbf{x}_1)$$

其中神经网络 $\mathbf{u}_\theta(\mathbf{x}, t) : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ 表示在时刻 t 和位置 \mathbf{x} 的速度向量， θ 是神经网络的参数， $\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) \in \mathbb{R}^d$ 表示目标的条件速度场， $\|\cdot\|$ 是 L_2 范数， $U[0, 1]$ 是区间 $[0, 1]$ 上的均匀分布。

可以证明条件速度场的学习和目标速度场（边缘速度场）的学习是等价的。有以下定理成立。所以，通过学习一个条件速度场，可以得到目标的（边缘）速度场，这就是流匹配方法的核心想法。

定理 35.3 最小化流匹配的目标函数 L_{FM} 等价于最小化条件流匹配的目标函数 L_{CFM} 。

$$\min_{\theta} L_{\text{FM}}(\theta) = \min_{\theta} L_{\text{CFM}}(\theta) \quad (35.18)$$

证明

损失函数 L_{FM} 可以写作

$$\begin{aligned} L_{\text{FM}}(\theta) &= \mathbb{E}_{t, \mathbf{x}} \|\mathbf{u}_\theta(\mathbf{x}, t) - \mathbf{v}_t(\mathbf{x})\|_2^2 \\ &= \mathbb{E}_{t, \mathbf{x}} [\|\mathbf{u}_\theta(\mathbf{x}, t)\|_2^2 - 2\mathbf{u}_\theta(\mathbf{x}, t)^T \mathbf{v}_t(\mathbf{x}) + \|\mathbf{v}_t(\mathbf{x})\|_2^2] \\ &= \mathbb{E}_{t, \mathbf{x}} [\|\mathbf{u}_\theta(\mathbf{x}, t)\|_2^2] - 2\mathbb{E}_{t, \mathbf{x}} [\mathbf{u}_\theta(\mathbf{x}, t)^T \mathbf{v}_t(\mathbf{x})] + \mathbb{E}_{t, \mathbf{x}} [\|\mathbf{v}_t(\mathbf{x})\|_2^2] \\ &= \mathbb{E}_{t, \mathbf{x}, \mathbf{x}_1} [\|\mathbf{u}_\theta(\mathbf{x}, t)\|_2^2] - 2\mathbb{E}_{t, \mathbf{x}} [\mathbf{u}_\theta(\mathbf{x}, t)^T \mathbf{v}_t(\mathbf{x})] + C' \end{aligned} \quad (35.19)$$

其中第三步根据数学期望性质；第四步第一项进一步对数据分布 $q(\mathbf{x}_1)$ 取期望，第三项是不依赖于参数 θ 的常数项 C' 。

进一步对其中的第二项进行推导。

$$\begin{aligned}
\mathbb{E}_{t,x} [\mathbf{u}_\theta(\mathbf{x}, t)^\top \mathbf{v}_t(\mathbf{x})] &= \int_0^1 \int p_t(\mathbf{x}) \mathbf{u}_\theta(\mathbf{x}, t)^\top \mathbf{v}_t(\mathbf{x}) d\mathbf{x} dt \quad (35.20) \\
&= \int_0^1 \int p_t(\mathbf{x}) \mathbf{u}_\theta(\mathbf{x}, t)^\top \left[\int \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) \frac{p_t(\mathbf{x}|\mathbf{x}_1) q(\mathbf{x}_1)}{p_t(\mathbf{x})} d\mathbf{x}_1 \right] d\mathbf{x} dt \\
&= \int_0^1 \int \int \mathbf{u}_\theta(\mathbf{x}, t)^\top \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) p_t(\mathbf{x}|\mathbf{x}_1) q(\mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x} dt \\
&= \mathbb{E}_{t,x,x_1} [\mathbf{u}_\theta(\mathbf{x}, t)^\top \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)]
\end{aligned}$$

其中第二步根据边缘速度场定义，第三步基于积分性质，第四步根据数学期望定义。

损失函数 L_{CFM} 可以写作

$$\begin{aligned}
L_{\text{CFM}}(\boldsymbol{\theta}) &= \mathbb{E}_{t,x,x_1} \|\mathbf{u}_\theta(\mathbf{x}, t) - \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)\|_2^2 \quad (35.21) \\
&= \mathbb{E}_{t,x,x_1} [\|\mathbf{u}_\theta(\mathbf{x}, t)\|_2^2 - 2\mathbf{u}_\theta(\mathbf{x}, t)^\top \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) + \|\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)\|_2^2] \\
&= \mathbb{E}_{t,x,x_1} [\|\mathbf{u}_\theta(\mathbf{x}, t)\|_2^2] - 2\mathbb{E}_{t,x,x_1} [\mathbf{u}_\theta(\mathbf{x}, t)^\top \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)] + \mathbb{E}_{t,x,x_1} [\|\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)\|_2^2] \\
&= \mathbb{E}_{t,x,x_1} [\|\mathbf{u}_\theta(\mathbf{x}, t)\|_2^2] - 2\mathbb{E}_{t,x,x_1} [\mathbf{u}_\theta(\mathbf{x}, t)^\top \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)] + (C' - C)
\end{aligned}$$

其中第三步根据数学期望定义，第四步中第三项是不依赖于 $\boldsymbol{\theta}$ 的常数项 $(C' - C)$ 。

比较式 (35.19) 和 (35.21)，利用式 (35.20)，得到

$$L_{\text{FM}}(\boldsymbol{\theta}) = L_{\text{CFM}}(\boldsymbol{\theta}) + C$$

其中 C 是常数项。 ■

间接优化目标函数 L_{CFM} 的好处在于无需知道或计算边缘速度场 $\mathbf{v}_t(\mathbf{x})$ 和边缘概率路径 $p_t(\mathbf{x})$ ，只需要从简单的条件分布 $p_t(\mathbf{x}|\mathbf{x}_1)$ 中采样，并利用已知的、有解析形式的条件速度场 $\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)$ 。这解决了直接优化目标函数 L_{FM} 困难的问题。

35.3.4 流匹配的算法

有了流匹配学习的一般方法，剩下的问题是如何定义具体的条件概率路径。通常假设它是高斯条件概率路径，得到常用的算法。

高斯条件概率路径，假设每一时刻 t 的条件概率分布定义为

$$p_t(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\alpha_t \mathbf{x}_1, \beta_t^2 \mathbf{I}).$$

等价地, 引入与 \mathbf{x}_1 独立的 $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, 并定义

$$\mathbf{x} = \alpha_t \mathbf{x}_1 + \beta_t \mathbf{x}_0.$$

其中 α_t 和 β_t 是参数。

设 $\alpha_0 = \beta_1 = 0$ 和 $\alpha_1 = \beta_0 = 1$, 即条件分布满足

$$p_0(\mathbf{x}|\mathbf{x}_1) = \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad p_1(\mathbf{x}|\mathbf{x}_1) = \delta_{\mathbf{x}_1}(\mathbf{x})$$

那么对应的边缘分布满足

$$p_0(\mathbf{x}) = \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad p_1(\mathbf{x}) = q(\mathbf{x})$$

这样, 有条件速度场可以诱导高斯条件概率路径

$$\begin{aligned} \mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathbf{x}_1 \sim q(\cdot), \frac{d\mathbf{x}}{dt} = \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) \Rightarrow \mathbf{x} \sim p_t(\cdot|\mathbf{x}_1) \\ \mathbf{x} = \alpha_t \mathbf{x}_1 + \beta_t \mathbf{x}_0 \end{aligned} \quad (35.22)$$

条件速度场为

$$\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t} \alpha_t \right) \mathbf{x}_1 + \frac{\dot{\beta}_t}{\beta_t} \mathbf{x} \quad (35.23)$$

其中 $\dot{\alpha}_t$ 表示 α_t 对时间 t 的导数, $\dot{\beta}_t$ 表示 β_t 对时间 t 的导数。证明留作习题。

当 \mathbf{x}_0 确定时可以推出

$$\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) = \dot{\alpha}_t \mathbf{x}_1 + \dot{\beta}_t \mathbf{x}_0$$

考虑一个具体情况: 条件概率路径是 \mathbf{x}_1 和 \mathbf{x}_0 的线性插值。也就是

$$\alpha_t = t, \beta_t = 1 - t$$

$$\mathbf{x} = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0$$

这样就有

$$\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) = \mathbf{x}_1 - \mathbf{x}_0$$

目标函数变成很简单的形式

$$\begin{aligned} L_{\text{CFM}}(\boldsymbol{\theta}) &= \mathbb{E}_{t, \mathbf{x}, \mathbf{x}_1} \|\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x}, t) - (\mathbf{x}_1 - \mathbf{x}_0)\|_2^2 \\ &= \mathbb{E}_{t, \mathbf{x}_0, \mathbf{x}_1} \|\mathbf{u}_{\boldsymbol{\theta}}(t\mathbf{x}_1 + (1 - t)\mathbf{x}_0, t) - (\mathbf{x}_1 - \mathbf{x}_0)\|_2^2 \end{aligned} \quad (35.24)$$

线性插值路径是在最优传输 (optimal transport) 意义下, 将先验分布连续地推送到数据分布的一条能量最小的传输路径。这一性质不仅在最优传输理论上具有最优性, 而且在实践中带来了更稳定的训练过程和更高效的采样效率, 使得流匹配模型在实际应用中得以广泛采用。

基于训练数据对该目标函数进行优化, 可以学习到与边缘速度场一致的神经网络。算法 35.1 给出了流匹配在线性插值情形下的学习算法。

算法 35.1 (流匹配 (线性插值) ——学习算法)

输入: 训练数据集 \mathcal{T} 。

输出: 速度场神经网络 $\mathbf{u}_\theta(\mathbf{x}, t)$ 。

- (1) 初始化神经网络参数 θ 。
- (2) 重复以下处理, 直到收敛:
 - (2-1) 从训练数据集 \mathcal{T} 中采样数据样本 $\mathbf{x}_1 \sim q(\cdot)$;
 - (2-2) 随机采样时间 $t \sim U[0, 1]$;
 - (2-3) 随机采样初始噪声 $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$;
 - (2-4) 构造中间状态样本 (线性插值路径):

$$\mathbf{x} = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0$$

- (2-5) 计算损失函数梯度并更新神经网络参数:

$$\nabla_\theta L(\theta) = \nabla_\theta \|\mathbf{u}_\theta(\mathbf{x}, t) - (\mathbf{x}_1 - \mathbf{x}_0)\|_2^2$$

- (3) 输出学习得到的速度场神经网络 $\mathbf{u}_\theta(\mathbf{x}, t)$ 。 ■

算法 35.2 给出了流匹配在线性插值情形下的生成算法。该算法采用欧拉法求解对应的 ODE, 其中速度场由学习得到的神经网络表示。原理上, 也可以采用更高精度的 ODE 数值积分求解方法。

算法 35.2 (流匹配 (线性插值) ——生成算法)

输入: 速度场神经网络 $\mathbf{u}_\theta(\mathbf{x}, t)$ 。

输出: 生成的样本 \mathbf{x}_1 。

超参数: 时间步数 T , 时间步长 $\Delta t = 1/T$ 。

- (1) 随机采样初始噪声 $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 。
- (2) For $(k = 0, 1, \dots, T - 1)\{$
 - (2-1) 设当前时间为 $t_k = k\Delta t$;
 - (2-2) 根据速度场计算样本的更新:

$$\mathbf{x}_{t_{k+1}} = \mathbf{x}_{t_k} + \Delta t \mathbf{u}_\theta(\mathbf{x}_{t_k}, t_k)$$

}

(3) 输出生成的样本 $\mathbf{x}_1 = \mathbf{x}_{t_T}$ 。 ■

35.4 流匹配模型与扩散模型的关系

本节首先指出流匹配是可以应用于扩散过程的更一般的框架，对流匹配模型和扩散模型进行比较。

35.4.1 扩散过程的流匹配模型

随机微分方程 (SDE)

扩散过程通过随机微分方程 (SDE) 描述样本在连续时间中的随机演化过程，包含前向过程（加噪）和对应的反向过程（去噪）。

前向过程预先定义，其 SDE 形式为：

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w} \quad (35.25)$$

其中， $\mathbf{f}(\mathbf{x}, t)$ 为漂移系数， $g(t)$ 为扩散系数， \mathbf{w} 是标准维纳过程。该过程从数据分布 $p_0(\mathbf{x})$ 的样本出发，通过逐步添加随机噪声，最终得到先验分布（如标准高斯分布 $p_1(\mathbf{x})$ ）的样本。

目标是学习反向过程。该过程从先验分布 $p_1(\mathbf{x})$ 的样本出发，通过逐步去除随机噪声，最终得到数据分布 $p_0(\mathbf{x})$ 的样本。根据随机过程理论，反向 SDE 的形式为：

$$d\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - g^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x})] dt + g(t)d\bar{\mathbf{w}} \quad (35.26)$$

其中， $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ 为分数函数或分数场， $\bar{\mathbf{w}}$ 是反向时间的标准维纳过程。学习目标变为估计每一时刻 t 的分数场。

福克-普朗克方程

扩散过程的前向或反向过程中每一时刻存在一个概率分布，其密度函数就是 $p_t(\mathbf{x})$ 。整体形成一个前向或反向的概率流或概率路径。概率路径满足一个偏微分方程，称为福克-普朗克方程 (Fokker-Planck equation) 或柯尔莫哥洛夫前向方程 (Kolmogorov forward equation)。

定理 35.4 (福克-普朗克方程) 对于扩散过程的前向过程(35.25)，若其漂移系数 $\mathbf{f}(\mathbf{x}, t)$ 和扩散系数 $g(t)$ 满足一定的光滑性条件 ($\mathbf{f}(\mathbf{x}, t)$ 对空

间变量 \mathbf{x} 局部 Lipschitz 连续、对时间变量 t 连续, 并满足线性增长; 且 $g(t)$ 对时间变量 t 连续且有界), 则其概率路径 $p_t(\mathbf{x})$ 满足如下福克-普朗克方程:

$$\frac{\partial}{\partial t} p_t(\mathbf{x}) = -\nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x}) \mathbf{f}(\mathbf{x}, t)) + \frac{g^2(t)}{2} \nabla_{\mathbf{x}}^2 p_t(\mathbf{x}) \quad (35.27)$$

其中, $\nabla_{\mathbf{x}} \cdot$ 是散度算子, $\nabla_{\mathbf{x}}^2$ 是拉普拉斯算子。反之, 若概率路径 $p_t(\mathbf{x})$ 满足此方程及初始条件 $p_0(\mathbf{x})$, 则它表示该扩散过程(35.25)的概率路径。

福克-普朗克方程(35.27)是概率质量守恒定律(连续性方程)在扩散过程中的具体形式。

常微分方程 (ODE)

上述扩散过程的概率路径或概率流, 可以用确定性的 ODE 描述。

推论 35.1 针对扩散过程的前向过程(35.25)和反向过程(35.26), 存在如下的速度场 $\mathbf{v}_t(\mathbf{x})$ 诱导概率密度函数 $p_t(\mathbf{x})$ 表示的概率路径。

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}_t(\mathbf{x}) = \mathbf{f}(\mathbf{x}, t) - \frac{g^2(t)}{2} \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \quad (35.28)$$

证明:

前向过程, 从福克-普朗克方程 (35.27) 出发:

$$\begin{aligned} \frac{\partial}{\partial t} p_t(\mathbf{x}) &= -\nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x}) \mathbf{f}(\mathbf{x}, t)) + \frac{g^2(t)}{2} \nabla_{\mathbf{x}}^2 p_t(\mathbf{x}) \\ &= -\nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x}) \mathbf{f}(\mathbf{x}, t)) + \frac{g^2(t)}{2} \nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x}) \nabla_{\mathbf{x}} \log p_t(\mathbf{x})) \\ &= -\nabla_{\mathbf{x}} \cdot \left[p_t(\mathbf{x}) \left(\mathbf{f}(\mathbf{x}, t) - \frac{g^2(t)}{2} \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \right) \right]. \end{aligned}$$

令 $\mathbf{v}_t(\mathbf{x}) = \mathbf{f}(\mathbf{x}, t) - \frac{g^2(t)}{2} \nabla_{\mathbf{x}} \log p_t(\mathbf{x})$, 上式即可写为连续性方程形式:

$$\frac{\partial}{\partial t} p_t(\mathbf{x}) = -\nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x}) \mathbf{v}_t(\mathbf{x}))$$

反向过程是前向过程的概率流给定条件下的时间反演过程。时间逆向从 $t = 1$ 推移到 $t = 0$, 其概率流 ODE 仍是(35.28)的形式。 ■

从推论 35.1 可知, 扩散模型在学到分数场 $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ 后, 可以直接使用 ODE(35.28)进行确定性的生成, 得到数据分布的样本, 而无需使用反向 SDE 进行随机性的采样。

表 35.1: 扩散模型与流匹配模型比较

	流匹配模型	扩散模型 (SDE 形式)
建模方式	确定性, 概率分布演化	随机性, 样本正向加噪 + 反向去噪
核心方程	常微分方程 (ODE)	随机微分方程 (SDE)
学习目标	速度场 $\mathbf{v}_t(\mathbf{x})$	分数场 $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ 或噪声
约束条件	连续性方程	福克-普朗克方程
采样效率	可用较少步数	需要较多步数; 也有确定性形式

例如, 在 DDPM 模型中, 前向 SDE 的系数为 $f(\mathbf{x}, t) = -\frac{\beta(t)}{2}\mathbf{x}$, $g(t) = \sqrt{\beta(t)}$ 。根据推论 35.1, 其对应的概率流 ODE 为:

$$\frac{d\mathbf{x}}{dt} = -\frac{\beta(t)}{2}(\mathbf{x} + \nabla_{\mathbf{x}} \log p_t(\mathbf{x})) \quad (35.29)$$

可以直接用这个概率流生成数据。DDIM 等加速采样算法, 本质上基于这个想法。

35.4.2 流匹配模型与扩散模型比较

流匹配模型与扩散模型都属于概率生成模型, 但两者在建模的基本原理上存在本质区别。流匹配模型通过确定性动力学刻画概率分布的确定性演化, 其过程由常微分方程 (ODE) 描述。这一过程由连续性方程约束。学习的是速度场 $\mathbf{v}_t(\mathbf{x})$, 将先验分布演化为目标数据分布。由于演化过程是确定性的, 采样通常可通过较少的步骤进行计算, 具有较高的采样效率。

相比之下, 扩散模型通过随机性动力学刻画数据样本的随机演化: 正向过程逐步对数据加噪, 将数据分布转化为先验分布, 由随机微分方程 (SDE) 描述; 反向过程, 逐步对数据去噪, 将先验分布转换为数据分布, 也由 SDE 描述。两个过程都由福克-普朗克方程约束。学习的是分数场 $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ 。由于反向过程是随机的, 采样通常需较多的步骤进行, 尽管它也有等价的确定性演化形式 (概率流 ODE)。

另外, 在扩散模型中, 其反向过程同时存在由分数函数决定的速度场。因此, 分数匹配可被视为一种在随机性动力学框架下对速度场的间接学习方式; 该速度场可与流匹配模型的建模框架相联系。

总之, 流匹配模型和扩散模型在建模方式、核心方程、学习目标、约束条件与采样行为上各有特点, 共同体现了概率生成模型中的动力学原理和强大生成能力 (表 35.1)。

本章概要

1. 流匹配模型的基本概念 定义在连续时间区间 $t \in [0, 1]$ 上的概率分布序列 $\{p_t(\mathbf{x})\}_{t \in [0, 1]}$, 称为概率路径或概率流。概率流在时刻 $t = 0$ 从先验分布 (如标准高斯分布) p_0 开始, 随时间连续演化, 在时刻 t 变为中间分布 p_t , 并在时刻 $t = 1$ 变为数据的真实数据分布 p_1 。

样本 \mathbf{x} 随时间的演化过程可以由一个常微分方程 (ODE) 描述:

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}(\mathbf{x}, t)$$

其中 $\mathbf{v}(\mathbf{x}, t)$ 是依赖于时间的速度场, 表示样本演化的方向与速率。

在一定的平滑性条件下, 速度场 $\mathbf{v}_t(\mathbf{x})$ 与初始分布 p_0 唯一决定一条概率路径 $p_t(\mathbf{x})$, 称该速度场诱导了该概率路径。

2. 连续性方程 连续性方程描述了概率分布演化过程中的质量守恒:

$$\frac{\partial p_t(\mathbf{x})}{\partial t} + \nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x}) \mathbf{v}_t(\mathbf{x})) = 0$$

在一定的平滑性条件下, 该方程是速度场 $\mathbf{v}_t(\mathbf{x})$ 诱导概率路径 $p_t(\mathbf{x})$ 的充要条件。

3. 流匹配方法的核心思想 流匹配方法构造条件速度场和条件概率路径, 可以证明: 由此得到的边缘速度场诱导目标的边缘概率路径; 学习目标边缘速度场等价于学习条件速度场。

3-1. 条件概率路径与条件速度场 给定数据样本 \mathbf{x}_1 , 定义条件速度场和条件概率路径, 并假设条件速度场诱导条件概率路径:

$$\mathbf{x}_0 \sim \pi(\cdot), \mathbf{x}_1 \sim q(\cdot), \frac{d\mathbf{x}}{dt} = \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) \Rightarrow \mathbf{x} \sim p_t(\cdot|\mathbf{x}_1)$$

设初始与终止条件分布为

$$p_0(\mathbf{x}|\mathbf{x}_1) = \pi(\mathbf{x}) \quad p_1(\mathbf{x}|\mathbf{x}_1) = \delta_{\mathbf{x}_1}(\mathbf{x})$$

其中 $\delta_{\mathbf{x}_1}$ 表示以 \mathbf{x}_1 为中心的狄拉克分布。

定义边缘分布

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x}|\mathbf{x}_1) q(\mathbf{x}_1) d\mathbf{x}_1$$

以及边缘速度场

$$\mathbf{v}_t(\mathbf{x}) = \int \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) \frac{p_t(\mathbf{x}|\mathbf{x}_1)q(\mathbf{x}_1)}{p_t(\mathbf{x})} d\mathbf{x}_1$$

在上述条件下, 边缘速度场诱导边缘概率路径:

$$\mathbf{x}_0 \sim \pi(\cdot), \frac{d\mathbf{x}}{dt} = \mathbf{v}_t(\mathbf{x}) \Rightarrow \mathbf{x} \sim p_t(\cdot)$$

3-2. 条件流匹配的学习目标 使用神经网络 $\mathbf{u}_\theta(\mathbf{x}, t)$ 拟合条件速度场, 学习的目标为

$$L_{\text{CFM}}(\theta) = \mathbb{E}_{t, \mathbf{x}, \mathbf{x}_1} \|\mathbf{u}_\theta(\mathbf{x}, t) - \mathbf{v}_t(\mathbf{x}|\mathbf{x}_1)\|_2^2$$

可以证明

$$\min_{\theta} L_{\text{FM}}(\theta) = \min_{\theta} L_{\text{CFM}}(\theta)$$

其中 $L_{\text{FM}}(\theta)$ 为学习边缘速度场时的目标。

3-3. 高斯条件概率路径 常用高斯条件概率路径:

$$\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad \mathbf{x}_1 \sim q(\cdot), \quad \mathbf{x} = \alpha_t \mathbf{x}_1 + \beta_t \mathbf{x}_0$$

其中 $\alpha_0 = 0, \beta_0 = 1, \alpha_1 = 1, \beta_1 = 0$ 。

其条件速度场为

$$\mathbf{v}_t(\mathbf{x}|\mathbf{x}_1) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t} \alpha_t \right) \mathbf{x}_1 + \frac{\dot{\beta}_t}{\beta_t} \mathbf{x}$$

4. 线性插值的流匹配算法 高斯条件路径中通常取 $\alpha_t = t, \beta_t = 1 - t$, 即采样线性插值路径

$$\mathbf{x} = t\mathbf{x}_1 + (1 - t)\mathbf{x}_0$$

学习的优化目标变成

$$\mathbb{E}_{t, \mathbf{x}_0, \mathbf{x}_1} \|\mathbf{u}_\theta(\mathbf{x}, t) - (\mathbf{x}_1 - \mathbf{x}_0)\|_2^2$$

采样通过数值积分

$$\mathbf{x}_{t_{k+1}} = \mathbf{x}_{t_k} + \Delta t \mathbf{u}_\theta(\mathbf{x}_{t_k}, t_k)$$

计算得到生成样本。

5. 扩散模型与概率流 ODE 考虑前向扩散过程

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t) dt + g(t) d\mathbf{w}_t$$

其概率分布满足福克-普朗克方程：

$$\frac{\partial}{\partial t} p_t(\mathbf{x}) = -\nabla_{\mathbf{x}} \cdot (p_t(\mathbf{x}) \mathbf{f}(\mathbf{x}, t)) + \frac{g^2(t)}{2} \nabla_{\mathbf{x}}^2 p_t(\mathbf{x})$$

学习得到分数场 $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ 之后，可使用概率流 ODE

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t) - \frac{1}{2} g^2(t) \nabla_{\mathbf{x}} \log p_t(\mathbf{x})$$

表示其概率分布的确定性演化过程。

继续阅读

流匹配模型的原始工作见文献 [1]，其后续的重要改进包括 Rectified Flow [2]。从方法论上看，流匹配模型与既有生成模型技术密切相关，其中包括归一化流 (Normalizing Flow) [3]、连续归一化流 (Continuous Normalizing Flow, CNF) [4]、基于 SDE 的扩散模型及其对应的概率流 ODE [5]，以及去噪扩散隐模型 (DDIM) [6]。关于流匹配模型的综述性与介绍性工作，可参见文献 [7-8]。

习题

35.1 写出流匹配模型的 ODE，并给出速度场和流的定义，回答它们在流匹配算法中的作用。

35.2 假设流 $\psi_t(\mathbf{x}_0)$ 是由 ODE 描述的微分同胚。如果从两个不同的初始位置 \mathbf{x}_0 和 \mathbf{x}'_0 在同一时刻 t 演化到了同一个位置 \mathbf{x} ，这违背了 ODE 的什么性质？

35.3 在流匹配算法中，假设路径是高斯条件路径，给出一种非线性路径设计，并讨论其可能的优缺点。

35.4 设速度场为常数向量 $\mathbf{v}_t(\mathbf{x}) = \mathbf{a}$ ，先验分布为 $p_0(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ 。
(1) 求 t 时刻的分布 $p_t(\mathbf{x})$ 的解析表达式；(2) 验证该分布满足连续性方程；
(3) 画出 $t = 0, 0.5, 1.0$ 时刻的分布（一维和二维）示意图。

35.5 流和概率流是不同的概念，列举其不同点。

35.6 证明高斯条件速度场的计算式(35.23)。

35.7 扩散过程的随机微分方程 (SDE) 的一般形式是

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + \mathbf{G}(\mathbf{x}, t)d\mathbf{w}$$

其中, $\mathbf{f}(\mathbf{x}, t)$ 是漂移系数, $\mathbf{G}(\mathbf{x}, t)$ 是扩散系数, $d\mathbf{w}$ 是标准维纳过程。漂移系数是向量, 扩散系数是矩阵。推导对应的一般形式的福克-普朗克方程。

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\nabla \cdot (\mathbf{f}(\mathbf{x}, t)p_t(\mathbf{x})) + \frac{1}{2}\nabla \cdot \nabla \cdot (\mathbf{G}(\mathbf{x}, t)\mathbf{G}^\top(\mathbf{x}, t)p_t(\mathbf{x}))$$

35.8 朗之万动力学的随机微分方程形式是

$$d\mathbf{x} = -\frac{1}{\gamma}\nabla_{\mathbf{x}}U(\mathbf{x})dt + \sqrt{\frac{2k_B T}{\gamma}}d\mathbf{w} \quad (35.30)$$

其中, γ 是摩擦系数, $U(\mathbf{x})$ 是势能函数, $\nabla_{\mathbf{x}}U(\mathbf{x})$ 代表保守力, k_B 是玻尔兹曼常数, T 是温度, $d\mathbf{w}$ 是维纳过程。写出对应的福克-普朗克方程。

参考文献

- [1] LIPMAN Y, CHEN R T, BEN-HAMU H, et al. Flow matching for generative modeling[C]//International Conference on Learning Representations, 2023.
- [2] LIU X, GONG C, et al. Flow straight and fast: Learning to generate and transfer data with rectified flow[C]//International Conference on Learning Representations, 2023.
- [3] KOBYZEV I, PRINCE S J, BRUBAKER M A. Normalizing flows: An introduction and review of current methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(11): 3964–3979.
- [4] CHEN R T, RUBANOVA Y, BETTENCOURT J, et al. Neural ordinary differential equations[C]//Advances in Neural Information Processing Systems, 2018.
- [5] SONG Y, SOHL-DICKSTEIN J, KINGMA D P, et al. Score-based generative modeling through stochastic differential equations[C]//International Conference on Learning Representations, 2021.
- [6] SONG J, MENG C, ERMON S. Denoising diffusion implicit models[C]//International Conference on Learning Representations, 2021.
- [7] LIPMAN Y, HAVASI M, HOLDERRIETH P, et al. Flow matching guide and code[J]. arXiv preprint arXiv:2412.06264, 2024.
- [8] HOLDERRIETH P, ERIVES E. An introduction to flow matching and diffusion models[J]. arXiv preprint arXiv:2506.02070, 2025.