Flow Matching II

Discussion #11

Written by:
Brent, CMK

1 Flow Matching Implementation

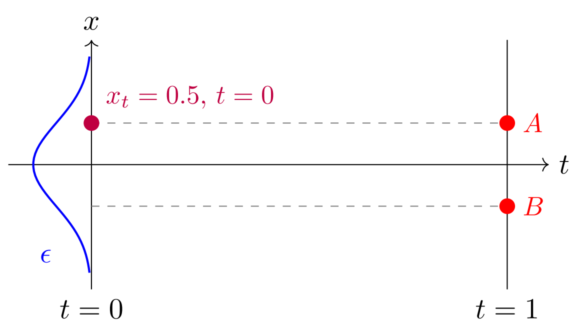
In last week's discussion, we explored geometric intuition for conditional and marginal flows. This week, we'll connect this idea to the actual implementation.

Consider a toy 1D flow matching problem. Our "dataset" has just two points: $A = +0.5$ and $B = -0.5$, each sampled with probability 0.5. We run a simple flow matching training loop:

L1: Sample $x_1 \sim$ data (A or B with prob 0.5 each)
L2: Sample $\epsilon \sim \mathcal{N}(0, 1)$
L3: Sample $t \sim$ Uniform$(0, 1)$
L4: Compute $x_t = (1 - t)\epsilon + t \cdot x_1$
L5: Compute target velocity: $u = x_1 - \epsilon$
L6: Update θ to minimize $\|u_\theta(x_t, t) - u\|^2$

For a given (x_t, t), the training loop sometimes supervises toward A and sometimes toward B.

1.1 Consider the case where $t = 0$ is sampled at L3, and $x_t = 0.5$ is computed at L4.



(a) If $x_1 = A$ was sampled at L1, what ϵ was sampled at L2? What if $x_1 = B$?

Both require $\epsilon = 0.5$. At $t = 0$, L4 gives $x_t = \epsilon$ regardless of x_1.

(b) Using L5, compute the target velocity u for each case. ($x_1 = A$ and $x_1 = B$).

For $x_1 = A$: $u = A - 0.5 = 0.5 - 0.5 = 0$
For $x_1 = B$: $u = B - 0.5 = -0.5 - 0.5 = -1$

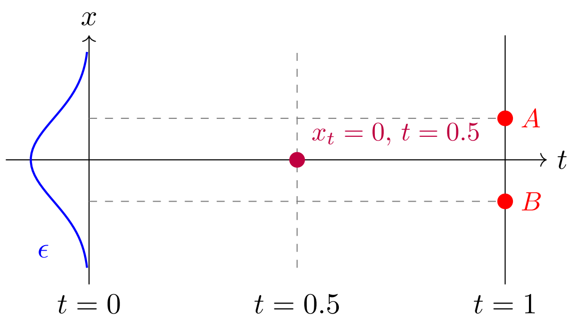
(c) How often does the training loop produce $(x_t = 0.5, t = 0, x_1 = A)$ vs. $(x_t = 0.5, t = 0, x_1 = B)$?

Equally often—both need the same ϵ, and L1 picks A vs B with 50/50 probability.

(d) What does $u_\theta(0.5, 0)$ converge to after training? Which x_1 does this velocity vector point at?

$u_\theta(0.5, 0) = \frac{1}{2}(0) + \frac{1}{2}(-1) = -0.5$.
Starting from $x_0 = 0.5$, this velocity vector points at $x_1 = 0.5 + (-0.5) = 0$.

1.2 Consider the case where $t = 0.5$ is sampled at L3, and $x_t = 0$ is computed at L4.



(a) If $x_1 = A$ was sampled at L1, what ϵ was sampled at L2? What if $x_1 = B$?

For A: $\epsilon = -0.5$. For B: $\epsilon = +0.5$. (Both equally likely under $\mathcal{N}(0, 1)$.)

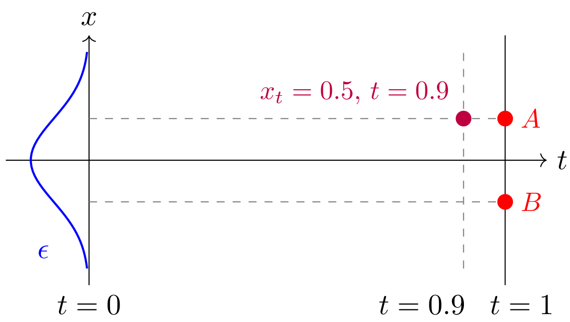
(b) Using L5, compute the target velocity u for each case.

For $x_1 = A$: $u = A - (-0.5) = 0.5 + 0.5 = 1$
For $x_1 = B$: $u = B - 0.5 = -0.5 - 0.5 = -1$

(c) What does $u_\theta(0, 0.5)$ converge to?

$u_\theta(0, 0.5) = \frac{1}{2}(1) + \frac{1}{2}(-1) = 0$

1.3 Consider the case where $t = 0.9$ is sampled at L3, and $x_t = 0.5$ is computed at L4.



(a) If $x_1 = A$ was sampled at L1, what ϵ was sampled at L2? What if $x_1 = B$?

For A: $\epsilon = 0.5$ (common). For B: $\epsilon = 9.5$ (essentially never sampled).

(b) Using L5, compute the target velocity u for each case.

For $x_1 = A$: $u = A - 0.5 = 0.5 - 0.5 = 0$
For $x_1 = B$: $u = B - 9.5 = -0.5 - 9.5 = -10$

(c) How often does the training loop produce supervision toward A vs. toward B at this (x_t, t)?

Almost always A. The B case requires $\epsilon = 9.5$, which is > 9 standard deviations from the mean.

(d) What does $u_\theta(0.5, 0.9)$ converge to?

$u_\theta(0.5, 0.9) \approx 0$ (the A case dominates)

1.4 Both 1.1 and 1.3 have $x_t = 0.5$. Why does u_θ converge to -0.5 at $t = 0$ but ≈ 0 at $t = 0.9$?

At $t = 0$, L4 computes $x_t = \epsilon$ regardless of x_1, so both targets appear equally often.
At $t = 0.9$, reaching $x_t = 0.5$ with target B requires an implausible ϵ at L2, so target A dominates.
The network learns a weighted average of target velocities. The weights depend on how likely each (x_t, t, x_1) combination is under the training distribution.

2 Mixing and Matching Models and Samplers

Suppose you have a pre-trained **ϵ-prediction model** $\epsilon_\theta(x_t, t)$ that predicts noise (like DDPM), and a sampler that implements **Euler integration** for flow matching:

$$x_{t+\Delta t} = x_t + \Delta t \cdot u_\theta(x_t, t)$$

The sampler expects a velocity u, but your model outputs ϵ.

2.1 Using the flow matching interpolation $x_t = (1-t)\epsilon + t \cdot x_{\text{clean}}$ and velocity definition $u = x_{\text{clean}} - \epsilon$, derive a formula for u in terms of x_t, ϵ, and t.

From the two equations:

$$\begin{aligned} x_t &= (1-t)\epsilon + t \cdot x_{\text{clean}} \\ u &= x_{\text{clean}} - \epsilon \end{aligned}$$

From the second equation: $x_{\text{clean}} = u + \epsilon$.
Substituting into the first:

$$\begin{aligned} x_t &= (1-t)\epsilon + t(u + \epsilon) \\ x_t &= (1-t)\epsilon + tu + t\epsilon \\ x_t &= \epsilon + tu \end{aligned}$$

Therefore: $\epsilon = x_t - tu$, which gives us $\boxed{u = \frac{x_t - \epsilon}{t}}$

2.2 Rewrite the Euler update step using ϵ_θ instead of u_θ.

Substituting $u = \frac{x_t - \epsilon}{t}$:

$$\boxed{x_{t+\Delta t} = x_t + \Delta t \cdot \frac{x_t - \epsilon_\theta(x_t, t)}{t}}$$