# Constructing a training target for Flow & Diffusion model.

Goal: Derive training algorithm. (Get a "good" vector field).

- Minimize MSE.

$$\mathcal{L}(\theta) = \| u_t^{\theta}(x) - \underbrace{u_t^{\text{target}}(x)}_{\text{Training target.}} \|^2$$

In linear reg, training target = label.

But what is it in our setting?

Today: Get a training target.

Some def:

Make sure you **understand the formulas** for:

| | | |
|---|---|---|
| **Conditional Probability Path** | **Conditional Vector Field** | **Conditional Score Function** |
| **Marginal Probability Path** | **Marginal Vector Field** | **Marginal Score Function** |

Conditional: "Per single data pt."
Marginal: "Across distribution of data pts."

Probability Paths. The path from Noise to Data.

Dirac Distribution: $z \in \mathbb{R}^d$. $X \sim \delta_z \Rightarrow X = z$.

Conditional prob path: $\boxed{p_t(\cdot|z) \quad \text{for fixed } z}$

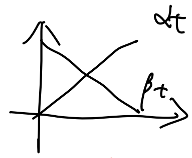
① $p_t(\cdot|z)$ is a distribution over $\mathbb{R}^d$.

② $p_0(\cdot|z) = p_{init}$. $\quad p_1(\cdot|z) = \delta_z$.

Eg. Gaussian prob path.

$p_t(\cdot|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$

Noise schedulers: $\alpha_t, \beta_t$. s.t. $\alpha_0 = 0 \quad \beta_0 = 1$
$\alpha_1 = 1. \quad \beta_1 = 0.$




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

Let's check it fullfill the requirement above,

$p_0(\cdot|z) = \mathcal{N}(\alpha_0 z, \beta_0^2 I_d) = \mathcal{N}(0, I_d)$

$p_1(\cdot|z) = \mathcal{N}(\alpha_1 z, \beta_1^2 I_d) = \mathcal{N}(z, 0) = \delta_z.$ (Only 1 pt)

---

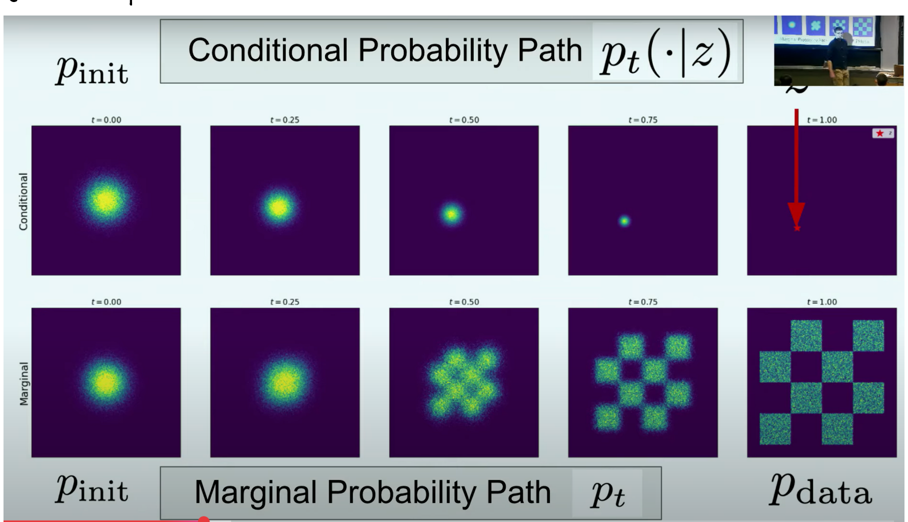
Marginal prob path: $\boxed{p_t(\cdot)}$ (for random $z \sim p_{data}$)

Sample $z \sim p_{data}$, $x \sim p_t(\cdot|z) \Rightarrow x \sim p_t.$

Forget $z$.

① $p_t(x) = \int p_t(x|z)\, p_{data}(z)\, dz.$

② $p_0 = p_{init}. \quad p_1 = p_{data}.$



## 3.2 Cond & Marginal Vector Field.

Conditional vector field.

$u_t^{target}(x|z)$ $\quad 0 \le t \le 1,\ x, z \in \mathbb{R}^d.$

$$p_{init} \xrightarrow{p_t(\cdot|z)} \delta_z \qquad p_{init} \xrightarrow{\text{ODE.}} \delta_z$$

s.t.

$$\boxed{\underbrace{X_0 \sim p_{init}}_{= p_0(\cdot|z).} \quad \frac{d}{dt} X_t = u_t^{target}(X_t|z) \Rightarrow X_t \sim p_t(\cdot|z) \quad (0 \le t \le 1)}$$

Follow the simulated ODE, the we got it.

$\cdot Z.$ $\quad p(x) \approx \sum_z p(x|z)\, p(z).$

Eg. For Gaussian

$$p_t(\cdot|z) = \mathcal{N}(\alpha_t z, \beta_t^2 I_d)$$

<u>Cond Gaussian VF</u> :

$$\underbrace{u_t^{target}(x|z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right) z + \frac{\dot{\beta}_t}{\beta_t} x}_{\text{A comb of } x \text{ \& } z.}$$

$$\dot{\alpha}_t = \frac{d}{dt}\alpha_t. \qquad \dot{\beta}_t = \frac{d}{dt}\beta_t.$$

<u>Thm. Marginalization Trick:</u>

The marginal Vector field by

$$\boxed{u_t^{target}(x) = \int u_t^{target}(x|z) \frac{p_t(x|z)\, p_{data}(z)}{p_t(x)}\, dz.}$$

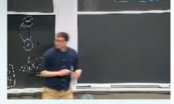
satisfies

$$\boxed{X_0 \sim p_{init},\ \frac{d}{dt} X_t = u_t^{target}(X_t) \Rightarrow X_t \sim p_t. \ (0 \le t \le 1)} \Rightarrow X_1 \sim p_{data}.$$

Just for knowledge:

## Continuity Equation

*Randomly initialized ODE*



**Given:** $X_0 \sim p_{\text{init}}, \quad \frac{\mathrm{d}}{\mathrm{d}t} X_t = u_t(X_t)$

---

**Follow probability path:**

$$X_t \sim p_t \quad (0 \le t \le 1)$$

*Marginals are p_t*

equivalent

**Continuity equation holds**

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t(x) = -\mathrm{div}(p_t u_t)(x)$$

*PDE holds*

Random in'it ODE $\Longleftrightarrow$ PDE

$$\mathrm{div}(V_t)(x) = \sum_{i=1}^{d} \frac{\partial}{\partial x_i} V_t(x) \qquad V_t : \mathbb{R}^d \to \mathbb{R}^d.$$

$$\underbrace{\frac{d}{dt} p_t(x)}_{\substack{\text{Change of} \\ \text{probability mass at } x}} = \underbrace{-\mathrm{div}(p_t u_t)}_{\text{Outflow} - \text{inflow.}}$$

pf of Marginalization:

$$\frac{d}{dt} p_t(x) = \frac{d}{dt} \int p_t(x|z)\, p_{data}(z)\, dz = \int \frac{d}{dt} p_t(x|z)\, p_{data}(z)\, dz$$

$$= \int -\mathrm{div}\left(p_t(\cdot|z)\, u_t^{target}(\cdot|z)\right)(x)\, p_{data}(z)\, dz$$

$$= -\mathrm{div}\left( \int p_t(x|z)\, u_t^{target}(x|z)\, p_{data}(z)\, dz \right)$$

(Continuity Equation is a prior knowledge).

$$= -\mathrm{div}\left( p_t(x) \int u_t^{target}(x|z) \frac{p_t(x|z)\, p_{data}(z)}{p_t(x)}\, dz \right)$$

$$= -\mathrm{div}\left(p_t u_t^{target}\right)(x)$$

(As we know)

$$\Rightarrow u_t^{target}(x) =$$

For Cond Gaussian VF

$$u_t^{\text{target}}(x|z) = \left(\dot{\alpha}_t - \frac{\dot{\beta}_t}{\beta_t}\alpha_t\right) z + \frac{\dot{\beta}_t}{\beta_t} x$$

Plug this in, $u_t^{\text{target}}(z) = \int \text{---} \; u_t^{\text{target}}(x|z)\, dz$.

---

## 3.3. Cond & Marginal Score.

Conditional Score: $\nabla_x \log p_t(x|z)$

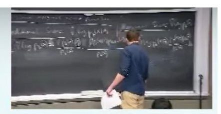
Marginal Score: $\nabla_x \log p_t(x)$

Formula: $$\nabla_x \log p_t(x) = \frac{\nabla_x p_t(x)}{p_t(x)} = \frac{\nabla_x \int p_t(x|z) p_{\text{data}}(z)\, dz}{p_t(x)} = \frac{\int \nabla_x p_t(x|z) p_{\text{data}}(z)\, dz}{p_t(x)}$$

$$= \int \nabla_x \log p_t(x|z) \frac{p_t(x|z)\, p_{\text{data}}(z)}{p_t(z)}\, dz$$

same trick.

**Marginal Prob. Path, Vector Field, and Score**



| | Notation | Key property | Formula |
|---|---|---|---|
| Marginal Probability Path | $p_t$ | Interpolates $p_{\text{init}}$ and $p_{\text{data}}$ | $\int p_t(x|z) p_{\text{data}}(z) \mathrm{d}z$ |
| 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 ODE target to SDE | $\int \nabla \log p_t(x|z) \frac{p_t(x|z) p_{\text{data}}(z)}{p_t(x)} \mathrm{d}z$ |

Eg. For Gaussian Path, $p_t(x|z) = \frac{1}{\sqrt{2\pi}\beta_t} \exp\left(-\frac{(x-\alpha_t z)^2}{2\beta_t^2}\right)$ $\qquad \alpha_t z$, $\beta_t^2 I_d$

Cond Score $= \nabla_x \log p_t(x|z) = \nabla_x \left(-\frac{(x-\alpha_t z)^2}{2\beta_t^2} - \log(\sqrt{2\pi}\beta_t)\right) = \boxed{-\frac{x-\alpha_t z}{\beta_t^2}}$ Gaussian Score.

## Why score is Useful?

$\boxed{\text{Thm. SDE extension trick.}}$

Let $u_t^{target}(x)$ be as before. Then for any $\sigma_t \geq 0$:

$$X_0 \sim p_{init}, \quad dX_t = \Big[u_t^{target}(X_t) + \underbrace{\frac{\sigma_t^2}{2}\nabla \log p_t(X_t)}_{\text{Add correction term.}}\Big]dt + \underbrace{\sigma_t \, dW_t}_{\text{diffusion term}}$$

$$\Rightarrow X_t \sim p_t.$$

$$\Rightarrow X_1 \sim p_{data}.$$