Rectified Flow (RF) Series: Discrete Normalizing Flows (DNF)s

Imagine wwt turn a simple distribution into a more complex one

Apply a sequence of invertible tranformations

$\{f_1, \text{———}, f_k\}, \ f_i : z_{i-1} \to z_i$ $\qquad$ $z_k = f_k \circ f_{k-1} \circ \text{———} \circ f_1(x)$

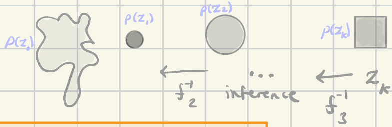
image $x = z_0, \ z_k \in \mathbb{R}^d, \ z_k \sim N(0, I)$ $\qquad$ $x = f_k^{-1} \circ f_{k-1}^{-1} \circ \text{———} \circ f_1^{-1}(z_k)$

$p(x) \ p(z)$

Let $p_x(x), p_z(z)$ be PDFs

$\Rightarrow$ Since f_i invertible,

$x = z_0 \xrightarrow{f_1} z_1 \xrightarrow{f_2} z_2 \ \cdots \xrightarrow{f_3} z_k$



Discrete change-of-variables (cov) (and log cov) formula

$$\log p(x) = \underbrace{\log p(z_k)}_{\text{Gaussian}} + \sum \log \left| \det \frac{\partial f_i}{\partial h_{i-1}} \right| \quad (A)$$

How do we get (A)? Specifically, the determinant

Aside: Probability and PDFs

Recall, a pdf by itself isn't a prob — it's "probability mass per infinitesimal volume"

So, in 1D $Pr(x \le X \le x + dx) \approx p_x(x)\,dx$ where dx is a small interval length

Say X is height $Pr(X = 193cm) = 0$

Now, the pdf has a value

$p_x(193) = 0.02$ (per cm) ← mass per unit volume

and to find the actual Prob, we evaluate the integral

$$Pr(a \le X \le b) = \int_a^b p_x(x)\,dx$$

Jacobian Matrix

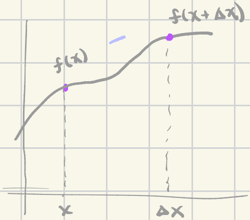
n=m in our case

DEFn for $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ $\quad f(z) = (f_1(z), \text{—}, f_m(z))$

$f(x+\Delta x) \approx f(x) + J_f(x)\Delta x$

the Jacobian matrix of f at a point x is

$$J_f(z) = \begin{pmatrix} \frac{\partial f_1}{\partial z_1} & \cdots & \frac{\partial f_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial z_1} & \cdots & \frac{\partial f_m}{\partial z_n} \end{pmatrix}$$



now make $\Delta x \ll 0$

Each row is a gradient of 1 output coordinate wrt inputs

Geometric meaning

It's the best linear approximation of f near the point z

iow $f(z + \Delta z) \approx f(z) + J_f(z)\Delta z$

Jacobian = local linearization of non linear map

$\frac{\partial f_1}{\partial z_1}$ what is the instantaneous rate of change f_1 wrt z_1. It's the slope of the mapping in dimension 1

stretch / compression factor

when $m=n \rightarrow |\det J_f(z)|$ = volume scaling factor
Not actual volume!

$|\det J_f(z)|$ is the factor by which an infinitesimal cube of space around z is stretched or squashed when mapped through f

Mass conservation:
If volume expands/contracts, density must compensate.

Geometrically,
if length doubles then density halves

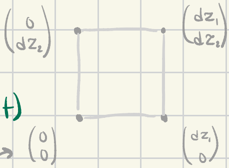
Density: Mass per unit volume
volume: The size of the region in question
mass: density $\times$ volume

For a linear map $A : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ (square matrix)
$|\det(A)|$ = factor by which A scales d-dimensional volumes
So if you apply A to any d-dimensional region its volume changes by exactly that factor

Area, determinant & Jacobian

Imagine a point at
$z = (z_1, z_2)$
The two "side" (displacement) vectors are

$$\begin{bmatrix} dz_1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ dz_2 \end{bmatrix}$$



$z = (z_1, z_2)$

$$\boxed{f(x+\Delta x) \approx f(x) + J_f(x)\Delta x}$$

Now apply $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$

$\Longrightarrow$ The two side vectors get transformed into

$$\left\{ \underbrace{J_f(z)\begin{bmatrix} dz_1 \\ 0 \end{bmatrix}}_{u}, \underbrace{J_f(z)\begin{bmatrix} 0 \\ dz_2 \end{bmatrix}}_{v} \right\}$$

span a parallelogram

what's the area of a parallelogram?

$$\text{Area} = |u_1 v_2 - u_2 v_1| = |\det[u, v]|$$

$$\det(AB) = \det(A)\det(B)$$

$$(u, v) = J_f(z) \begin{pmatrix} dz_1 & 0 \\ 0 & dz_2 \end{pmatrix}$$

$$\det|(u,v)| = \det(J_f(z)) \det \begin{pmatrix} dz_1 & 0 \\ 0 & dz_2 \end{pmatrix}$$

$$\boxed{\text{Area} = \det|(u,v)| = |\det J_f(z)| \, dz_1 dz_2}$$

Back to DNFs. How do we get (A)?

$$\lg p(x) = \lg p(z) - \sum_{i=1}^{k} \lg \left| \det \frac{\partial f_i}{\partial h_{i-1}} \right| \quad (A)$$

Change of variables theorem

$$\int_A p_X(x)\,dx = \int_{f^{-1}(A)} p_Z(z)\,dz$$

for any measurable set A.

Shrinking A down to an infinitesimal interval we get

$$p_X(x)\,dx \approx p_Z(z)\,dz$$

But folks write

$$\boxed{p_X(x)\,dx = p_Z(z)\,dz}$$

the differential form of the formal integral

$f: z \to x$

$$\boxed{\begin{gathered} x = f(z) \\ \frac{dx}{dz} = \frac{df}{dz} \end{gathered}}$$

Case 2D

Now $z \in \mathbb{R}^2$, $x = f(z)$ $\quad f: z \longrightarrow x$

- a little square patch in z-space has area $dz_1 dz_2$
- It maps under f to a parallelogram in x-space with area:

$$dx_1 dx_2 = \left| \det \frac{\partial f}{\partial z} \right| dz_1 dz_2$$

$$\Longrightarrow \quad \frac{dz_1 dz_2}{dx_1 dx_2} = \left| \det \frac{\partial f}{\partial z} \right|^{-1}$$

Probability Conservation:

$$\rho_X(x)\, dx_1 dx_2 = \rho_Z(z)\, dz_1 dz_2$$

So $\rho_X(x) = \rho_Z(z) \dfrac{dz_1 dz_2}{dx_1 dx_2}$

$$= \rho_Z(z) \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$$

$$= \rho_Z(z) \left| \det J_f(z) \right|^{-1}$$

Inverse form

$$= \rho_Z(f^{-1}(x)) \left| \det J_{f^{-1}(x)}(z) \right|$$

Higher dimensions

For $z \in \mathbb{R}^d$, a small hypercube volume $dz_1 \text{---} dz_d$ maps to a distorted parallelepiped in x-space

It's volume is scaled by

$$dx_1 \text{---} dx_d = \left| \det \frac{\partial f(z)}{\partial z} \right| dz_1 \text{---} dz_d$$

Prob Conservation

$$\rho_X(x)\, dx = \rho_Z(z)\, dz$$

General formula is

$f: Z \longrightarrow X$

$$p_X(x) = p_Z(z) \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$$

$\Longrightarrow$ DNFs

$z \sim N(0, I)$ and $x = f_K \circ \text{———} \circ f_1(z)$

f invertible $\rightarrow$ you can compute densities by (COV) formula

$$\lg p(x) = \lg p(z) - \sum_{i=1}^{K} \log \left| \det \frac{\partial f_i}{\partial h_{i-1}} \right| \quad (B)$$

where h_{i-1} is an intermediate latent

How do we get (B) using (COV)?

Compose multiple steps

$z_0 \xrightarrow{f_1} z_1 \xrightarrow{f_2} z_2 \cdots \xrightarrow{f_K} z_K \sim N(0, I)$

data

$$p(z_1) = p(z_0) \det \left| \frac{\partial f_1}{\partial z_0} \right|^{-1}$$

$$p(z_2) = p(z_1) \det \left| \frac{\partial f_2}{\partial z_1} \right|^{-1}$$

substitute recursion ...

$$\Longrightarrow \quad p(z_K) = p(z_0) \prod_{i=1}^{K} \det \left| \frac{\partial f_i}{\partial z_{i-1}} \right|^{-1}$$

solve for $p(z_0)$ $\Longrightarrow$ $p(z_0) = p(z_K) \prod_{i=1}^{K} \det \left| \frac{\partial f_i}{\partial z_{i-1}} \right|$

let our data $x = z_0$

$$\Longrightarrow \quad p(x) = p(z_K) \prod_{i=1}^{K} \det \left| \frac{\partial f_i}{\partial z_{i-1}} \right|$$

Multiply by log to get (A)

$$\boxed{\log p(x) = \log p(z_K) + \sum \log \left| \det \frac{\partial f_i}{\partial h_{i-1}} \right| \qquad (A)}$$

How to train? $\max_{\theta} \mathbb{E}_{x \sim data} \log p_\theta(x)$

$$\ell(x;\theta) = \log p(z_K) + \sum \log \left| \det \frac{\partial f_{\theta i}}{\partial h_{i-1}} \right|$$ ← back propagate

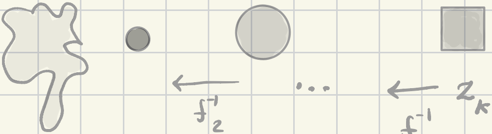
Problems: (1) f must be invertible

(2) computing determinants for big Jacobians is very expensive

Training $\quad z_K = f_K \circ f_{K-1} \circ \text{——} \circ f_1(z_0)$

Inference $\quad x = f_K^{-1} \circ f_{K-1}^{-1} \circ \text{——} \circ f_1^{-1}(z_K)$

$$x = z_0 \xrightarrow{f_1} z_1 \xrightarrow{f_2} z_2 \cdots \xrightarrow{f_3} z_K$$



$$\xleftarrow{f_2^{-1}} \cdots \xleftarrow{f_3^{-1}} z_K$$