

GGI ML Workshop

Intro to Normalizing Flows and Some applications to LHC & GaiaMonday Sept 12, 2022

By now many talks on NFs & their applications to many domains
Clearly a hot topic!

Each talk - very brief descr of NFs

Aim here: more pedagogical introduction to NFs

+ applications to LHC & Gaia I've worked on
as time permits.

NFs are powerful new tools ~~to~~ for density estimation
and generative modeling



DE: Given data ~~points~~ $D = \{\vec{x}_i\}$ $\rightarrow$ what is underlying $p(x)$ they were drawn from?

Conditional DE: Also: given $\{(\vec{x}_i, \vec{y}_i)\}$ $\rightarrow$ what is $p(x|y)$?

Generative modeling: can sample from $p(x)$ & produce more data that ^{or} looks like D ?

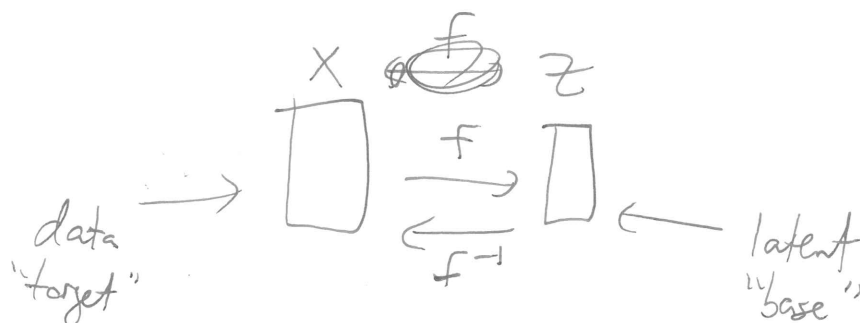
(2)

Note: some approaches $\nabla O \in$ but not gen modely
 " " gen modely but no $O \in \rightarrow p(x)$ only known explicitly
 (GANs, VAEs)

- ↓
- NFs ~~provide~~ one reason so powerful — can do both in principle!

- Will not be able to cover every aspect of NFs, more advanced topics
 (e.g. w/ symmetries — see Unis's talk)
 (or score based models — more recent variant)
 ↓
 just some "basics"

- Idea of NF: learn ^{smooth,} invertible transf. $x \rightarrow z = f(x)$



w/ $z \sim$ simple dist'n e.g. unif or gaussian

$$p(x) dx = p(z = f(x)) dz$$

$$\Rightarrow \boxed{p(x) = p(z = f(x)) \left| \frac{\partial z}{\partial x} \right|} = p(z = f(x)) \left| \frac{\partial f}{\partial x} \right|$$

specific of base dist'n

Jacobian of transf.
 need tractable Jac! more soon.

- Given f : $\begin{cases} z = f(x) & : \text{density est.} \\ x = f^{-1}(z) & : \text{gen modeling (samples)} \end{cases}$

- Loss fn for NFs: since NF provides access to $p(x)$ can train w/ MLE!

$$L = - \sum_{x_i \in D} \log p(x_i | \theta)$$

← pars of f

→ optimal
→ stable
→ convergent

and get metric for quality of density est. model selection.

← much improved vs GANs!

$$\theta^* = \underset{\theta}{\operatorname{argmin}} L$$

note: in some settings, no data for samples to train from (eg Latticed) but instead have $p(x)$ or $z \cdot p(x)$ already train gen NF as gen modeling KL div, or somethg like it

- Need a family of fns f that is

- expressive

- invertible

- tractable Jacobian

$$\rightarrow \left| \frac{\partial f}{\partial x} \right| \quad x \in \mathbb{R}^d$$

↳ $\mathcal{O}(d^3)$ operations generically

Key Idea: ^{expressive} build ~~compose~~ f out of sequence of simpler transformations

"Flow"
gradually transform x to z

$$f = f_1 \circ f_2 \circ f_3 \circ \dots$$

$$\det J_f = \det J_{f_1} \det J_{f_2} \dots$$

Another key idea:

$f \rightarrow$ Autoregressive transformation

$$\begin{cases} z_1 = f_1(x_1) \\ z_2 = f_2(x_1, x_2) \\ z_3 = f_3(x_1, x_2, x_3) \\ \vdots \end{cases}$$

$$\frac{\partial z_i}{\partial x_j} = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

$$\det \frac{\partial z_i}{\partial x_j} = \prod \frac{\partial z_i}{\partial x_i}$$

$\rightarrow$ So $\mathcal{O}(d^3) \rightarrow \mathcal{O}(d)$
determinant becomes tractable!

- Invertibility: if f_n invertible as fn of $x_n \forall x_1, \dots, x_{n-1}$ then f is invertible

$$\text{i.e.: } \begin{cases} f_1^{-1}(z_1) = x_1 \\ f_2^{-1}|_{x_1}(z_2) = x_2 \\ f_3^{-1}|_{x_1, x_2}(z_3) = x_3 \\ \vdots \end{cases} \rightarrow \text{in that sense, have } p(x_n | x_1, \dots, x_{n-1})$$

from $z_n \sim \mathcal{N}(0, 1)$

$\rightarrow$ autoregressive property.

- Popular (and effective) class of transf.:

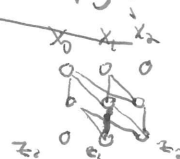
"affine"

$$z_n = a_n(x_1, \dots, x_{n-1})x_n + b_n(x_1, \dots, x_{n-1}) \quad \text{as long as } a_n \neq 0$$

this satisfies invertibility req. above!

idea of MAF: a_n & b_n given by MLPs!

$\rightarrow$ can unify into single MLP $a_n \rightarrow \vec{a}, b_n \rightarrow \vec{b}$ w/ masking $\rightarrow$ FAST



mult. by binary mtx

$$x_n = \frac{z_n - b_n}{a_n}$$

$$\Delta x_n = \sigma_n(x_1, \dots, x_{n-1}) z_n + \mu_n(x_1, \dots, x_{n-1})$$

$$\text{So } \left[P(x_n | x_1, \dots, x_{n-1}) = \mathcal{N}(\mu_n(x_1, \dots, x_{n-1}), \sigma_n(x_1, \dots, x_{n-1})) \right]$$

"mixture density network"

- each step of NF is a MDN!

~~• $\sigma_n(x_1, \dots, x_{n-1}), \mu_n(x_1, \dots, x_{n-1})$ given by DNNs~~

$$z_n = \sigma_n(x_1, \dots, x_{n-1}) \cdot z = a(x)x + b(x) \quad x \rightarrow z \text{ fast one call to MLP}$$

$$x = \sigma(x)z + \mu(x) \quad z \rightarrow x \text{ slow d calls to MLP}$$

$$x_0 = \sigma_0 z_0 + \mu_0$$

$$x_1 = \sigma_1(x_0)z_1 + \mu_1(x_0) \quad \sim d \text{ times slower!}$$

Can also define IAF ~~Reverse as~~ for $z \rightarrow x$ fast

$x \rightarrow z$ slow

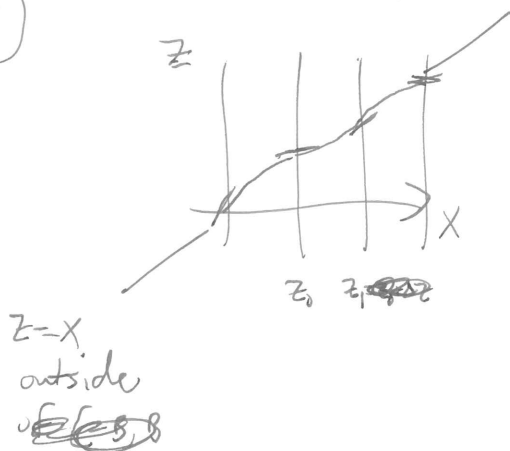
$\hookrightarrow$ log likelihood tracing becomes impossible!

	MAF	IAF
$x \rightarrow z$	slow fast	slow
$z \rightarrow x$	slow	fast

Can chain together / stack MLP/transfs
"MADE blocks"
for even more expressive transfs

Variations

- Rational Quadratic Splines (RQS)



→ more expressive than affine!

monotonic, invertible
piecewise smooth
continuous & 1st der

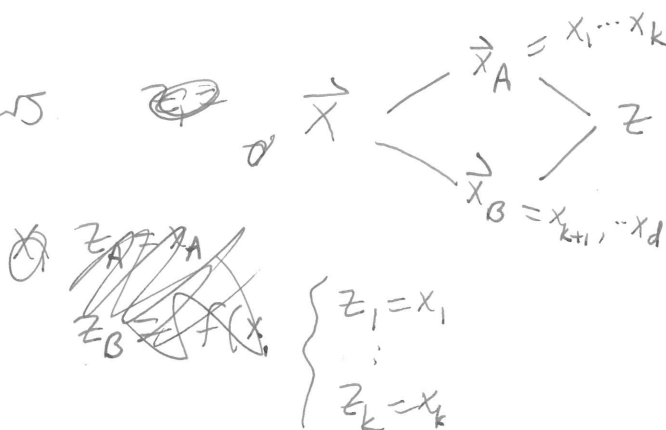
$$z = z_0 + \frac{(z_1 - z_0)}{5\xi^2 + d_0\xi(1-\xi)} \left(5\xi^2 + d_0\xi(1-\xi) \right)$$

$$5 + (d_1 + d_0 - 2d_0)\xi(1-\xi)$$

$$\xi = \frac{x - x_0}{x_1 - x_0}, \quad \eta = \frac{z_1 - z_0}{x_1 - x_0}$$

$$z_{0,1}; x_{0,1}; d_{0,1} \leftarrow \text{MLPs}$$

- Coupling layers



equally fast neither dir
since $x_i = z_i$!

$$z_{k+1} = a(x_1 \dots x_k) x_{k+1} + b(x_1 \dots x_k)$$

$\vdots$

$z_d = a$

but much less expressive than MAF/IAF.



Applications of NFs $\rightarrow$ LHC, Gaia (in work)

LHC $\rightarrow$ everyone knows

Gaia $\rightarrow$ ESA satellite 2013 $\rightarrow$ 2025
aims to ^{produce most complete} ~~map~~ catalog MW stars - full 6d phase space
DR3 - $\approx 1.5B$ sources w/ $(\alpha, \delta, \mu_\alpha, \mu_\delta, G, B, R)$
much smaller # ($\approx 33M$) w/ ~~parallax & rv~~
 $\swarrow$ full 6d phase space (parallax, rv)
 $\sim 3 \text{ kpc}$
 $\rightarrow$ interesting apps to DM.

	Generative Modeling/ Density Est	Anomaly Detection
LHC	CaloFlow	ANODE, CATMODE
Gaia	Galactic mass density Upsampling Gaia rock catalogs	Via Machinae (ANODE for stellar streams)

Explain/sketch idea of each project