GGI ML Workshop

Intro to Normalizing Flows and Some applications to LHC & Gaia

Monday 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 for density estimation
ad generative modeling

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

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

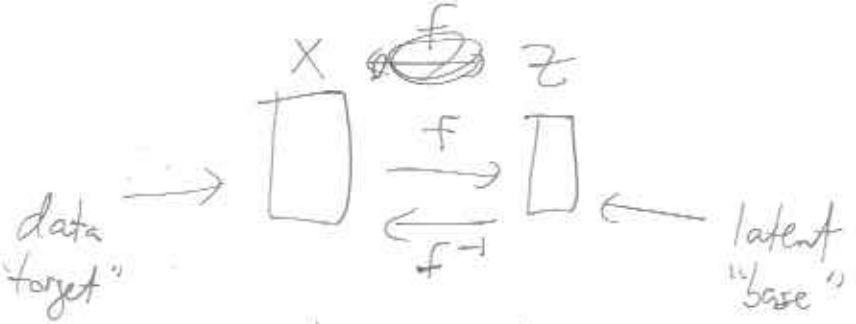
Generative modeling: can sample from $p(x)$ & produce more data that looks like D?

Note: some approaches give DE but not gen modeling
" gen modeling but no DE → p(x) only known explicitly (GANs, VAEs)

- NFs ~~promise~~ 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 Max's talk) (or score based models more recent variant)
  ↓ just some "basics"

- Idea of NF: learn smooth, invertible transf. $x \longrightarrow z = f(x)$



w/ $z \sim$ simple distn e.g. unif or gaussian

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

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

specified base distn

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 set metric for quality of density est. ← much improved vs GANs!
model selection.

$$\theta^* = \arg\min_\theta L$$

note: in some settings, models see samples to train from (e.g. Lattice QCD) but instead have $p(x)$ or $\tilde{z} \cdot p(x)$ already train NFs as gen model using KL div. or something like it

- Need ∅ a family of fns f that is
  - expressive
  - invertible
  - tractable Jacobian → $\left|\frac{\partial f}{\partial x}\right|$ $x \in \mathbb{R}^d$
    ↳ $O(d^3)$ operations generically

Key Idea: build expressive f out of sequence of simpler transformations

$f = f_1 \circ f_2 \circ f_3 \circ \cdots$ $\det J_f = \det J_{f_1} \det J_{f_2} \cdots$

"flow" gradually transform x to z

Another key idea:

$f \rightarrow$ Autoregressive transformation

$$\begin{cases} z_1 = f_1(x_1) \\ z_2 = f_2(x_1, x_2) \\ z_3 = f_3(x_1, x_2, x_3) \\ \vdots \end{cases} \longrightarrow \frac{\partial z_i}{\partial x_j} = \begin{pmatrix} \ddots & 0 \\ \vdots & \ddots \end{pmatrix}$$

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

So $O(d^2) \rightarrow O(d)$

$\longrightarrow$ determinant becomes tractable!

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

i.e.:
$$\begin{cases} f_1^{-1}(z_1) = x_1 \\ f_2^{-1}\big|_{x_1}(z_2) = x_2 \\ f_3^{-1}\big|_{x_1, x_2}(z_3) = x_3 \\ \vdots \end{cases}$$

$\rightarrow$ in that sense, have $p(x_n \mid x_1, \dots x_{n-1})$ from $z_n \sim N(0,1)$

$\rightarrow$ autoregressive property.

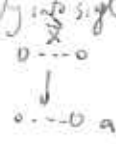
- Popular (and effective) class of transf!

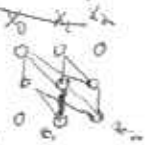
"affine" $z_n = a_n(x_1, \dots x_{n-1}) x_n + b_n(x_1, \dots x_{n-1})$ as long as $s_n \neq 0$ this satisfies invertibility req. above!

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/ mask

$\rightarrow$ FAST





$$\begin{cases} x_1 = \frac{z_1 - b_1}{a_1} \\ x_n = \sigma_n(x_1, \dots x_{n-1}) z_n + \mu_n(x_1, \dots x_{n-1}) \end{cases}$$

So $\boxed{p(x_n | x_1, \dots x_{n-1}) = \mathcal{N}(\mu_n(x_1, \dots x_{n-1}), \sigma_n(x_1, \dots x_{n-1}))}$

"mixture density network"

- each step of NF is a MDN!
- ~~each $p_n(x_n | x_1, \dots x_{n-1})$ given by RNNs~~

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

- $x = \sigma(x) z + \mu(x)$ $z \to x$ slow
d calls to MLP

$x_0 = \sigma_0 z_0 + \mu_0$

$x_1 = \sigma_1(x_0) z_1 + \mu_1(x_0)$

~ d times slower!

Can also define IAF ~~reverse~~ for $z \to x$ fast

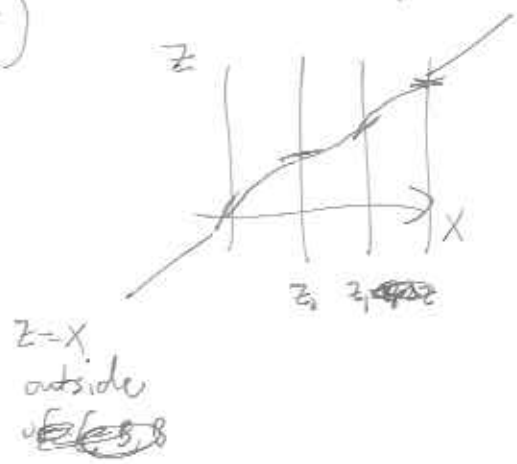
$x \to z$ slow
↳ log likelihood training becomes impossible!

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

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

Variations

- Rational Quadratic Splines (RQS) → more expressive than affine!



$z = x$ outside

monotonic, invertible
piecewise smooth
continuous & 1st der

$z = z_0 + (z_1 - z_0)\left(s\xi^2 + d_0\xi(1-\xi)\right)$

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

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

$z_{0,1}$; $x_{0,1}$; $d_{0,1}$ ← MLPs

- Coupling layers

$z_1 = x_1$
$\vdots$
$z_k = x_k$

$z_{k+1} = a_1(x_1 \cdots x_k)\, x_{k+1} + b_1(x_1 \cdots x_k)$
$\vdots$
$z_d = a$

← equally fast neither dir since $x_i = z_i$!

- but much less expressive than MAF/IAF.

Applications of NFs → LHC, Gaia (my work)

LHC → everyone knows

Gaia → ESA satellite 2013 → 2025+

aims to ~~complete~~ (almost complete) catalog MW stars — ~~full 6d phase space~~

DR3 — ≈ 1.5B sources w/ $(\alpha, \delta, \mu_\alpha, \mu_\delta, G, B, R)$

much small # (≈ 33M) w/ ~~parallax~~ full 6d phase space (parallax, rv)

↓

~3 kpc

→ interesting apps to DM.

	Generative Modeling / Density Est	Anomaly Detection
LHC	CaloFlow	ANODE, CATHODE
Gaia	Galactic mass density Upsampling Gaia mock catalogs	VicMachinae (ANODE for stellar streams)

Explain/sketch idea of each project