Wednesday, April 21, 2021 10:15 AM

Lecture 24

Last time (Claudius)

- Idea of density estimation — learn $p(x)$ from data
- Simple approaches
  - histograms
  - KDE
  - GMM
- Modern approaches — Normalizing Flows

$z, x \in \mathbb{R}^d$

$$z \xrightarrow{f} x$$

"base" dist'n e.g. $z \sim N(0,1)$ or uniform$(0,1)$

"target" dist'n $x \sim p_{data}(x)$.

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

f must be
- piecewise diff.
- invertible

Wednesday, April 21, 2021 10:30 AM

If we have $\begin{cases} f: z \to X & \longrightarrow \text{sample} \\ f^{-1}: x \to z & \longrightarrow \text{density estimation} \end{cases}$

(assuming $\left|\frac{\partial f^{-1}}{\partial x}\right|$)

<u>NEED</u>: - f invertible

- det J fast compute Jacobian

$\downarrow$ d×d matrix $\downarrow$ ~~fastest~~ algo ~ d^3.

Recent breakthroughs (2015+)

- f's can be composed! $f = f_1 \circ f_2 \circ f_3 \cdots$ still bijective

$$\det J_f = (\det J_{f_1} \cdot \det J_{f_2} \cdots)$$

could try to build expressive f out of simple ones (flow idea)

Wednesday, April 21, 2021 10:37 AM

E.g.

- planar flow

nonlinearity like tanh.

$\in \mathbb{R}$

$$f(z) = z + u\,h(w^T z + b)$$

weights $\in \mathbb{R}^d$

Jacobian is not d^3, fast

invertible given some conditions on h.

- radial flow

$$f(z) = z + \beta \frac{z - z_0}{\alpha + \|z - z_0\|}$$

learnable weights

good first step

not very expressive

not SOTA. for higher dim'l problems.

Wednesday, April 21, 2021 10:43 AM

Autoregressive Flows

- First, solve $\mathcal{O}(d^3)$ det problem w/ following trick

$$x_1 = f_1(z_1)$$
$$x_2 = f_2(z_1, z_2)$$
$$\vdots$$
$$x_i = f_i(z_1, z_2, \cdots z_i)$$
$$\vdots$$
$$x_d = f_d(z_1, \cdots z_d)$$

Then $\frac{\partial f}{\partial z} = \begin{pmatrix} \ast & \cdots & \ast \\ 0 & \ddots & \vdots \\ 0 \cdots & 0 & \ast \end{pmatrix}$ Jacobian is upper triangular!

$$\det J = \prod_{i=1}^{d} \frac{\partial f_i}{\partial z_i}$$ now $\det J \sim \mathcal{O}(d)$ operations!

Still need f to be invertible ... again make some simple choices for f.

Wednesday, April 21, 2021 10:47 AM

e.g. —

could be NN's.

$$x_i = z_i\, \sigma_i(x_1, \ldots x_{i-1}) + \mu_i(x_1, \ldots x_{i-1})$$

or $z_1, \ldots z_{i-1}$ $\quad z_1, \ldots z_{i-1}$

invertible $\quad z_i = \dfrac{x_i - \mu_i(x_1, \ldots x_{i-1})}{\sigma_i(x_1, \ldots x_{i-1})}$

Another perspective on Autoregressive Flows:

$$P(x_i \mid x_1, \ldots x_{i-1}) = \underbrace{N(x_i \mid \mu_i, \sigma_i)}_{\text{"mixture density network"}}$$ simple parametric density

$$\boxed{P(x) = \prod_{i=1}^{d} P(x_i \mid x_1, \ldots x_{i-1})}$$

- Problem w/ Auto. Flows is could depend heavily on ordering of x's.
- Can chain together A.F.'s to build more complex transformations
- Can permute order of x's btw successive A.F.'s.

Wednesday, April 21, 2021 10:58 AM

Chaining together successive AF's w/ normal dist's

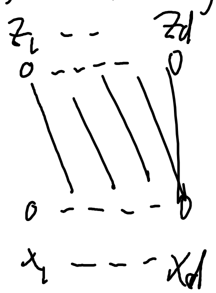
$\longrightarrow$ "Masked Autoregressive Flows"

or "Inverse Autoregressive Flows"

dep. μ_i, σ_i are fns of x or z.

μ_i and σ_i are NN's

$\mu_i(z_1, \cdots z_{i-1})$



take ordinary DNN
apply binary masks to all weights
to enforce autoregressive property.

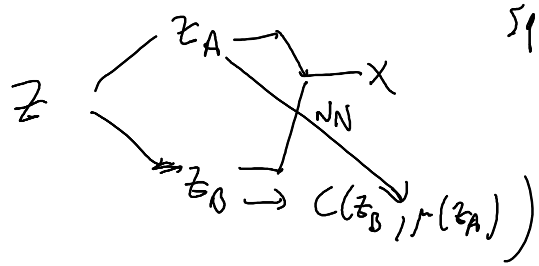
"coupling fns"

↳ C's: affine layer "RealNVP"

PW linear, quadratic, etc. "Neural Imp. Sampling"

rational quadratic splines "Neural Spline Flows"

• Coupling Layers

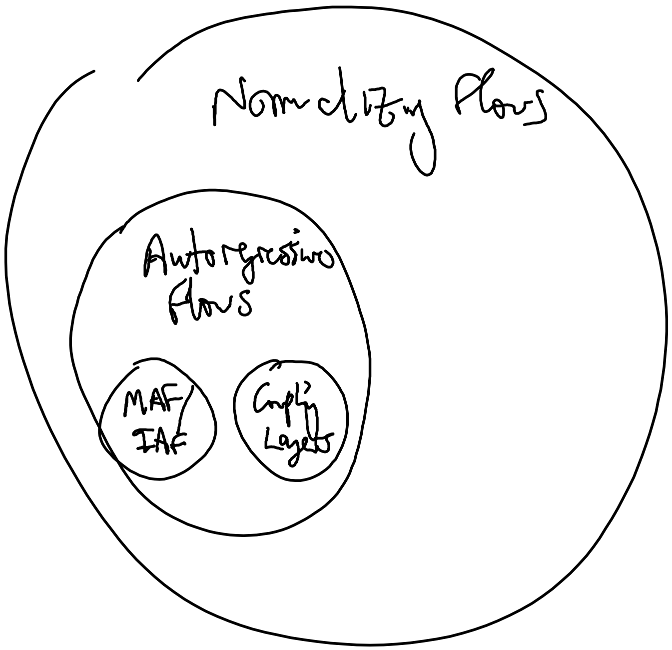


Autoregressive!

split data into 2 groups

$$\begin{cases} x_1 = z_1 \\ \vdots \\ x_k = z_k \\ x_{k+1} = C_{k+1}(z_{k+1}, \mu_{k+1}(z_1 \cdots z_k)) \\ \vdots \\ x_d = C_d(z_d, \mu_d(z_1, \cdots z_k)) \end{cases}$$

Wednesday, April 21, 2021 11:06 AM



Training a density estimator $p(x|\theta)$

MLE: $\prod_{i=1}^{N_{data}} p(x_i|\theta)$

neg. log likelihood: $L = -\sum_{i=1}^{N_{data}} \log p(x_i|\theta)$

goal: $\min_{\theta} L$