The Jacobian is computed with Neural autoregressive flow. The NF is decomposed in a way:

\[ p_Y(y) = p_X(x) \prod_{i=1}^{d} p_{k|X}(x_i | x_{<i}) \]

Usually, a density is decomposed in an estimation:

\[ p_Y(y) = \frac{1}{Z} \exp(f(Y)) \]

The NF is decomposed in:

\[ y_i = f_i^{	heta}(x_i; x_{<i}) = f_i^{	heta}(y, \theta | x_{<i}) \]

The NF is composed in:

\[ J_{f_i^{	heta}} = \frac{\partial y_i}{\partial x_i} = \frac{\partial f_i^{	heta}}{\partial x_i} = \frac{\partial f_i^{	heta}}{\partial y_i} \frac{\partial y_i}{\partial x_i} \]

\[ \text{Invertibility depends on the transformers} \]

\[ \text{Trivially invertible transformations may not be express enough} \]

\[ \text{Neural autoregressive flow (NAF) by Huang et al. (2018): replaces hand-crafted transformers with invertible neural networks!} \]

\[ \text{The Jacobian is computed with backpropagation:} \]

\[ J_{f_i^{	heta}} = \left| \frac{\partial y_i}{\partial x_i} \right| \left| \frac{\partial y_{k+i|y}}{\partial x_i} \right| \ldots \left| \frac{\partial y_i}{\partial x_i} \right| \]

**ADVANTAGES:**

- NAFs are universal approximators of density functions

**DRAWBACKS:**

- NAFs are hyper-networks and therefore the number of parameters scale quadratically.

**SOLUTION:**

our model a universal approximator of density functions with single feed-forward network!

- we model each \( f_i \) directly as an NN without a conditioner
- we employ affine transformations with positive weight to process \( x_i \) ensuring strict monotonicity and thus invertibility

For each affine layer, the weight matrix \( W \) is a lower-triangular block matrix with strictly positive diagonal blocks:

\[ W = \begin{bmatrix} \exp(B_{11}) & 0 & \ldots & 0 \\ B_{12} & \exp(B_{22}) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ B_{1d} & B_{2d} & \ldots & \exp(B_{dd}) \end{bmatrix} \]

- Universal approximator of densities: we can arbitrarily increase the hidden layer dimension
- Autoregressive: lower triangular Jacobian
- Stable: the det-Jacobian can be computed in the log-domain
- Efficient: fewer parameters than NAF and easy-to-compute Jacobian

**Results**

- Comparison with Glow (Kingma and Dhariwal, 2018) on density estimation
- Discontinuities and low-density regions are better modelled by B-NAF

- Comparison with Planar Flows (Rezende and Mohamed, 2015) on density matching
- 32 layers of planar flows work better than 2 layers of B-NAF

- More shallow!
- Faster training!

**Contact Information**

Nicola De Cao
Ph.D. Candidate at University of Amsterdam
nicola.decao@gmail.com
https://nicola-decao.github.io
https://twitter.com/nicola_decao

**References**


**Acknowledgements**

We would like to thank George Papamakarios and Luca Falorsi for insightful discussions. This project is supported by SAP Innovation Center Network, ERC Starting Grant Broadcast (677284) and the Dutch Organization for Scientific Research (NWO) VIDI 639.022.518. Wilker Aziz is supported by the European Unions Horizon 2020 research and innovation programme under grant agreement No 825299 (Gourmet).