NICE Formulas

January 21, 2021

1 Forward Pass

1. Split input x vector: $x = [x_{I1}; x_{I2}]$
2. Apply Additive Coupling layer 1:
$$\begin{cases} x_{I1} \to h_{I1}^{(1)} \\ x_{I2} + m^{(1)}(x_{I1}) \to h_{I2}^{(1)} \end{cases}$$
3. Apply Additive Coupling layer 2:
$$\begin{cases} h_{I1} + m^{(2)}(x_{I2}) \to h_{I1}^{(2)} \\ h_{I1}^{(1)} \to h_{I2}^{(2)} \end{cases}$$
4. Apply Additive Coupling layer 3:
$$\begin{cases} h_{I1}^{(2)} \to h_{I1}^{(3)} \\ h_{I2} + m^{(3)}(x_{I1}) \to h_{I2}^{(3)} \end{cases}$$
5. Apply Additive Coupling layer 4:
$$\begin{cases} h_{I1} + m^{(4)}(x_{I1}) \to h_{I1}^{(4)} \\ h_{I2}^{(3)} \to h_{I2}^{(4)} \end{cases}$$
6. Join $h_{I1}^{(4)}$ and $h_{I2}^{(4)}$: $\text{concat}(h_{I1}^{(4)}; h_{I2}^{(4)}) \to h^{(4)}$
7. Apply scale layer: $exp(S) \odot h^{(4)} \to h$
8. Apply prior such as logistic distribution where $p_H \backsim f(x; 0, 1)$ and get log-likelihood:
$$\log(p_H(h)) = \sum_{d=1}^{D} -\log(1 + \exp(h_d)) - \log(1 + \exp(-h_d))$$
9. Maximize log-likelihood: $\min \big[L = -\big(\log(p_H(h) + \sum_{i=1}^{D} S_{ii} \big) \big]$

2 Inverse Pass

1. Sample prior $p_H \backsim f(x; 0, 1)$: $p_H \rightarrow h$
2. Apply inverse scale layer: $h \odot \exp(-S) \rightarrow h^{(4)}$
3. Split $h^{(4)}$: $h^{(4)} \rightarrow [h_{I1}^{(4)}; h_{I2}^{(4)}]$
4. Apply inverse Additive Coupling layer 4:

$$\begin{cases} h_{I1}^{(4)} - m^{(4)}(h_{I2}^{(I4)}) \rightarrow h_{I1}^{(3)} \\ h_{I2}^{(4)} \rightarrow h_{I2}^{(3)} \end{cases}$$

5. Apply inverse Additive Coupling layer 3:

$$\begin{cases} h_{I1}^{(3)} \rightarrow h_{I1}^{(2)} \\ h_{I2}^{(3)} - m^{(3)}(h_{I1}^{(I3)}) \rightarrow h_{I2}^{(2)} \end{cases}$$

6. Apply inverse Additive Coupling layer 2:

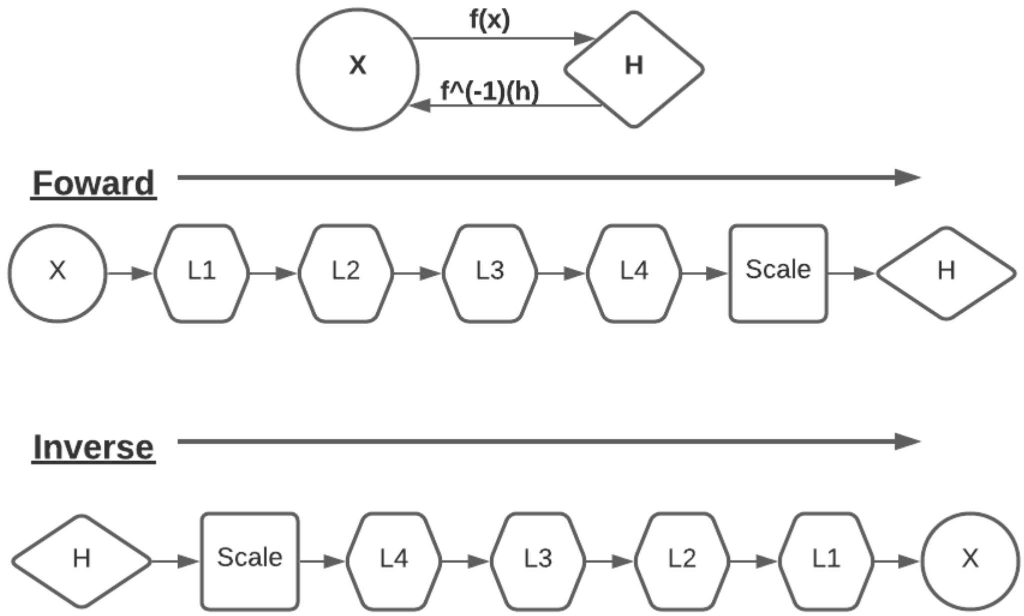
$$\begin{cases} h_{I1}^{(2)} - m^{(2)}(h_{I2}^{(I2)}) \rightarrow h_{I1}^{(1)} \\ h_{I2}^{(2)} \rightarrow h_{I2}^{(1)} \end{cases}$$

7. Apply inverse Additive Coupling layer 1:

$$\begin{cases} h_{I1}^{(1)} \rightarrow x_{I1} \\ h_{I2}^{(1)} - m^{(1)}(h_{I1}^{(I1)}) \rightarrow x_{I2} \end{cases}$$

8. Join x_{I1} and x_{I2}: $\text{concat}(x_{I1}; x_{I2}) \rightarrow x$

3 NICE Diagram



4 Method of Transformations

For monotonic transformations $h_1, \dots h_n$;

$$f_U(u_1, \dots, u_n) = f_X(h(x_1), \dots h(x_n))|J|$$

5 Method of Transformations 1D

Say $f : X \to U$ and f is monotonic increasing or decreasing then

$$f_U(u) = f_X(h^{-1}(u)) \times |\frac{dh^{-1}u}{du}|$$

Proof:

1. Denote $F_U(u) = P(U \leq u)$ and $U = h(x)$
2. $P(U \leq u) = P(h(X) \leq u) = P(X \leq h^{-1}(u)) = F_X(h^{-1}(u))$
3. $f_U(u) = \frac{d}{du}F_U(u) = \frac{d}{du}F_X(h^{-1}(u)) = f_X(h^{-1}(u)) \times |\frac{dh^{-1}u}{du}|$ $\square$

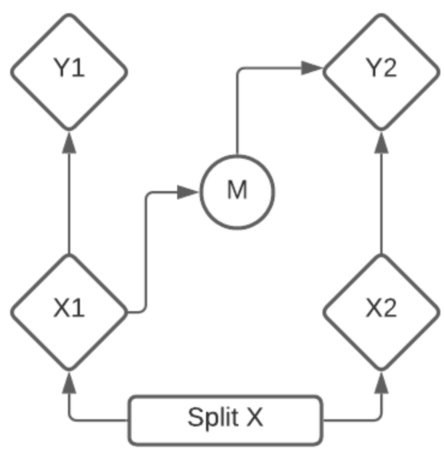
6 Additive Coupling

- Let x be a sample from your collected data
- Let m be a Neural Net with hidden ReLU layers and a linear output layer
- Split $x \to [x_1; x_2]$
- Forward:

$$\begin{cases} x_1 \to y_1 \\ x_2 + m(x_1) \to y_2 \end{cases}$$

- Inverse:

$$\begin{cases} y_1 \to x_1 \\ y_2 - m(y_1) \to x_2 \end{cases}$$



7 Prior

Set the final transformation to a simple distribution such as a standard logistic taken to the log:

$$\log f(x; 0, 1) = \log(p_H(h)) = -\log(1 + \exp(h_d)) - \log(1 + \exp(-h_d))$$

8 Log Likelihood using Additive Coupling

$$\log(p_X(x)) = \sum_{d=1}^{D} \log(p_{Hd}(f_d(x))) + \log(|\frac{\partial f(x)}{\partial x}|)$$

By using the Additive Coupling layer:

$$\frac{\partial y}{\partial x} = \begin{bmatrix} I_d & 0 \\ \frac{\partial y_{I_2}}{\partial x_{I_1}} & \frac{\partial y_{I_2}}{\partial x_{I_2}} \end{bmatrix} \text{ and } \frac{\partial y_{I_2}}{\partial x_{I_2}} = 1$$

This causes the $\sum_{d=1}^{D} \log(|\frac{\partial f(x)}{\partial x}|) = \sum_{d=1}^{D} \log(1) = 0$ since $\frac{\partial y}{\partial x}$ makes a triangular matrix whose determinate is the cumulative product of the diagonal elements, which are all ones.

Therefore, the log-likelihood becomes:

$$\log(p_X(x)) = \sum_{d=1}^{D} \log(p_{Hd}(f_d(x)))$$

9 Scaling Layer

To allow for shrinking or expanding, given the shape of the density of the data, a learnable diagonal matrix is multiplied by the output of the Additive Coupling layers. The diagonal matrix is multiplied by exp to force the outputs to be positive, given that we are outputting probabilities at the end.

$$\exp(S) \odot h^{(4)} \to h \quad \text{where} \quad S = \begin{bmatrix} s_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s_{DD} \end{bmatrix}$$

By adding the Scaling layer, the log-likelihood becomes:

$$\log(p_X(x)) = \sum_{d=1}^{D} \log(p_{Hd}(f_d(x))) + S_{dd}$$

10 Objective Function

$$\min \quad -\Big[\sum_{d=1}^{D} \log(p_{Hd}(f_d(x))) + S_{dd}\Big]$$

11 NICE citation

Dinh, Laurent, David Krueger, and Yoshua Bengio. "Nice: Non-linear independent components estimation." arXiv preprint arXiv:1410.8516 (2014).