Real NVP

Let $x = (x_1, x_2, x_3, x_4) \sim P_X$ and $z = (z_1, z_2, z_3, z_4) \sim \mathcal{N}(0, 1)$. The goal is to design an invertible network such that it can transform data distribution P_X to the normal distribution $\mathcal{N}(0, 1)$.

Give the boolean vector $b = (1, 1, 0, 0)$ and two mappings

$$\begin{aligned} s &: \mathbb{R}^4 \to \mathbb{R}^4 \\ t &: \mathbb{R}^4 \to \mathbb{R}^4, \end{aligned}$$

the coupling layer $z = c(x)$ is defined as the transformation

$$\begin{aligned} c &: \mathbb{R}^4 \to \mathbb{R}^4 \\ & x \to b \odot x + (1 - b) \odot [x \odot \exp(s(b \odot x)) + t(b \odot x)] \end{aligned}$$

where exp is component-wise, and $\odot$ is the Hadamard component-wise product. Particularly,

$$\begin{aligned} z_1 &= x_1 \\ z_2 &= x_2 \\ z_3 &= x_3 \cdot \exp(s_3(b \odot x)) + t_3(b \odot x) \\ z_4 &= x_4 \cdot \exp(s_4(b \odot x)) + t_4(b \odot x) \end{aligned}$$

where $s_i(b \odot x)$ and $t_i(b \odot x)$ is the i-th component of $s(b \odot x)$ and $t(b \odot x)$, respectively. By arranging terms,

$$\begin{aligned} x_1 &= z_1 \\ x_2 &= z_2 \\ x_3 &= [z_3 - t_3(b \odot x)] \cdot \exp(-s_3(b \odot x)) \\ x_4 &= [z_4 - t_4(b \odot x)] \cdot \exp(-s_4(b \odot x)) \end{aligned}$$

Note that $b \odot x = (1, 1, 0, 0) \odot (x_1, x_2, x_3, x_4) = (x_1, x_2, 0, 0) = (z_1, z_2, 0, 0) = b \odot z$, we have the inverse mapping $x = c^{-1}(z)$

$$\begin{aligned} x_1 &= z_1 \\ x_2 &= z_2 \\ x_3 &= [z_3 - t_3(b \odot z)] \cdot \exp(-s_3(b \odot z)) \\ x_4 &= [z_4 - t_4(b \odot z)] \cdot \exp(-s_4(b \odot z)) \end{aligned}$$

or

$$c^{-1}(z) = b \odot z + (1 - b) \odot [x - t(b \odot z)] \odot \exp(-s(b \odot z))$$

The Jacobian matrix of the invertible mapping $c(z)$ is determined by[1]

$$\begin{aligned} \frac{\partial z_1}{\partial x_1} &= 1, \frac{\partial z_1}{\partial x_2} = \frac{\partial z_1}{\partial x_3} = \frac{\partial z_1}{\partial x_4} = 0 \\ \frac{\partial z_2}{\partial x_2} &= 1, \frac{\partial z_2}{\partial x_1} = \frac{\partial z_2}{\partial x_3} = \frac{\partial z_2}{\partial x_4} = 0 \\ \frac{\partial z_3}{\partial x_3} &= \exp(s_3(b \odot x)), \frac{\partial z_3}{\partial x_4} = 0 \\ \frac{\partial z_4}{\partial x_4} &= \exp(s_4(b \odot x)), \frac{\partial z_4}{\partial x_3} = 0 \end{aligned}$$

1. We have used the fact that $b \odot x = (x_1, x_2, 0, 0)$ is constant in term of x_3 and x_4.

That is,

$$\frac{\partial z}{\partial x} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ ? & ? & \exp(s_3(b \odot x)) & 0 \\ ? & ? & 0 & \exp(s_4(b \odot x)) \end{pmatrix}$$

and the log of absolute value of its determinant is

$$\begin{aligned} \log\left[\left|\det\left(\frac{\partial z}{\partial x}\right)\right|\right] &= \log|1 \cdot 1 \cdot \exp(s_3(b \odot x)) \cdot \exp(s_4(b \odot x))| \\ &= 0 + 0 + s_3(b \odot x) + s_4(b \odot x) \\ &= \sum_{i=1}^{4} [(1-b) \odot s(b \odot x)]_i \end{aligned}$$

where $[(1-b) \odot s(b \odot x)]_i$ denotes the i-th component of $(1-b) \odot s(b \odot x)$.

We also have

$$\begin{aligned} 1 &= \int p_Z(z)dz \\ &= \int p_Z(c(x))\left|\det\left(\frac{\partial z}{\partial x}\right)\right| dx \\ &= \int p_X(x)dx \\ &\Rightarrow p_X(x) = p_Z(c(x))\left|\det\left(\frac{\partial z}{\partial x}\right)\right| \end{aligned}$$

and thus,

$$\begin{aligned} \log p_X(x) &= \log p_Z(c(x)) + \log\left|\det\left(\frac{\partial z}{\partial x}\right)\right| \\ &= \log p_Z(c(x)) + \sum_{i=1}^{4} [(1-b) \odot s(b \odot x)]_i \end{aligned}$$