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

$$s : \mathbb{R}^4 \to \mathbb{R}^4$$
$$t : \mathbb{R}^4 \to \mathbb{R}^4,$$

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

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

where exp is component-wise, and ⊙ is the Hadamard component-wise product. Particularly,

$$z_1 = x_1$$

$$z_2 = x_2$$

$$z_3 = x_3 \cdot \exp(s_3(b \odot x)) + t_3(b \odot x)$$

$$z_4 = x_4 \cdot \exp(s_4(b \odot x)) + t_4(b \odot x)$$

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,

$$x_{1} = z_{1}$$

$$x_{2} = z_{2}$$

$$x_{3} = [z_{3} - t_{3}(b \odot x)] \cdot \exp(-s_{3}(b \odot x))$$

$$x_{4} = [z_{4} - t_{4}(b \odot x)] \cdot \exp(-s_{4}(b \odot x))$$

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 $x = c^{-1}(z)$

$$x_{1} = z_{1}$$

$$x_{2} = z_{2}$$

$$x_{3} = [z_{3} - t_{3}(b \odot z)] \cdot \exp(-s_{3}(b \odot z))$$

$$x_{4} = [z_{4} - t_{4}(b \odot z)] \cdot \exp(-s_{4}(b \odot z))$$

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

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

^{1.} We have useed 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

$$\log\left[\left|\det\left(\frac{\partial z}{\partial x}\right)\right|\right] = \log|1 \cdot 1 \cdot \exp(s_3(b \odot x)) \cdot \exp(s_4(b \odot x))|$$

$$= 0 + 0 + s_3(b \odot x) + s_4(b \odot x)$$

$$= \sum_{i=1}^{4} [(1-b) \odot s(b \odot x)]_i$$

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

$$1 = \int p_{Z}(z)dz$$

$$= \int p_{Z}(c(x)) \left| \det \left(\frac{\partial z}{\partial x} \right) \right| dx$$

$$= \int p_{X}(x)dx$$

$$\Rightarrow p_{X}(x) = p_{Z}(c(x)) \left| \det \left(\frac{\partial z}{\partial x} \right) \right|$$

and thus,

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