Derivation of the Loss in the CDVAE

Antonio Almudévar

$$\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{c}) = \iint q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x}) \log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{c}) \; d\boldsymbol{z_l} \; d\boldsymbol{z_u} \tag{1}$$

$$= \iint q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x}) \log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z_l}, \boldsymbol{z_u}, \boldsymbol{c}) p_{\boldsymbol{\theta}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{c})}{p_{\boldsymbol{\theta}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{c}, \boldsymbol{x})} \; d\boldsymbol{z_l} \; d\boldsymbol{z_u} \tag{2}$$

$$= \iint q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x}) \log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z_l}, \boldsymbol{z_u}, \boldsymbol{c}) p_{\boldsymbol{\theta}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{c}) q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x})}{p_{\boldsymbol{\theta}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{c}, \boldsymbol{x}) q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x})} \; d\boldsymbol{z_l} \; d\boldsymbol{z_u} \tag{3}$$

$$\begin{aligned} &= \iint q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x}) \log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z_l}, \boldsymbol{z_u}, \boldsymbol{c}) \; d\boldsymbol{z_l} \; d\boldsymbol{z_u} \\ &\quad + \iint q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x}) \log \frac{q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x})}{p_{\boldsymbol{\theta}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{c}, \boldsymbol{x})} \; d\boldsymbol{z_l} \; d\boldsymbol{z_u} \\ &\quad - \iint q_{\boldsymbol{\phi}}(\boldsymbol{z_l}|\boldsymbol{x}) q_{\boldsymbol{\phi}}(\boldsymbol{z_u}|\boldsymbol{x}) \log \frac{q_{\boldsymbol{\phi}}(\boldsymbol{z_l}|\boldsymbol{x}) q_{\boldsymbol{\phi}}(\boldsymbol{z_u}|\boldsymbol{x})}{p_{\boldsymbol{\theta}}(\boldsymbol{z_l}|\boldsymbol{c}) p_{\boldsymbol{\theta}}(\boldsymbol{z_u})} \; d\boldsymbol{z_l} \; d\boldsymbol{z_u} \end{aligned} \tag{4}$$

$$\begin{aligned} &= \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x})} \left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z_l}, \boldsymbol{z_u}, \boldsymbol{c}) \right] + D_{KL} \left(q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x}) || p_{\boldsymbol{\theta}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{c}, \boldsymbol{x}) \right) \\ &\quad - \int q_{\boldsymbol{\phi}}(\boldsymbol{z_u}|\boldsymbol{x}) \int q_{\boldsymbol{\phi}}(\boldsymbol{z_l}|\boldsymbol{x}) \log \frac{q_{\boldsymbol{\phi}}(\boldsymbol{z_l}|\boldsymbol{x})}{p_{\boldsymbol{\theta}}(\boldsymbol{z_l}|\boldsymbol{c})} \; d\boldsymbol{z_l} \; d\boldsymbol{z_u} \\ &\quad - \int q_{\boldsymbol{\phi}}(\boldsymbol{z_l}|\boldsymbol{x}) \int q_{\boldsymbol{\phi}}(\boldsymbol{z_u}|\boldsymbol{x}) \log \frac{q_{\boldsymbol{\phi}}(\boldsymbol{z_u}|\boldsymbol{x})}{p_{\boldsymbol{\theta}}(\boldsymbol{z_u})} \; d\boldsymbol{z_l} \; d\boldsymbol{z_u} \end{aligned} \tag{5}$$

$$\begin{aligned} &= \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x})} \left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z_l}, \boldsymbol{z_u}, \boldsymbol{c}) \right] + D_{KL} \left(q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x}) || p_{\boldsymbol{\theta}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{c}, \boldsymbol{x}) \right) \\ &\quad - D_{KL} \left(q_{\boldsymbol{\phi}}(\boldsymbol{z_l}|\boldsymbol{x}) || p_{\boldsymbol{\theta}}(\boldsymbol{z_l}|\boldsymbol{c}) \right) - D_{KL} \left(q_{\boldsymbol{\phi}}(\boldsymbol{z_u}|\boldsymbol{x}) || p_{\boldsymbol{\theta}}(\boldsymbol{z_u}) \right) \end{aligned} \tag{6}$$

Since the term $D_{KL} \left(q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x}) || p_{\boldsymbol{\theta}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{c}, \boldsymbol{x}) \right)$ is positive, we have that an ELBO of $\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{c})$ is:

$$\begin{aligned} &\mathcal{L}_{\boldsymbol{\theta}, \boldsymbol{\phi}} \left(\boldsymbol{x}, \boldsymbol{c}, \boldsymbol{z_l}, \boldsymbol{z_u} \right) = \\ &\quad \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x})} \left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z_l}, \boldsymbol{z_u}, \boldsymbol{c}) \right] - D_{KL} \left(q_{\boldsymbol{\phi}}(\boldsymbol{z_l}|\boldsymbol{x}) || p_{\boldsymbol{\theta}}(\boldsymbol{z_l}|\boldsymbol{c}) \right) - D_{KL} \left(q_{\boldsymbol{\phi}}(\boldsymbol{z_u}|\boldsymbol{x}) || p_{\boldsymbol{\theta}}(\boldsymbol{z_u}) \right) \end{aligned} \tag{7}$$

Moreover, since we have imposed that the decoder does not depend on the class, we have that:

$$\mathcal{L}_{CDVAE} = \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z_l}, \boldsymbol{z_u}|\boldsymbol{x})} \left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z_l}, \boldsymbol{z_u}) \right] - D_{KL} \left(q_{\boldsymbol{\phi}}(\boldsymbol{z_l}|\boldsymbol{x}) || p_{\boldsymbol{\theta}}(\boldsymbol{z_l}|\boldsymbol{c}) \right) - D_{KL} \left(q_{\boldsymbol{\phi}}(\boldsymbol{z_u}|\boldsymbol{x}) || p_{\boldsymbol{\theta}}(\boldsymbol{z_u}) \right) \tag{8}$$