

The Variational Autoencoder exploits the methods described in the previous section to define a probabilistic model whose elbo is highly reminiscent of the objective optimised in a traditional autoencoder. In particular we define a generative model where we assume that the i^{th} observation was generated by first sampling a latent variable $z_i \sim N(0, I)$ and that the each observation vector real-valued observation $x_i \sim N(\mu_\theta(z_i), \sigma_\theta^2(z_i))$ where

$$\mu_\theta(z_i) = h_i W_\mu^{(q)} + b_\mu^{(q)}, \quad (1)$$

$$\log \sigma_\theta^2(z_i) = h_i W_\sigma^{(q)} + b_\sigma^{(q)}, \quad (2)$$

$h_i \in \mathbb{R}^{1 \times D_h}$ is the output at the final hidden layer of the “decoder MLP”, $W_\mu^{(q)}, W_\sigma^{(q)} \in \mathbb{R}^{D_h \times D_z}$ are matrices mapping from the D_h hidden units to the D_z dimensional latent space. Similarly, $b_\mu^{(q)}, b_\sigma^{(q)} \in \mathbb{R}^{1 \times D_z}$ are row vector biases. Note that the variances are parameterised implicitly through their logs to ensure that they are correctly valued only on the positive reals.

If the output vectors are to be treated as either probabilities or are binary values then $x_i \sim \text{Bernoulli}(f_\theta(z_i))$ where again given h_i , the output at the final hidden layer of the decoder MLP for latent-space value z_i

$$f_\theta(z_i) = \left[1 + \exp\left(-h_i W^{(q)} - b^{(q)}\right) \right]^{-1}, \quad (3)$$

where $W^{(q)} \in \mathbb{R}^{D_h \times D_z}$ and $b^{(q)} \in \mathbb{R}^{1 \times D_z}$

The recognition model is given by

$$q_\phi(z|x) = N(z | \mu_\phi(x_i), \sigma_\phi^2(x_i)), \quad (4)$$

where the distributional parameters $\mu_\phi(x_i)$ and $\sigma_\phi^2(x_i)$ are again given by an MLP whose input is x_i . Note that whenever a variance is parameterised by an MLP,