

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 $\mathbf{z}_i \sim N(0, I)$ and that the each observation vector real-valued observation $\mathbf{x}_i \sim N(\mu_\theta(\mathbf{z}_i), \sigma_\theta^2(\mathbf{z}_i))$ where

$$\mu_\theta(\mathbf{z}_i) = \mathbf{h}_i W_\mu^{(dec)} + \mathbf{b}_\mu^{(dec)}, \quad (1)$$

$$\log \sigma_\theta^2(\mathbf{z}_i) = \mathbf{h}_i W_\sigma^{(dec)} + \mathbf{b}_\sigma^{(dec)}, \quad (2)$$

$\mathbf{h}_i \in \mathbb{R}^{1 \times D_h}$ is the output at the final hidden layer of the “decoder MLP”, $W_\mu^{(dec)}, W_\sigma^{(dec)} \in \mathbb{R}^{D_h \times D_z}$ are matrices mapping from the D_h hidden units to the D_z dimensional latent space. Similarly, $\mathbf{b}_\mu^{(dec)}, \mathbf{b}_\sigma^{(dec)} \in \mathbb{R}^{1 \times D_z}$ 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 binary valued then $x_i \sim \text{Bernoulli}(f_\theta(\mathbf{z}_i))$ where again given \mathbf{h}_i , the output at the final hidden layer of the decoder MLP for latent-space value \mathbf{z}_i

$$f_\theta(\mathbf{z}_i) = \left[1 + \exp \left(-\mathbf{h}_i W^{(dec)} - \mathbf{b}^{(dec)} \right) \right]^{-1}, \quad (3)$$

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

The recognition model is given by

$$q_\phi(\mathbf{z} | \mathbf{x}) = N(\mathbf{z} | \mu_\phi(\mathbf{x}_i), \sigma_\phi^2(\mathbf{x}_i)), \quad (4)$$

where the distributional parameters $\mu_\phi(\mathbf{x}_i)$ and $\sigma_\phi^2(\mathbf{x}_i)$ are again given by an MLP, which will be referred to as the “encoding MLP”, whose input is \mathbf{x}_i .

In this case, there is a tractable closed form solution for the KL-divergence between the variational posterior $q_\phi(\mathbf{z} | \mathbf{x})$ and the prior $p_\theta(\mathbf{z})$ meaning that we can approximate the ELBO using equation ?? . Additionally we have that

$$g_\phi(\mathbf{x}, \epsilon) = \mu_\phi(\mathbf{x}_i) + \sigma_\phi(\mathbf{x}_i) \odot \epsilon, \quad \epsilon \sim N(0, I). \quad (5)$$

The algorithm to perform inference in the Variational Autoencoder is provided in algorithm 1.

Algorithm 1 Procedure by which inference is performed in the Variational Autoencoder.

$\theta, \phi \leftarrow$ Initialisation.

while not converged in θ, ϕ **do**

 Pick subset of size $\mathbf{x}_{1:M}$ from the full dataset uniformly at random.

 Compute $\mu_{1:M}, \log \sigma_{1:M}^2$ using the encoding MLP.

 For all $i \in \{1, \dots, M\}$, sample $\epsilon_i \sim N(0, I)$.

 For all $i \in \{1, \dots, M\}$, $\mathbf{z}_i \leftarrow \mu_i + \sigma_i \odot \epsilon_i$.

 For all $i \in \{1, \dots, M\}$, compute $\log p(\mathbf{x}_i | \mathbf{z}_i)$ using the decoder MLP and likelihood function.

 Compute $g \leftarrow \nabla_{\theta, \phi} \mathcal{L}^B(\theta, \phi; \mathbf{x}_{1:M}, \epsilon_{1:M})$

 Update θ, ϕ using noisy gradient estimate g .

end while
