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 $z_i \sim \mathrm{N}\left(0, I\right)$ and that the each observation $x_i \sim \mathrm{N}\left(\mu_\theta\left(z_i\right), \sigma^2_\theta\left(z_i\right)\right)$ if the observations are real-valued, or $x_i \sim \mathrm{Bernoulli}\left(f_\theta\left(z_i\right)\right)$ if they are binary or valued on $[0,1]$. In both cases the distribution parameters $\mu_\theta\left(z_i\right), \sigma^2_\theta\left(z_i\right), f_\theta\left(z_i\right)$ are parameterised in terms of a multi-layer perceptron (MLP) whose input is z_i, which will be referred to as the "decoder MLP". Specifically, define h_i to be the vector output at the final hidden layer of the decoder MLP when provided input z_i, then

$$\mu_\theta\left(z_i\right) = h_i W^{(q)}_\mu + b^{(q)}_\mu, \tag{1}$$

$$\log \sigma^2_\theta\left(z_i\right) = h_i W^{(q)}_\sigma + b^{(q)}_\sigma. \tag{2}$$

The recognition model is given by

$$q_\phi\left(z \,|\, x\right) = \mathrm{N}\left(z \,\middle|\, \mu_\phi\left(x_i\right), \sigma^2_\phi\left(x_i\right)\right), \tag{3}$$

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