Workshop 7

COMP90051 Machine Learning

Semester 2, 2020

Learning Outcomes

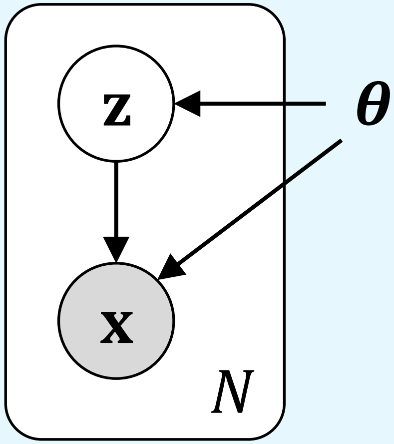
By the end of this workshop you should be able to:

1. explain how VAEs can be used to (approximately) fit latent variable models
2. compute the objective function for VAEs—the evidence lower bound
3. implement a VAE in Keras/TensorFlow

Variational autoencoders (VAEs)

Problem: fit a latent variable model $p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) = p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})p_{\boldsymbol{\theta}}(\mathbf{z})$ where

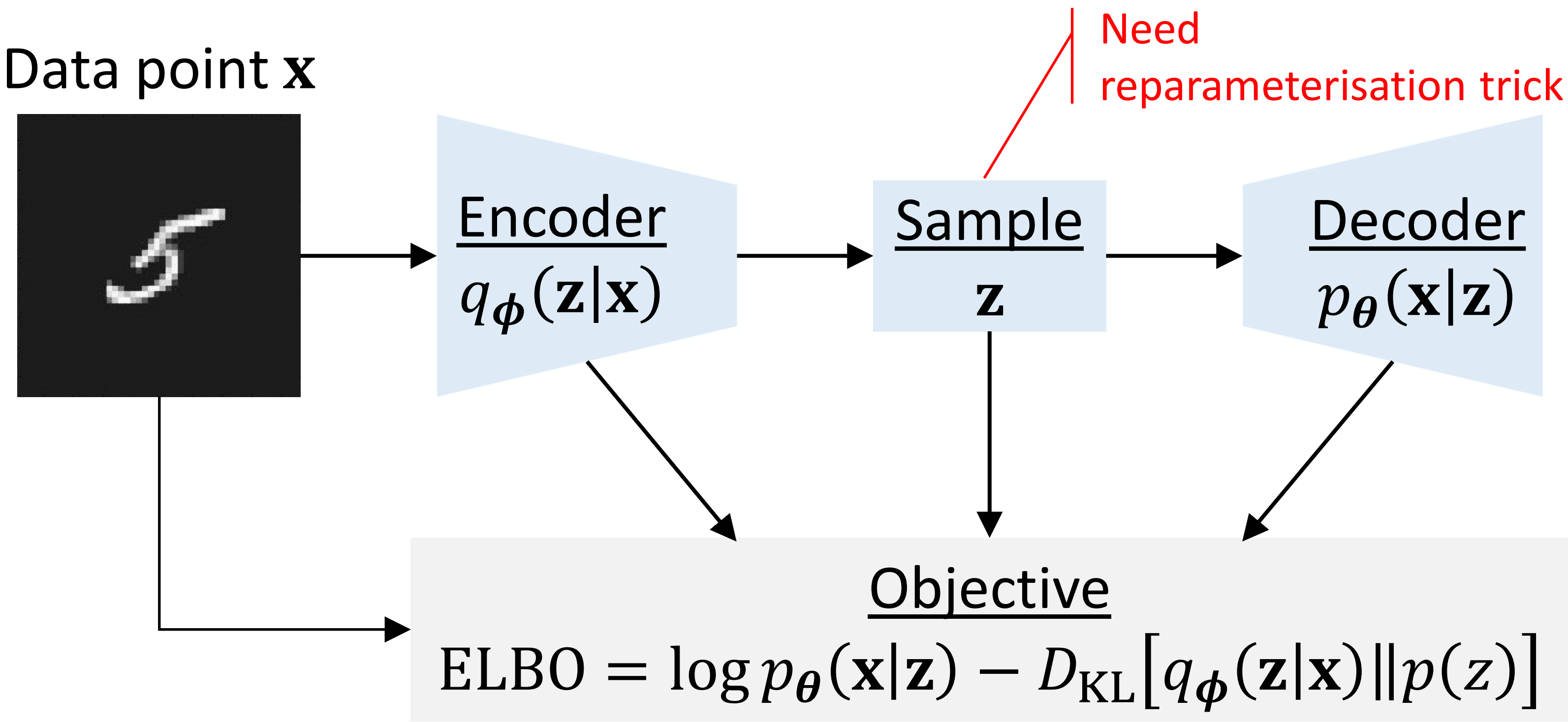
- $\mathbf{x}$ is the observed data,
- $\mathbf{z}$ is an (unobserved) latent variable,
- $\boldsymbol{\theta}$ is a set of unknown parameters



- VAEs can handle complex models where the likelihood $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$ is parameterised by a deep neural net
- Other methods break down in this case, because the posterior $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$ and marginal $p_{\boldsymbol{\theta}}(\mathbf{x})$ become intractable

Variational autoencoders (VAEs)

- Since $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$ is intractable, approximate it by $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$
- Can do inference using stochastic gradient descent



VAE objective function

- Maximise the evidence lower bound (ELBO) with respect to the model parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$:

$$\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\phi}}(\mathbf{x}) = \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}[\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})] - D_{\mathrm{KL}}\left[q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) \| p(z)\right]$$

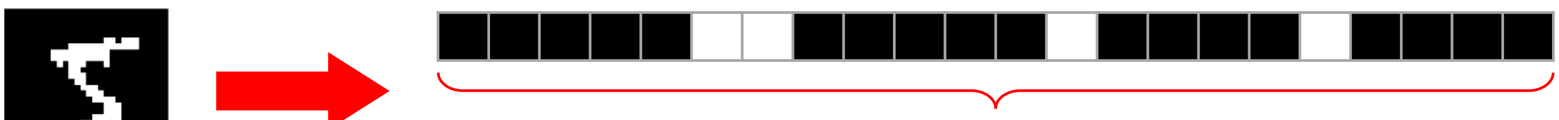
- The ELBO simultaneously
  - approximately maximises the marginal likelihood $p_{\boldsymbol{\theta}}(\mathbf{x})$
  - minimises the KL divergence from the true posterior $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$ to the approximation $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$

- We'll minimise the negative ELBO in our implementation:

$$-\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\phi}}(\mathbf{x}) = \underbrace{-\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}[\log p_{\theta}(\mathbf{x}|\mathbf{z})]}_{\text{reconstruction loss}} + \underbrace{D_{\mathrm{KL}}\left[q_{\phi}(\mathbf{z}|\mathbf{x}) \| p(z)\right]}_{\text{regularisation term}}$$

A generative model for MNIST

- Express an image as a flattened array $\mathbf{x} \in \{0,1\}^d$
  - Note: we're assuming a 1-bit (black/white) image



$d = 28 \times 28$ pixels

- Assume an independent Bernoulli likelihood

$$p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z}) = \prod_{i=1}^{d} \pi_i^{x_i}(1-\pi_i)^{1-x_i} \text{ where}$$

$$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_d) = \mathrm{DecoderNN}_{\boldsymbol{\theta}}(\mathbf{z})$$

- Assume latent variable $\mathbf{z} \in \mathbb{R}^m$ has a spherical Gaussian prior

$$p_\theta(\mathbf{z}) = \mathcal{N}(\mathbf{z}; 0, \mathbf{I})$$

Posterior approximation

- The VAE framework requires us to specify a tractable approximation $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ for the posterior $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$
- We assume $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$ can be well-approximated by a factorised Gaussian:

$$q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}, \mathrm{diag}(\boldsymbol{\sigma})) \text{ where}$$
$$(\boldsymbol{\mu}, \log \boldsymbol{\sigma}) = \mathrm{EncoderNN}_{\boldsymbol{\phi}}(\mathbf{x})$$

- VAE will tune $\boldsymbol{\phi}$ to make $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x}) \approx q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$

Worksheet 7