Importance Weighted Autoencoders

Federico Barbero, Kaiqu Liang, Haoran Peng

University of Cambridge

Background: Variational Autoencoders (VAEs)

📖 Generative model capable of learning latent representations z from data x

To train a VAE, we maximise the evidence lower bound $\mathcal{L}^{\text{VAE}}$:

$$\mathcal{L}^{\text{VAE}}(\mathbf{x}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log\left(\frac{p_\theta(\mathbf{x},\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})}\right)\right]$$

$$\leq \log \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\left(\frac{p_\theta(\mathbf{x},\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})}\right)\right] = \log p_\theta(\mathbf{x})$$

where $\mathbf{X}$ is the observed data, $\mathbf{Z}$ is the latent variable, and $q(\mathbf{z}|\mathbf{x})$ is the variational posterior that approximates the true posterior $p(\mathbf{z}|\mathbf{x})$.

Importace Weighted Autoencoders (IWAEs)

💡 Optimise a tighter lower bound than the ELBO

To train a IWAE, we maximise a different variational lower bound $\mathcal{L}_k^{\text{IWAE}}$:

$$\mathcal{L}_k^{\text{IWAE}} = \mathbb{E}_{\mathbf{z}^1,\ldots,\mathbf{z}^k\sim q(\mathbf{z}|\mathbf{x})}\left[\log \frac{1}{k}\sum_{i=1}^{k}\frac{p(\mathbf{x},\mathbf{z}^i)}{q(\mathbf{z}^i|\mathbf{x})}\right]$$

$$\leq \log \mathbb{E}_{\mathbf{z}^1,\ldots,\mathbf{z}^k\sim q(\mathbf{z}|\mathbf{x})}\left[\frac{1}{k}\sum_{i=1}^{k}\frac{p(\mathbf{x},\mathbf{z}^i)}{q(\mathbf{z}^i|\mathbf{x})}\right] = \log p(\mathbf{x})$$

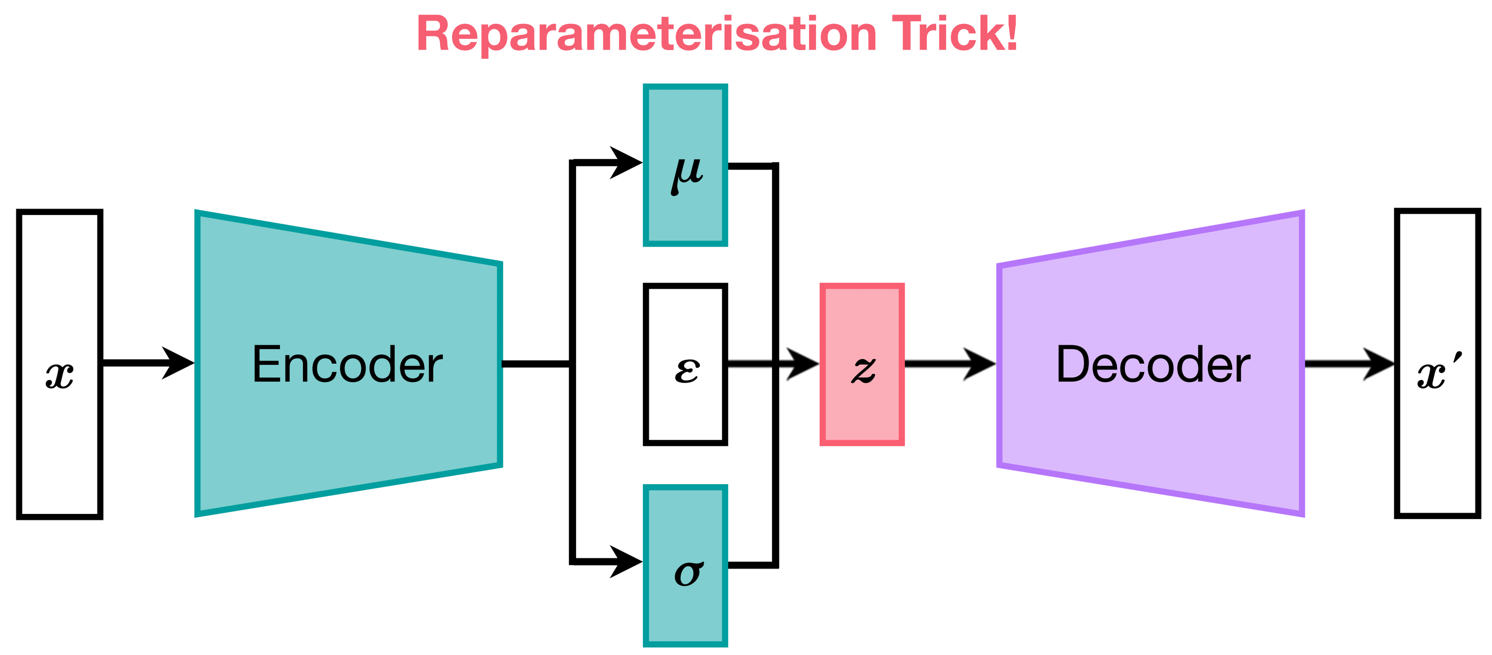
We can see that $\mathcal{L}^{\text{VAE}} = \mathcal{L}_k^{\text{IWAE}}$ when $k = 1$.
Burda et al. (2015) showed that for all k:

$$\mathcal{L}_k^{\text{IWAE}} \leq \mathcal{L}_{k+1}^{\text{IWAE}} \leq \log p(\mathbf{x})$$

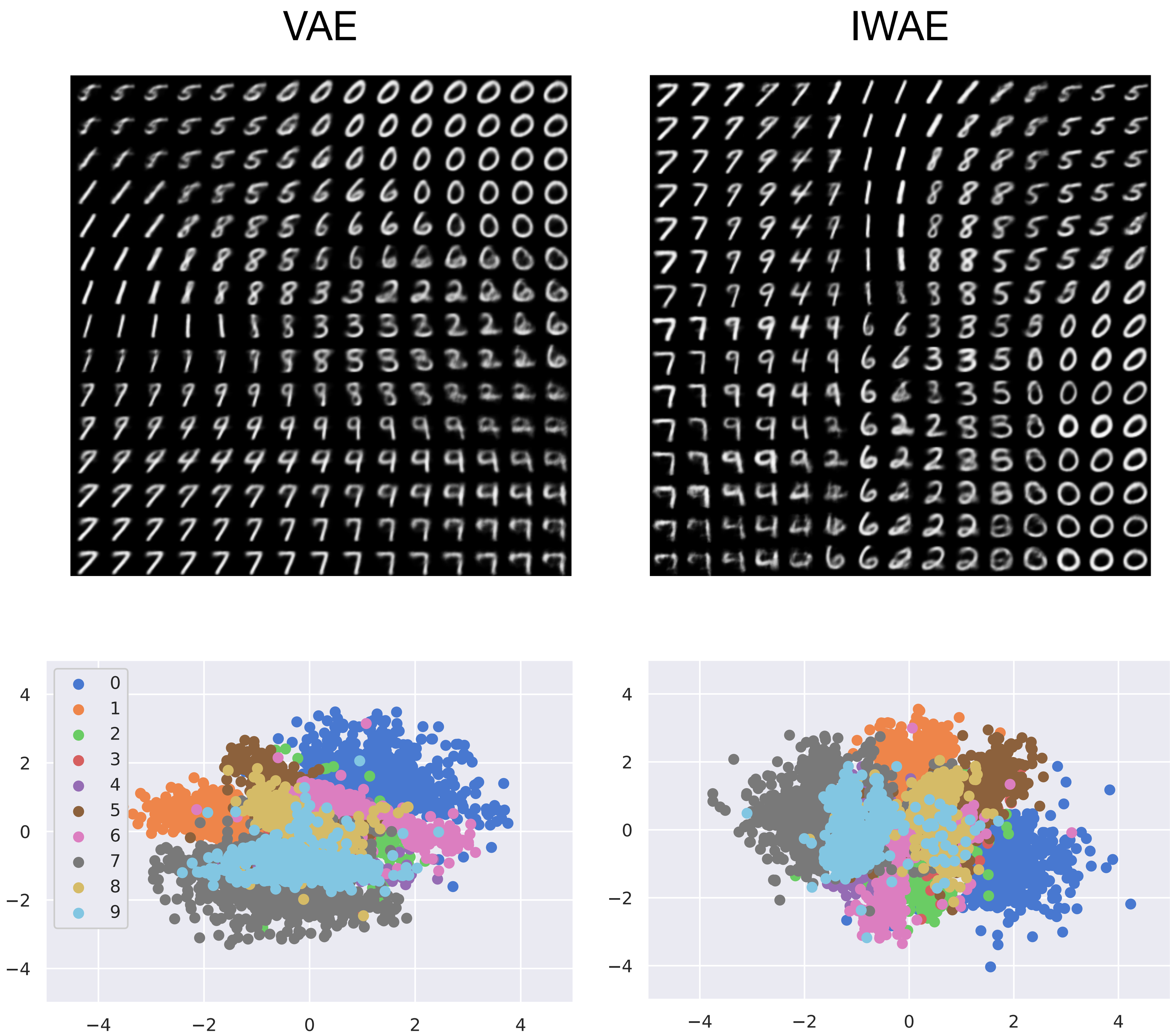
And the variational lower bound approaches the true log marginal likelihood as $k \to \infty$. IWAE is able to use more samples to improve the tightness of the bound.

Architecture

🔨 We use the same architectures as in Burda et al. (2015)

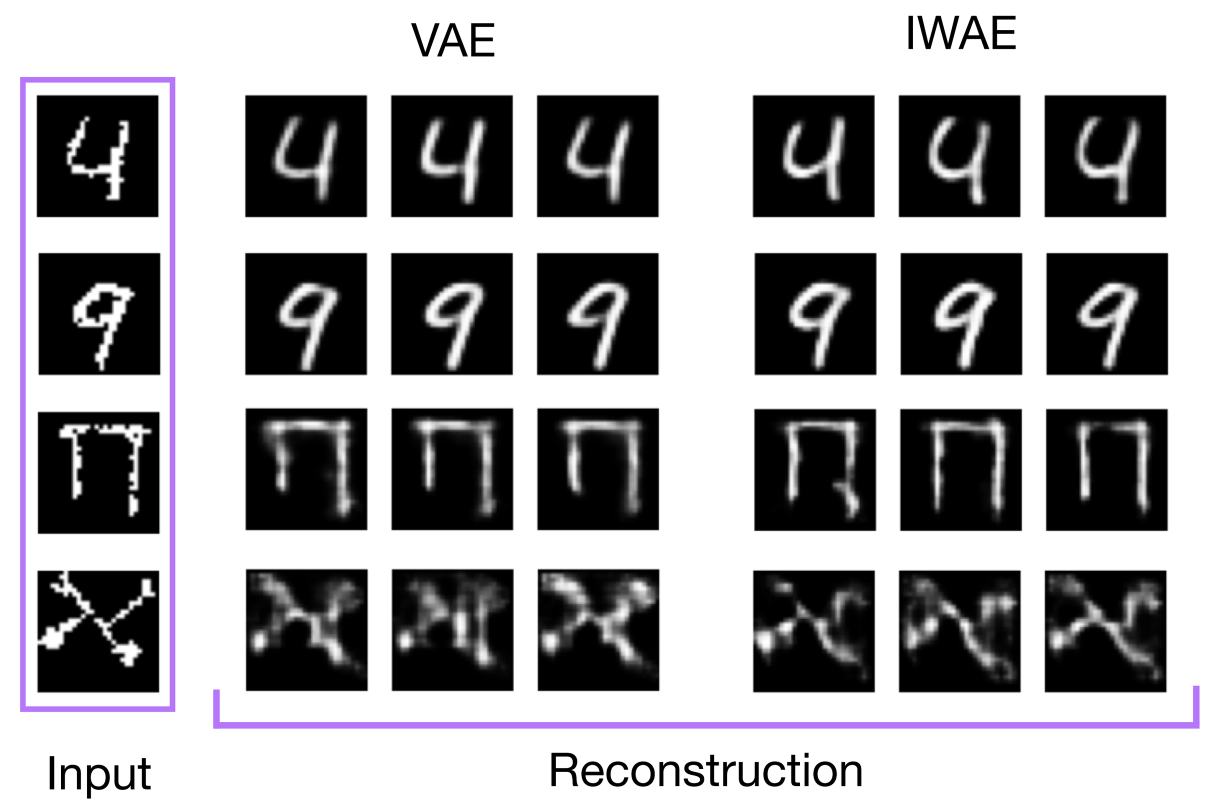


Latent Representations



🔎 Visualisation of a 2 dimensional latent space

Reconstruction



Sampling

🧪 We can generate new samples from the latent space by sampling from a Gaussian prior

Density Estimation

		MNIST		Omniglot	
# stoch. layers	k	VAE	IWAE	VAE	IWAE
1	1	87.68	87.68	112.36	112.36
	5	87.51	85.98	111.67	108.33
	50	87.54	**85.37**	112.05	**106.84**
2	1	87.18	87.18	111.31	111.31
	5	86.87	85.03	111.23	107.77
	50	86.93	**84.42**	111.15	**105.59**

🚀 The negative log likelihood is lower for IWAEs

References

1.Kingma, D. P., and Welling M., "Auto-encoding variational bayes." arXiv preprint arXiv:1312.6114 (2013).

2.Yuri B., Roger G. and Ruslan S., "Importance Weighted Autoencoders" arXiv:1509.00519 (2015)