The Reparamaterisation Trick

Greg Feldmann

August 23, 2020

1 Quick Overview of Variational Autoencoders

Let $\mathbf{z}$ be a vector of latent variables, $\mathbf{x}$ be a row vector from a dataset X, $q_\phi(\mathbf{z}|\mathbf{x})$ be the encoder and $p_\theta(\mathbf{x}|\mathbf{z})$ be the decoder. The loss function used is referred to as the evidence lower bound (ELBO). ELBO is defined as follows

$$\mathcal{L}_{\theta,\phi}(\mathbf{x}) = log(p_\theta(\mathbf{x}|\mathbf{z})) - D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z}|\mathbf{x})) \tag{1}$$

where D_{KL} is the Kullback-Leibler divergence.

See here for more on ELBO.

2 Optimising ELBO with Stochastic Gradient Descent

To optimise our VAE with stochastic gradient descent (SGD), we need to be able to take gradients of the expected value of ELBO with respect to the network weights θ and ϕ. This is easy enough for θ:

$$\nabla_\theta \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\mathcal{L}_{\theta,\phi}(\mathbf{x})] = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\nabla_\theta \mathcal{L}_{\theta,\phi}(\mathbf{x})] \tag{2}$$

Once the grad operator is inside the expectation, all we have to do is approximate the expectation with a sample.

Things are trickier with ϕ. As the expectation assumes that $p(\mathbf{z}|\mathbf{x}) = q_\phi(\mathbf{z}|\mathbf{x})$, we cannot simply move ∇_ϕ in and out of the expectation arbitrarily.

$$\nabla_\phi \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\mathcal{L}_{\theta,\phi}(\mathbf{x})] \neq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\nabla_\phi \mathcal{L}_{\theta,\phi}(\mathbf{x})] \tag{3}$$

That the left and right hand sides are not equal can be seen as follows:

$$\nabla_\phi \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\mathcal{L}_{\theta,\phi}(\mathbf{x})] = \nabla_\phi \int q_\phi(\mathbf{z}|\mathbf{x})\mathcal{L}_{\theta,\phi}(\mathbf{x})d\mathbf{x} \tag{4}$$

$$= \int \nabla_\phi (q_\phi(\mathbf{z}|\mathbf{x})\mathcal{L}_{\theta,\phi}(\mathbf{x}))d\mathbf{x} \tag{5}$$

$$= \int [q_\phi(\mathbf{z}|\mathbf{x})(\nabla_\phi \mathcal{L}_{\theta,\phi}(\mathbf{x})) + \mathcal{L}_{\theta,\phi}(\mathbf{x})(\nabla_\phi q_\phi(\mathbf{z}|\mathbf{x}))]d\mathbf{x} \tag{6}$$

$$= \int q_\phi(\mathbf{z}|\mathbf{x})(\nabla_\phi \mathcal{L}_{\theta,\phi}(\mathbf{x}))d\mathbf{x} + \int \mathcal{L}_{\theta,\phi}(\mathbf{x})(\nabla_\phi q_\phi(\mathbf{z}|\mathbf{x}))d\mathbf{x} \tag{7}$$

$$= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\nabla_\phi \mathcal{L}_{\theta,\phi}(\mathbf{x})] + \int \mathcal{L}_{\theta,\phi}(\mathbf{x})(\nabla_\phi q_\phi(\mathbf{z}|\mathbf{x}))d\mathbf{x} \tag{8}$$

This isn't in a form that we can easily handle. In particular, $\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})$, meaning $\mathbf{z}$ is stochastic rather than deterministic. This is where the reparamaterisation trick comes into play. We can make $\mathbf{z}$ deterministic by making it a function of a stochastic variable, ϵ:

$$\mathbf{z} = g(\epsilon, \phi, \mathbf{x}) \tag{9}$$

After replacing $\mathbf{z}$ with $g(\epsilon, \phi, \mathbf{x})$, the expectation is then taken over the distribution of ϵ. The stochasticity is now separated from θ, allowing us to calculate $\nabla_\phi \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\mathcal{L}_{\theta,\phi}(\mathbf{x})]$ as follows:

$$\nabla_\phi \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\mathcal{L}_{\theta,\phi}(\mathbf{x})] = \nabla_\phi \mathbb{E}_{p(\epsilon)}[\mathcal{L}_{\theta,\phi}(\mathbf{x})] \tag{10}$$

$$= \mathbb{E}_{p(\epsilon)}[\nabla_\phi \mathcal{L}_{\theta,\phi}(\mathbf{x})] \tag{11}$$

References

[1] Carl Doersch, *Tutorial on variational autoencoders*, 2016.

[2] Diederik P. Kingma and Max Welling, *An introduction to variational autoencoders*, 2019.