DL-Assignment 10

Marta Gulida	5585808	mg776
Tillman Heisner	4517815	th273
Erik Bode	4505799	kb301

Exercise 2.1: Closed form forward diffusion process

derive μ and Σ for $q(x_3|x_0) = \mathcal{N}(x_3|\mu x_0, \Sigma I)$

As defined on the sheet, we know

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t|\sqrt{\alpha_t}\, x_{t-1}, (1-\alpha_t) I),$$

Using the reparametrisation trick, we can represent x_t given x_{t-1} and $\varepsilon_{t-1} \sim \mathcal{N}(\epsilon|0,I)$ as:

$$x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, \varepsilon_{t-1}$$

now, expanding x_{t-1} in a similar fashion, given x_{t-2}, $\varepsilon_{t-2} \sim \mathcal{N}(\varepsilon_{t-2}|0,I)$

$$x_t = \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_{t-1}}\, \varepsilon_{t-2}\right) + \sqrt{1-\alpha_t}\, \varepsilon_{t-1}$$

$$x_t = \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \sqrt{\alpha_t(1-\alpha_{t-1})}\, \varepsilon_{t-2} + \sqrt{1-\alpha_t}\, \varepsilon_{t-1}$$

We can now use the re. trick again, defining the distributions A and B

$$A := \mathcal{N}(0, \alpha_t(1-\alpha_{t-1})I) \quad B := \mathcal{N}(0, 1-\alpha_t I)$$

Adding both distributions: $\mathcal{N}(0, [\alpha_t(1-\alpha_{t-1})] + [1-\alpha_t] I)$

$$\rightsquigarrow \mathcal{N}(0, (1-\alpha_t\alpha_{t-1}) I)$$

Using the rep. trick again:

$$x_t = \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\, \varepsilon \quad \text{with } \varepsilon \sim \mathcal{N}(0,I)$$

Expanding x_{t-2} and substituting $\alpha_t \alpha_{t-1} = \gamma$

$$x_t = \sqrt{\gamma}\left(\sqrt{\alpha_{t-2}}\, x_{t-3} + \sqrt{1-\alpha_{t-2}}\, \varepsilon_{t-3}\right) + \sqrt{1-\gamma}\, \varepsilon$$

$$x_t = \sqrt{\gamma \alpha_{t-2}}\, x_{t-3} + \sqrt{\gamma(1-\alpha_{t-2})}\, \varepsilon_{t-3} + \sqrt{1-\gamma}\, \varepsilon$$

Using the same rep. trick, sum of G. and rep. trick as before;

$$x_t = \sqrt{\gamma \alpha_{t-2}}\, x_{t-3} + \sqrt{1-\gamma\alpha_{t-2}}\, \varepsilon$$

resubstitution:

$$x_t = \sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2}}\, x_{t-3} + \sqrt{1-\alpha_t\alpha_{t-1}\alpha_{t-2}}\, \varepsilon$$

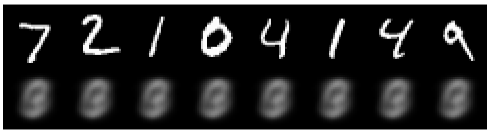
Setting t to 3, we now get the following values:

$$\mu = \alpha_3 \alpha_2 \alpha_1 \qquad \Sigma = 1 - \alpha_3 \alpha_2 \alpha_1$$

Exploring KL-divergence on the loss

1) What do you observe?

Results for kl_loss_weight set to 30:



The results are very blurry and it is difficult to distinguish any digits.

Results for kl_loss_weight set to 0:

Results are quite good, for some digits reconstruction results are still not so straightforward (4 or 9).

2) How can these results be explained?

kl_loss_weight is used as a hyperparameter which helps to balance KL-divergence loss and reconstruction loss. Setting the high value for this hyperparameter may lead to underfitting, when VAE is working more on minimizing the KL-divergence loss than actually reconstructing the input data.
When we are setting kl_loss_weight to 0, we are ignoring KL-divergence loss component (our regularization term). And this leads to the absence of possibility for VAE to act as a generative model (it will be focused only on minimizing the reconstruction loss).

3) What is the role of the KL divergence term?

It measures the difference between the learned distribution of the latent variables (the output of the encoder) and the prior distribution. It works like a regularization term.