Deeplearning HK10 Participants. 1° Hanzhi Zhuang 2° Rui Song 3 Tengyunhao Yang
5578347 5576534 5587382

1. 4) The train result of VAE with KL divergence weight 0 is better than that of VAE with KL divergence weight 30. In VAE with KL divergence weight 0, the features of different numbers are distinct, their positions are also more far away than those on VAE with $W_{KL}=30$. The generated samples exhibit more variability.

<u>explain</u>: A higher weight on KL divergence enforces a stronger regularization on latent space, making it more compact, reduce risk of overfitting. The model is encouraged to learn a more structured, organized latent space representation

<u>role of KL divergence</u>: 1° Regularization 2° Variational Inference.
3° control over latent space.

2. 1) For t=1:

$$q(x_1|x_0) = \mathcal{N}(x_1 \mid \sqrt{a_1}\, x_0 + \sqrt{1-\alpha_1}\, \epsilon_0 \;,\, (1-\alpha_1)I)$$

For t=2:

$$q(x_2|x_1) = \mathcal{N}(x_2 \mid \sqrt{a_2}\, x_1 + \sqrt{1-\alpha_2}\, \epsilon_1 \;,\, (1-\alpha_2)I)$$

So $$q(x_2|x_0) = \mathcal{N}(x_2 \mid \sqrt{a_2}(\sqrt{a_1}\, x_0 + \sqrt{1-\alpha_1}\, \epsilon_0) + \sqrt{1-\alpha_2}\, \epsilon_1 \;,\, (1-\alpha_2)I)$$

For t=3:

$$q(x_3|x_2) = \mathcal{N}(x_3 \mid \sqrt{a_3}\, x_3 + \sqrt{1-\alpha_3}\, \epsilon_2 \,,\, (1-\alpha_3)I)$$

So,

$$q(x_3|x_0) = \mathcal{N}\left(x_3 \,\middle|\, \sqrt{\alpha_3}\cdot\sqrt{a_2}\cdot\sqrt{a_1}\cdot x_0 + \sqrt{a_3}\cdot\sqrt{a_2}\cdot\sqrt{1-\alpha_1}\,\epsilon_0 + \sqrt{a_3}\sqrt{1-\alpha_2}\,\epsilon_1 + \sqrt{1-\alpha_3}\,\epsilon_2,\ (1-\alpha_3)I\right)$$

Simplification:

Let $\mu_{x_3} = \sqrt{a_3 a_2 a_1}\, x_0 + \sqrt{a_3 a_2}\sqrt{1-\alpha_1}\,\epsilon_0 + \sqrt{a_3}\sqrt{1-\alpha_2}\,\epsilon_1 + \sqrt{1-\alpha_3}\,\epsilon_2$

$\Sigma_3 = (1-\alpha_3)$

the closed form solution is: $q(x_3|x_0) = \mathcal{N}(x_3 | \mu_{x_3}, \Sigma_3 I)$

Generalization:

$$\mu_{x_t} = \sqrt{a_t!}\cdot x_0 + \sum_{i=0}^{t-1} \sqrt{\frac{a_t!}{a_{i+1}!}}\cdot\sqrt{1-\alpha_{i+1}}\,\epsilon_i$$

$$\Sigma = (1-\alpha_t)I$$