DEEP LEARNING EX 10

Uralp Ergin 5975013, İsmail Karabaş 7654321, Ahmet Yaşar Ayfer 5986167

Exploring the effect of KL divergence on Loss:

- What do we observe?

  When we include KL divergence in the loss with a weight factor of 30, we observe that the reconstruction performs poorly compared to the loss without KL divergence.

- How can these results be explained?

  KL divergence loss dominates the overall loss due to its high weight in penalizing the model when it deviates from the standard normal distrubition. So the encoder tries to minimize the KL loss term instead of learning the data and reconstruction patterns effectively. As a result the model can’t capture the underlying data structure that leads to poor reconstruction quality.

- What is the role of KL term?

  It helps regularizing the models latent space representation by approximating it to a standard normal distribution.

- We were using the env from the previous exercise (ex-9) so,we didn’t install the requirements txt and the ffmpeg package as it is stated in the beginning of the pdf (either use the previous env or create new one and install packages). It turns out the video creation part is not working with the current env setting and we didn’t want to gamble with the environment setting in order not to break anything. Thats why we are missing the video.

- Diffusion models are a class of generative models which has the idea of adding noise to the data incrementaly and try to learn how to generate the data from the noise. They differ from GANs in terms of training process and data generation where GANs use two networks namely generator and discriminator (one generates data from noise by random sampling and the other tries to distinguish fake data) and Diffusion Models generates data by reversing the diffusion process without any adverserial process

  Advantages:
- Higher Quality
- Better Coverage of Data
- Stability

Disadvantages:

- Computationally Expensive
- Slower Inference
- Parametrization Complexity

....../....../.......

- $x_1 = \sqrt{\alpha_1}\ x_0 + \sqrt{1-\alpha_1}\ \epsilon_0$

  $x_2 = \sqrt{\alpha_2}\ x_1 + \sqrt{1-\alpha_2}\ \epsilon_1$

  $x_3 = \sqrt{\alpha_3}\ x_2 + \sqrt{1-\alpha_3}\ \epsilon_2$

With sum of Gaussians :

$$\mu_3 = \sqrt{\alpha_3} \cdot \sqrt{\alpha_2} \cdot \sqrt{\alpha_1}\ x_0$$

$$\Sigma_3 = \sum_{t=1}^{3} (1-\alpha_t)$$

where $q(x_3|x_0)$:

$= \mathcal{N}(x_3 | \mu x_0, \Sigma I)$

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

for each t this can be expanded as:

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

we get $x_t = \prod_{i=1}^{t} \sqrt{\alpha_i}\ x_0 + \sum_{i=1}^{t} \sqrt{1-\alpha_i} \prod_{j=i+1}^{t} \sqrt{\alpha_j}\ \epsilon_{i-1}$

for convenience :

$\bar{\alpha}_t = \Pi_{i=1}^{t} \alpha_i$

$\bar{\beta}_t = 1 - \bar{\alpha}_t$

final expression for x_t :

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon$$

mean : $\mu_t = E[x_t] = \sqrt{\bar{\alpha}_t}\, x_0$

Variance : $\Sigma_t = Var(x_t) = 1-\bar{\alpha}_t$

so $q(x_t \mid x_0) = N(x_t \mid \sqrt{\bar{\alpha}_t}\, x_0 , (1-\bar{\alpha}_t) I)$