KAIST

AI502 Deep Learning

Homework 4 - VAE (Variational Autoencoders)

Cappa Victor
20206080
School of computing
victor.cappa@kaist.ac.kr

HW4 - Intro on Variational Autoencoders and MNIST dataset

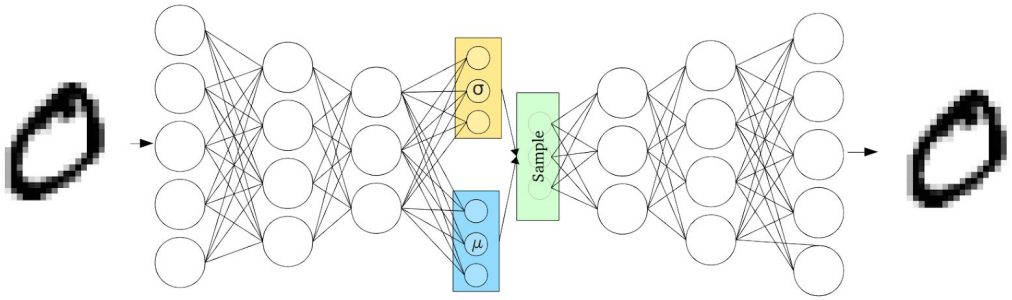


Figure1 - VAE (Variational Autoencoder)

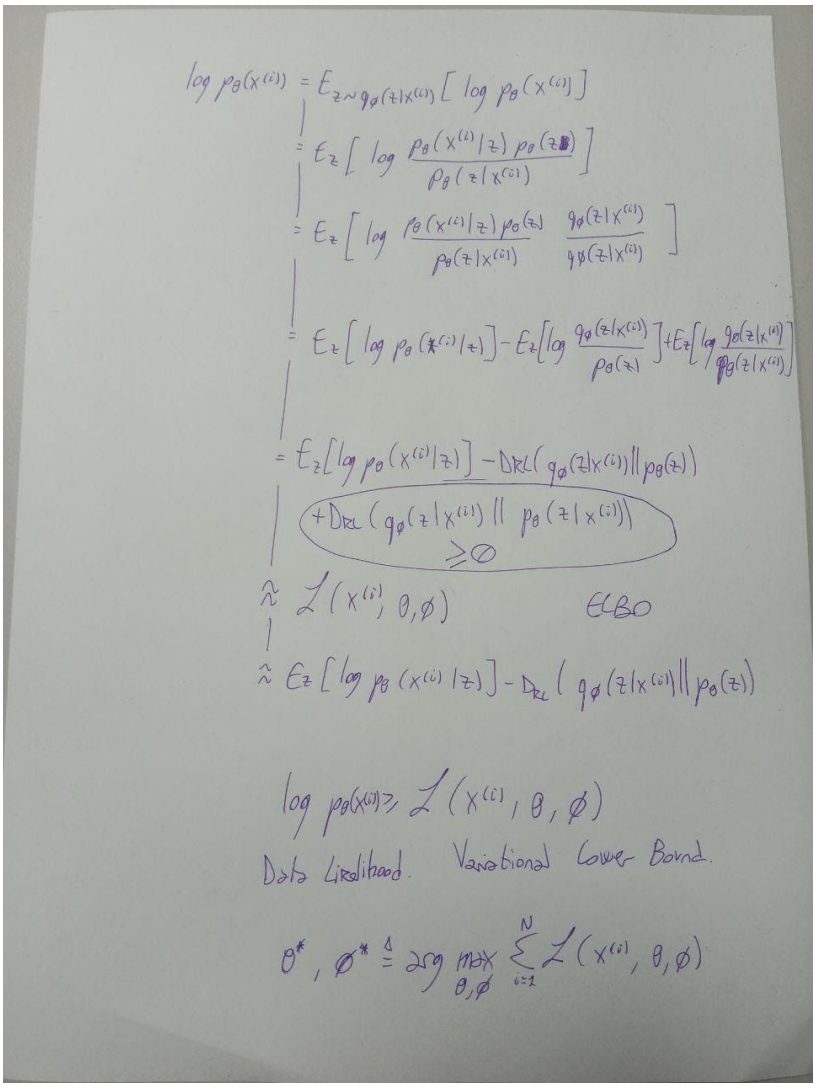
The Variational Autoencoder is a deep generative model performing explicit-approximate density estimation, able during inference to generate new samples starting from a fixed-size input (the latent space).
It can be divided in two different main components: the decoder and the encoder, both of them are fully-connected neural networks. During inference we are only interested in the decoder (generating the output image), while during training we are interested in both of the components. The loss function we want to maximize is given by the sum of two different terms, the reconstruction term (measuring the “distance” between the inputs and outputs) and the KLD term also called regularization term (regularization because it makes the output of the encoder, the latent space, “regular”, meaning we want it to be distributed according to a prior probability distribution, usually a Normal distribution with unitary variance and zero mean).The size of the hidden dimension can vary and also the weight of the KLD term can be tuned as an hyperparameter.
The Reparametrization trick is used in order to make the gradient descent possible despite the random sapling of the latent space.
The code implementation details can be found in the “HW4_20206080.ipyb” file.
The MNIST dataset is a handwritten digits dataset of images, composed of a total of 70000 samples (60000 as the training set and 10000 as the test set). Each image is of size 28x28.

HW4-Problem1 - Derivation of objective function

The following results are obtained from Stanford lecture 13 on Generative models (https://www.youtube.com/watch?v=5WoItGTWV54).

The purpose of VAE is maximize the likelihood of our data, and by deriving it we find out that we can maximize over a lower bound ("ELBO"), due to the fact that Kullback Leibler Divergence terms are always positive.

The results are reported in the following picture:



HW4-Problem2 - Qualitative comparison the results of VAE according to the size of the Hidden dimension Z

For original VAE (weight parameter of KLD term equal to 1), we perform a qualitative comparison of the generated random sample images with respect to the size of the hidden dimension Z and describe the experimental results for #Z = 2, 10, 25, and 50. The values of the Z dimension are sampled from a Normal probability distribution with zero mean and unit variance.

For #Z equal to 2 (left) and 10 (right) we generate the following images:

For #Z equal to 25 (left) and 50 (right) we generate the following images:

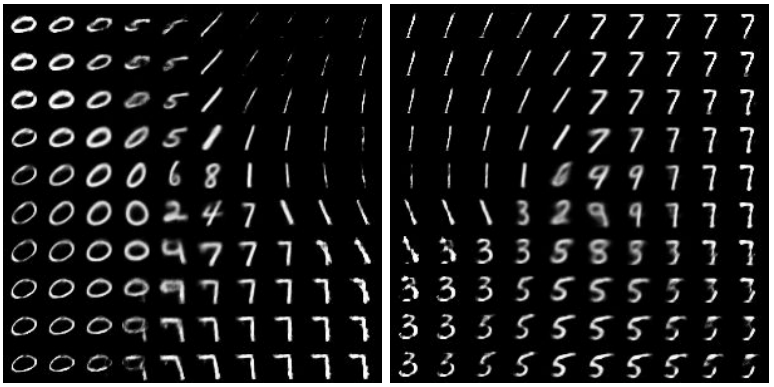
As we can see the images are randomly generated. We can observe that the quality of the output images varies according to the size of the latent space, and in particular by increasing its size we don't generate more high quality images. For #Z=25, 50 the experimental results are very similar, with some very clear and understandable and some non-understandable numbers, proposing that having a higher dimensional space for Z doesn't correspond to higher quality image inferences. It is interesting to observe that for #Z=2 most of the numbers are readable (with a few exceptions) but the numbers are blurrier in general compared to the other configurations. This is due to the smaller size of the space of Z, where different numbers of the input dataset could overlap on each others due to the smaller space size, thus we have blurrier generated images.

By having a higher dimensional latent space we could generate more clear images, but we risk also to generate not readable ones, and by having a small dimensional latent space we are more likely to generate blurrier, less definite images.

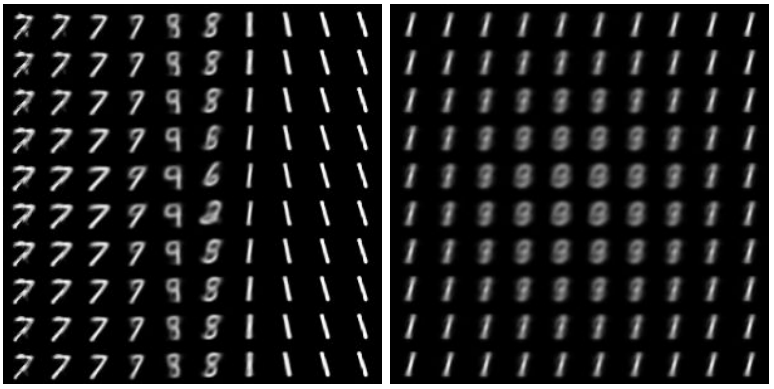
HW4-Problem3 - Qualitative comparison the results of VAE according to the weight of KLD term

Finally, we perform a qualitative comparison and description of the generated sample images by keeping the size of the hidden dimension Z equal to 2 and by varying the weight of the KLD term in loss function for 1, 5, 10, and 40. Moreover, we will choose the values of the Z hidden dimension (with #Z=2 fixed) in a grid fashion, meaning that we will not sample its values from a gaussian probability distribution but we will generate its values with the torch.linspace() function. In particular, we will choose values in range -4 to +4 and analyze the experimental results.

For weight equal to 1 (left) and 5 (right) we generate the following images:



For weight equal to 10 (left) and 40 (right) we generate the following images:



By varying the weight parameter we are varying the strength of the optimization of the KLD term (regularization term) with respect to the log-likelihood (reconstruction term). As a result, in our implementation, by increasing the value of the weight we increase the strength of the regularization term, thus making the decoder approximate the input images as much as possible to the Normal distribution with unit variance and zero mean. This theoretical explanation can be demonstrated with our experimental results, where for the highest regularization strength (weight=50) we have an overlapping of different input numbers in the center of the grid-latent space where both z1 and z2, with Z=[z1, z2], are close to zero, so during inference we generate blurrier images. On the other hand for smaller values of the weight we have clearer generated numbers, proposing that maybe the best value for this parameter should be found in this interval.

In situations were the input data is more complex and the hidden dimension is larger we can increase the weight parameter such that the input data is more prone to be encoded in the latent space with respect to the Normal distribution with unit variance and zero mean, so we are allowing the input data to be less widespread in the latent dimension space..