VAE Problem Set

Group U8: Lake Yin, Zhiqi Wang

April 22, 2023

In the lecture of VAE, we've learned that the variational autoencoder (VAE) as a tweak of autoencoder with given objective. In this problem set we will explore how ELBO is the objective function of VAE.

Problem 1

The equation of log likelihood we used in the lecture is:

$$\log p_\theta(x) = \mathcal{L}_{\theta,\phi}(x) + KL(q_\phi(z|x)||p_\theta(z|x))$$

The first term on RHS is the ELBO, which is the objective function to use for VAE. Rewrite this equation to explain why that would be called evidence lower bound (ELBO).

Problem 2

Here's the code implementation of VAE in PyTorch. (csci4968-VAE-project). In assignment 4, we have done an reconstruction of Fashion-MNIST dataset. In this problem, we will build a VAE on assignment 4 based on our implementation of VAE.

1. VAE class

First, you should take our VAE class object and modify it to fit your code in assignment 4. Feel free to start your implementation from scratch if you want.

For the VAE class to work, you will need to implement `decoder` and `encoder` class. Both of those class will need a `forward()`. You don't have to use convolution layers in your encoder and decoder, but you can if you want. You can also use fully connected layers as what we did in assignment 4.

2. Train VAE

The training of VAE is similar to the training of autoencoder. If you want to use many layers (unlike in assignment 4, we've only used 2 layers, which is not too computational expensive), you could try to train with gpu.

3. Reconstruction

As what we've done for assignment 4, you will be reconstructing image in the test set and show the reconstruction loss. You should also reconstruct 10 images in a 2 x 5 grid as how you did in assignment 4.

Keep in mind that even though we are using VAE to reconstruct images, this is not what VAE is for. VAE is a generative model as we use the decoder with new latent variables that's unseen from training.

4. Generating new data

Now you need to generate new image by sampling from normal distribution. Remember that you should sample latent variables z from normal distribution and then use the decoder to generate a new image. You should generate 10 images in a 2 x 5 grid.