Homework 4

June 12, 2020

[AI502/KSE527] Deep Learning
TA Jongwoo Ko

HW Description

In this homework, you will study about variational autoencoder networks by using MNIST dataset.

The contents of this homework consist of:

- **Task1** : Design the autoencoder structured network for MNIST
- **Task2** : Design the loss function for autoencoder with weight of KLD term
- **Task3** : Qualitative comparison depends on hidden dimension size, weight of KLD term

After you make a code for each task, you should write a report according to the each problem.

Source Code
Download homework4.zip from KLMS that contains the source code for three tasks.

Before tasks : GPU setting

Because this homework takes a lot of time if you do on CPU, we recommend you to use GPU. If you do not have a GPU machine, please use Google Colab. The manual about how to use GPU on Colab is included in homework2.zip file. (Also, the provided code is GPU-based code.)

Before tasks : Change the indent size (For Colab User)

For python, the indent size is very important. If indent is not proper, program will not run because of syntax error. Anyway, the indent size of our code is 4. When you use Colab first, this size might be set 2. Therefore, you have to change the indent size from 2 to 4 on Colab. In our Colab manual file, there is the way to change this size.

Task1 : Design the VAE structured network for MNIST

Please implement the fully connected network depicted in the picture below. In stochastic sampling procedure, it would be nice to generate a random variable for reparameterization trick in the standard normal distribution, $\mathcal{N}(0,1)$.

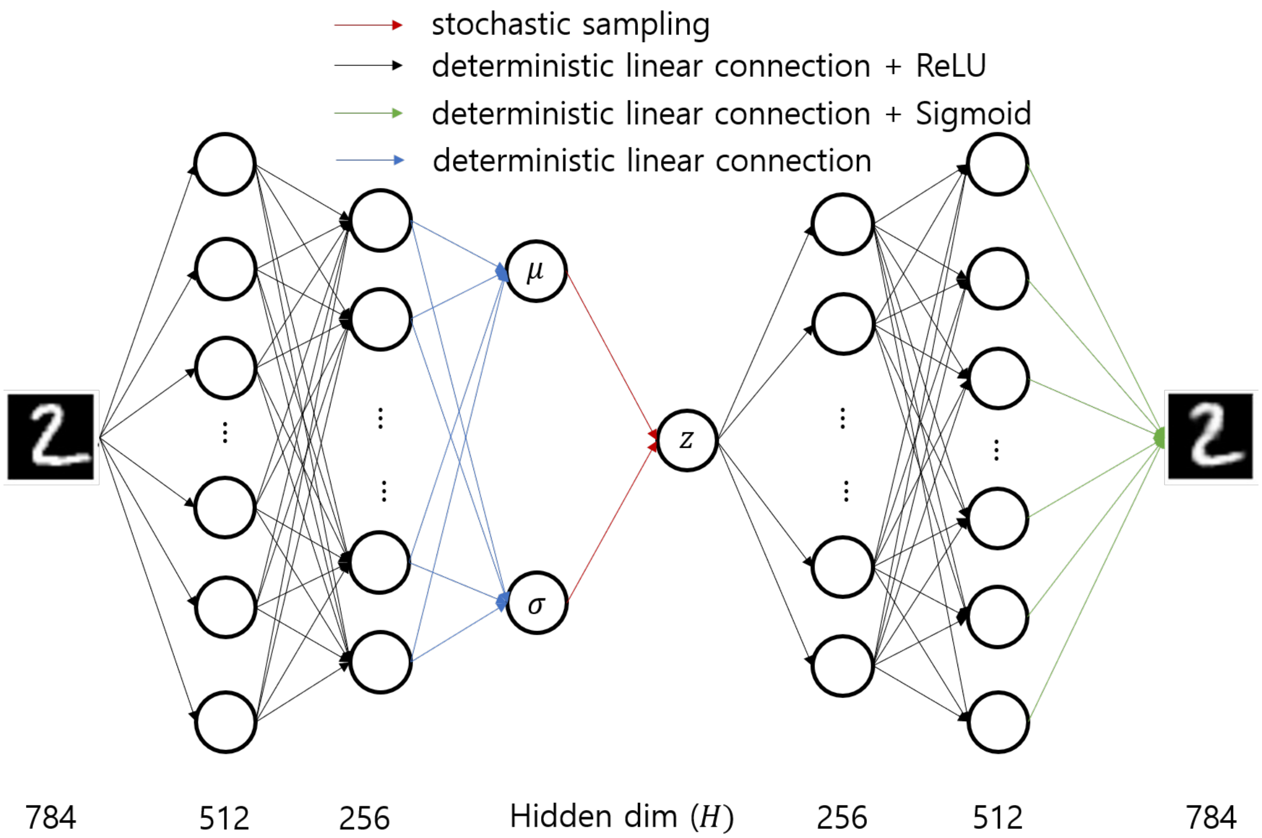


Figure 1: Variational AutoEncoder for MNIST

Task2: Design the loss function for autoencoder with weight of KLD term

In original VAE paper, loss function for VAE is as follows:

$$\mathcal{L}(\theta, \phi; \mathbf{x}^{(i)}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x}^{(i)})}[\log p_\theta(\mathbf{x}^{(i)}|\mathbf{z})] - D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}^{(i)})||p_\theta(\mathbf{z})),$$

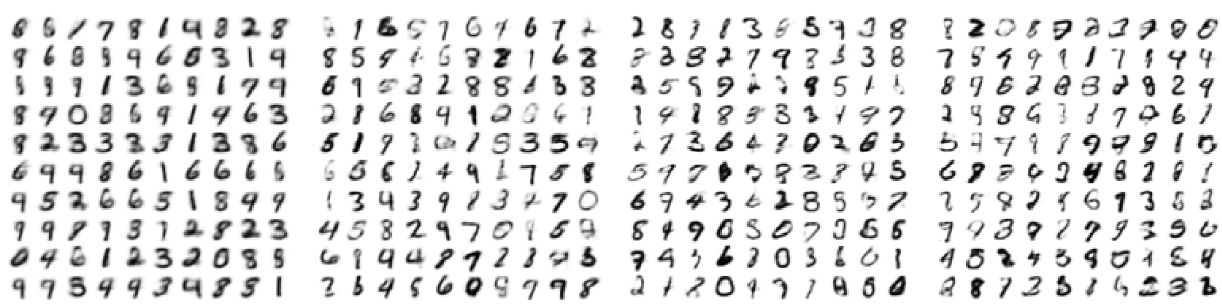
where the ϕ is the set of variational parameter and θ is the set of generative parameters. For optimizing these parameter sets, we want to differentiate and optimize the lower bound $\mathcal{L}(\theta, \phi; \mathbf{x}^{(i)})$. However, we will use slightly different loss function for VAE:

$$\mathcal{L}^*(\theta, \phi; \mathbf{x}^{(i)}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x}^{(i)})}[\log p_\theta(\mathbf{x}^{(i)}|\mathbf{z})] - \lambda D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}^{(i)})||p_\theta(\mathbf{z})),$$

Please implement the loss function for VAE and optimize your own loss function.

Task3: Qualitative comparison depends on hidden dimension size, weight of KLD term

We want to conduct some ablation studies using our designed network and loss function. For qualitative comparison, you have to plot the generated images for random sampled latent points (hidden dimension) or grid sampled latent points (weight of KLD term).



(a) 2-D latent space (b) 5-D latent space (c) 10-D latent space (d) 20-D latent space

Figure 2: Examples of generated images for random sampled latent points

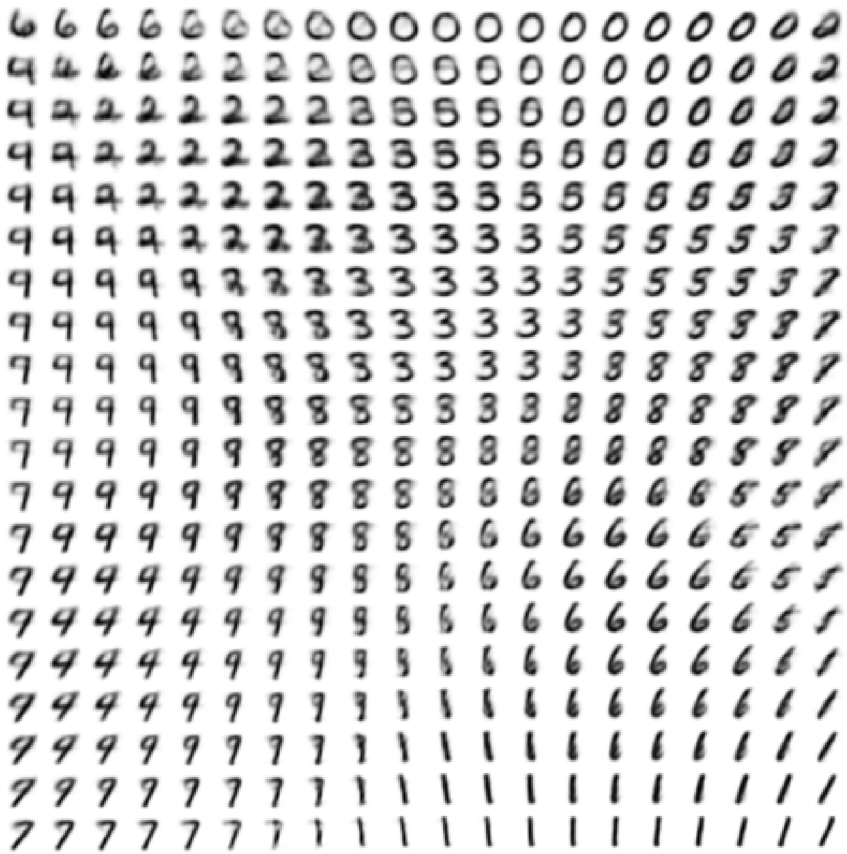


Figure 3: For latent variable $\mathbf{z}$ with hidden dimension size 2, location for each image means value of corresponding latent variable. In problem 3, you need to sample the images as in this figure. (NOT random sampling)

Homework 4

June 12, 2020

Problem 1 (6pt)

- The original objective function is $\min_{\phi,\theta} D_{KL}(q_\phi(z|x)||p_\theta(z|x))$. Derive $\text{argmin}_{\phi,\theta} D_{KL}(q_\phi(z|x)||p_\theta(z|x)) = \text{argmax}_{\phi,\theta}\mathcal{L}(\theta,\phi;\mathbf{x}^{(i)})$ (3 points)
- For constrained optimization problem, $\min_{\phi,\theta} D_{KL}(q_\phi(z|x)||p_\theta(z|x)) \quad \text{s.t. } D_{KL}(q_\phi(z|x)||p(z)) < \epsilon$ can be approximated to $\max_{\phi,\theta} \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - \lambda\, D_{KL}(q_\phi(z|x)||p(z))$. Derive the formula by using Lagrange multiplier method. Note that ϵ is positive real number. (You can use above result) (3 points)

Problem 2 (4pt)

- For original VAE ($\lambda = 1$), perform a qualitative comparison of samples according to the size of the hidden dimension and describe the results. ($H = 2, 10, 25, 50$) (2 points)
- What can be strength and weakness for VAE with large hidden dimension and small hidden dimension? (2 points)

Problem 3 (5pt)

- Perform a qualitative comparison the results of VAE according to the weight of KLD term in loss function for $H = 2$. ($\lambda = 1, 5, 10, 40$) In this problem, you **NEED** to sample the latent variables in a row for each axis as described in Figure 3. (2 points)
- What is the difference between VAEs with different KLD term weights? What is the advantage of increasing the weight of the KLD term in a situation where the data is more complex and the hidden dimension is larger? (3 points)

About the Submission

- The deadline for submission is **23:59 on 26 June (Fri)**, and late submission is not permitted.
- You have to submit **.zip file** including both **.ipynb file** and **.pdf file**.
  (Please convert .doc file to **.pdf file**)
- File name should be **[hw4]student_ID.zip** (e.g., [hw4]20201234.zip)
  (If you do not keep this naming, there will be a disadvantage.)

References