Tsinghua University

Department of Computer Science and Technology

Deep Learning

Homework 7

Sahand Sabour

2020280401

Class Conditional Variational Autoencoder

In this homework, we were required to imeplement a class-conditional VAE model using the ZhuSuan library,and test it on the MNIST dataset. In the MNIST dataset, there are 10 possible labels for the samples (0-9). Binarizing the labels with the one-hot encoding method, gives a sequence of 10 digits with one 1 and nine 0s. Hence, there could be 10 locations for the 1; the probability of a label l to be one of the 10 labels L would be p(l = L) = $\frac{1}{10}$ = 0.1. According to the lecture notes, the variational lower bound for the normal case of VAE was obtained as:

$$L(\theta, x) = E_{q(z|x)}[logp(z, x; \theta) - logq(z|x)] = E_{q(z|x)}[logp(x|z; \theta)] - KL(q(z|x)||p(z; \theta))$$

However, it can be noticed that the output of this equation is only dependent on the latent variable z and therefore, does not produce any specific results, which is not practical for our case. Hence, we should modify the lower bound to include the label l of the sample we would like to generate likewise.

$$L(\theta, x, l) = E_{q(z|x,l)}[logp(x, l|z; \theta)] - KL(q(z|x, l)||p(z; \theta))$$

Since $z \sim \mathcal{N}(0, 1)$ for Gaussian, the KL-divergence is as follows:

$$-KL(q(z|x, l)||p(z; \theta)) = \frac{1}{2}(1 + log\sigma^2 - \mu^2 - \sigma^2) \tag{1}$$

Consequently, the expected log-likelihood would be

$$E_{q(z|x,l)}[logp(x, l|z; \theta)] = E_{q(z|x,l)}[-\sum_j \frac{1}{2}log\sigma_j^2 + \frac{(x_{ij} - \mu_{xi})^2}{\sigma^2}] \tag{2}$$

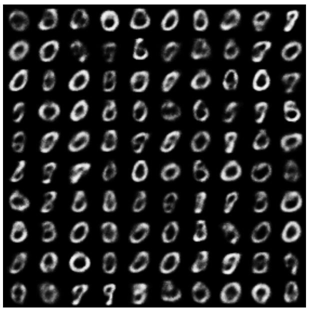
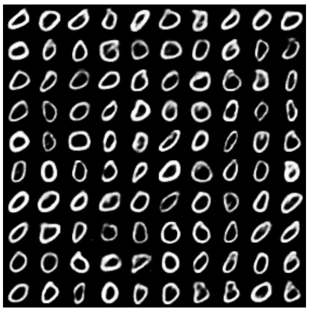
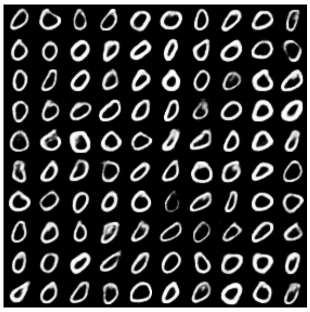
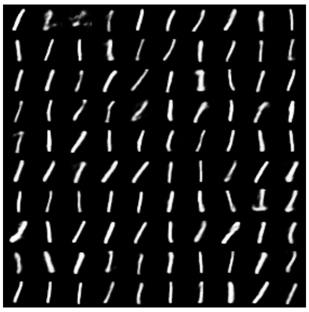
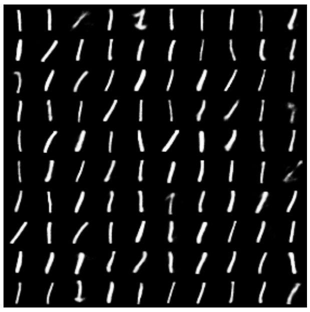
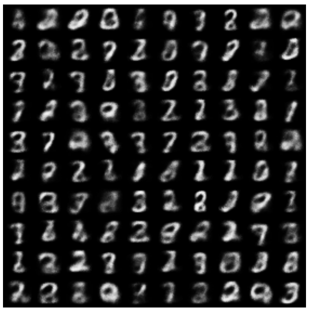
Approximating the above equation with Monte Carlo methods gives

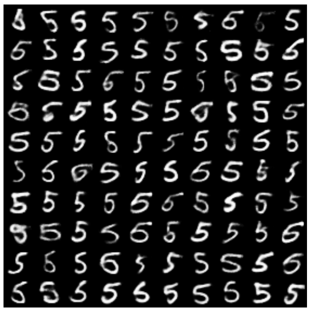
$$E_{q(z|x,l)}[logp(x, l|z; \theta)] \approx \frac{1}{L}\sum_k logp(x, l|z^{(k)}) \quad \text{where} \quad z^{(k)} \sim q(z|x, l) \tag{3}$$

where $z^{(k)}$ is a random variable, which cannot be used for back-propagation. Hence, by utilizing re-parameterization techniques, we have $z^{(k)} = \mu(x, l) + \sigma(x, l).\epsilon^{(k)} = g(x, l, \epsilon^{(k)})$, where g is a deep neural network. The lower bound becomes

$$L(\theta, x, l) = E_{p(\epsilon)}[log\frac{p(g(x, l, \epsilon), x; \theta)}{q(g(x, l, \epsilon)|x; \theta)}] - KL(q(z|x, l)||p(z; \theta))$$

$$L(\theta, x, l) = \frac{1}{L}\sum_k logp(x, l|z^{(k)}) + \frac{1}{2}\sum_{i=1}^{j}[1 + log\sigma^2 - \mu^2 - \sigma^2]$$

Digit	Epoch 1	Epoch 50	Epoch 100
0			
1			
2			
3			
4			

Digit	Epoch 1	Epoch 50	Epoch 100
5			
6			
7			
8			
9			

The model was trained on the MNIST dataset and the obtained lower bound values for different number of epochs are provided respectively in the table below:

Epoch	1	10	25	50	100
Lower Bound	-167.45	-97.954	-92.546	-90.108	-88.361

Table 1: Table of lower bound based on given epoch

In addition, the obtained digit generation results are provided respectively above. Tensorflow 1.15 was used to run this implementation.