Variational Autoencoder by ZhuSuan

Instructor: Prof. Jun Zhu

TA: Cheng Lu (lucheng.lc15@gmail.com)

1 Introduction

MNIST[1] digits dataset is a widely used dataset for image classification and density modeling in machine learning field. It contains 60,000 training examples and 10,000 testing examples. The digits have been size-normalized and centered in a fixed-size image. Each example is a 784×1 matrix, which is transformed from an original 28×28 grayscale image. Digits in **MNIST** range from 0 to 9. Some examples are shown below. **Note**: During training, information about testing examples should never be used in any form.

Variational Autoencoder (VAE) [1] is one of the most widely used deep generative models. In this homework, you are to imeplement a VAE model using the ZhuSuan library[2], and run it on the **MNIST** dataset. **We will use Jupyter Notebook in this homework**. You need to implement the requirements and write your codes in **vae.ipynb**. To get started with ZhuSuan, follow this tutorial for basic concepts in ZhuSuan.

2 Variational Autoencoders

The generative model is defined as follows:

$$\begin{aligned} \mathbf{z} &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d) \\ \mathbf{x}_{\text{logits}} &= \text{NN}_\theta(\mathbf{z}) \\ \mathbf{x} &\sim \text{Bernoulli}(\mathbf{x}|\text{sigmoid}(\mathbf{x}_{\text{logits}})) \end{aligned} \tag{1}$$

where $\mathbf{z}$ and $\mathbf{x}$ are random variables. $\mathbf{z} \in \mathbb{R}^d$ is the latent representation. $\mathbf{x} \in \{0, 1\}^{784}$ denotes the image. $\text{NN}_\theta(\cdot)$ is a mapping from $\mathbf{z}$ to $\mathbf{x}$, parameterized by a neural network with parameter θ.

[1] http://yann.lecun.com/exdb/mnist/

[2] https://zhusuan.readthedocs.io/en/latest/

In the problem we fix $d = 40$. Given the training set of **MNIST** images $\mathcal{D} = (\mathbf{x}_i)_{i=1}^N$, we need to do maximum likelihood learning of the network parameters

$$\max_{\boldsymbol{\theta}} \log p_\theta(\mathcal{D}) = \sum_{i=1}^N \log p_\theta(\mathbf{x}_i).$$

However, the model we defined has not only the observation $\mathbf{x}$ but also latent representation $\mathbf{z}$. This makes it hard for us to compute $p_\theta(\mathbf{x})$, which we call the marginal likelihood of $\mathbf{x}$, because we only know the joint likelihood of the model:

$$p_\theta(\mathbf{x}, \mathbf{z}) = p(\mathbf{z}) p_\theta(\mathbf{x}|\mathbf{z})$$

while computing the marginal likelihood requires an integral over latent representation, which is generally intractable:

$$p_\theta(\mathbf{x}) = \int p_\theta(\mathbf{x}, \mathbf{z}) \mathrm{d}\mathbf{z}$$

The intractable integral problem is a fundamental challenge in learning latent variable models like VAEs. One method to address this problem is Variational Inference. The main idea is to maximize a lower bound of $\log p_\theta(\mathbf{x})$:

$$\mathcal{L}_{\theta,\phi}(\mathbf{x}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \mathrm{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z})), \tag{2}$$

where $q_\phi(\mathbf{z}|\mathbf{x})$ is a user-specified distribution of $\mathbf{z}$ (called variational posterior) that is chosen to match the true posterior $p_\theta(\mathbf{z}|\mathbf{x})$. This lower bound is usually called Evidence Lower Bound (ELBO).

In variational autoencoder, the variational posterior $q_\phi(\mathbf{z}|\mathbf{x})$ is also parameterized by a neural network with parameter ϕ, which accepts input $\mathbf{x}$, and outputs the mean and variance of a Normal distribution:

$$\mu_{\mathbf{z}}(\mathbf{x}; \phi), \log \sigma_{\mathbf{z}}(\mathbf{x}; \phi) = \mathrm{NN}_\phi(\mathbf{x})$$
$$q_\phi(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}|\mu_{\mathbf{z}}(\mathbf{x}; \phi), \sigma^2_{\mathbf{z}}(\mathbf{x}; \phi))$$

To train this model, ZhuSuan has implemented the **Stochastic Gradient Variational Bayes** (SGVB) estimator from the original paper of variational autoencoders[1]. This estimator takes benefits of a clever reparameterization trick to greatly reduce the variance when estimating the gradients of the ELBO. The only thing you need to do is to construct the generative model and the variational posterior model.

3 Requirements

1. Prove that $\log p_\theta(\mathbf{x}) \geq \mathcal{L}_{\theta,\phi}(\mathbf{x})$, and the equation is satisfied if and only if $q_\phi(\mathbf{z}|\mathbf{x}) = p_\theta(\mathbf{z}|\mathbf{x})$ for $\forall \mathbf{z} \in \mathbb{R}^d$, i.e. the variational posterior equals to the true posterior.
2. Following the instructions in **vae.ipynb**, implement the model using ZhuSuan, and train the model on the whole training set of **MNIST**.
3. Visualize the generations of your learned model. Include a few samples in your report.

4 Attention

You are required to complete the '# TODO' parts in **vae.ipynb**. You need to submit **vae.ipynb** and a short report **named as report.pdf** with the following requirements:

- We recommend that you typeset your report using appropriate software such as LaTeX. If you submit your handwritten version, please make sure it is cleanly written up and legible.
- Do not paste a lot of codes in your report (only some essential lines could be included).
- Please include experiment results using figures or tables in your report, instead of asking the TA to run your code.
- **Plagiarism is not permitted**. Please finish your homework independently.

References

[1] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014.