Tradeoffs in Neural Variational Inference

Ashwin D'Cruz[1], Sebastian Nowozin[2], Bill Byrne[1]
[1]Cambridge Engineering Department, [2]Microsoft Research Cambridge

OBJECTIVES

There have been several proposals on how to improve the inference network of the variational auto-encoder which would improve the model's capabilities. We will compare these different approaches on a range of applications to provide practical advice on the different trade-offs involved with each model with regards to accuracy and runtime.

Introduction

Variational auto-encoders (VAEs) [1] are an efficient probabilistic deep learning method that has gained a lot of popularity recently. They are widely applied due to their ease of use and promising results.
The basic VAE model is shown in Figure 4. The quality of inference and the generative process are dependent on the accuracy of the inference network.

Improving Variational Auto-Encoders

We will compare the following models:

- Auxiliary Deep Generative Model (ADGM) [2]
- Householder Flow [3]
- Importance Weighted Auto-Encoder (IWAE) [4]
- Non Linear Independent Components Estimation (NICE) [5]
- Normalizing (Planar) Flow [6]
- Markov Chain Monte Carlo (MCMC) and Variational Inference (VI) [7]

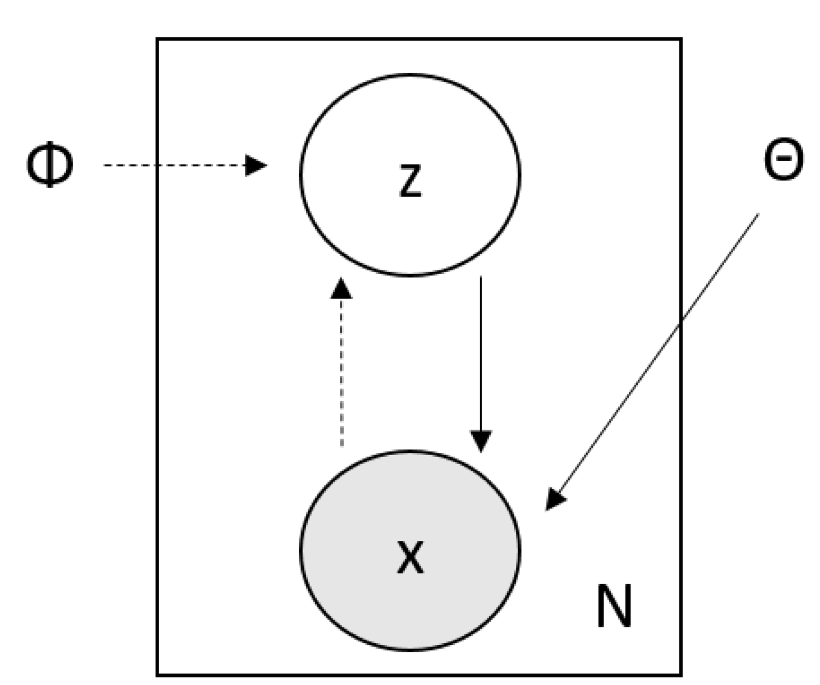

Figure: VAE model.

IWAE

The IWAE maximizes a different lower bound that is shown to be a tighter approximation than the standard VAE lower bound. This tighter bound is given by:

$$\mathcal{L}_k(x) = \mathbb{E}_{z\sim q(z|x)}\left[log\frac{1}{k}\sum_{i=1}^{k}\frac{p(x,z)}{q(z|x)}\right] \quad (1)$$

Flow Transformations

Normalizing flows can be divided into general normalizing flows and volume preserving flows. The tradeoff between these two is that the former is computationally more expensive but allows for more flexible posteriors. We will consider the Planar Flow for the general normalizing flow case and the Householder Flow for the case of a volume preserving flow.

Householder Flow Transformation

The transformation is defined as follow:

$$\begin{aligned} z^{(t)} &= \left(I - 2\frac{v_t v_t^T}{||v_t||^2}\right) z^{(t-1)} \\ &= H_t z^{(t-1)} \end{aligned} \quad (2)$$

v is known as the Householder vector. Importantly, H_t is an orthogonal matrix which simplifies the objective calculations.

Householder Flow Pipeline

The pipeline can be described with the following:

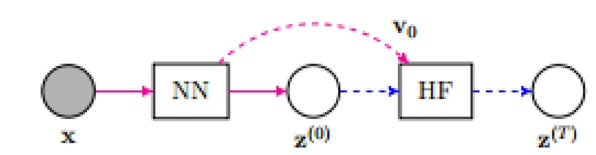

Figure: Encoder network and Householder flow. Image obtained from [3].

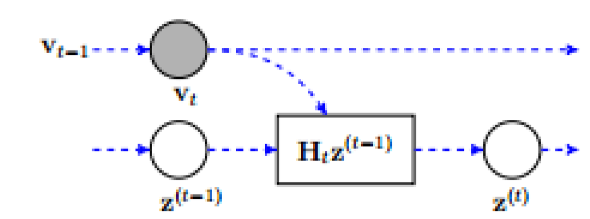

Figure: Single iteration of the Householder flow. Image obtained from [3].

Planar Flow

We will consider transformations of the following form:

$$f(z) = z + uh(w^T z + b) \quad (3)$$

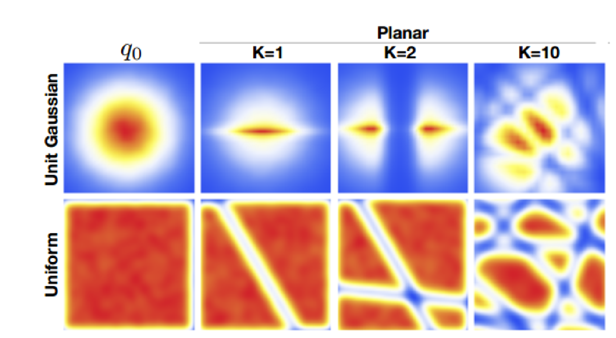

Figure: Effect of planar flow on two distributions. Image obtained from [6].

Sampled Data

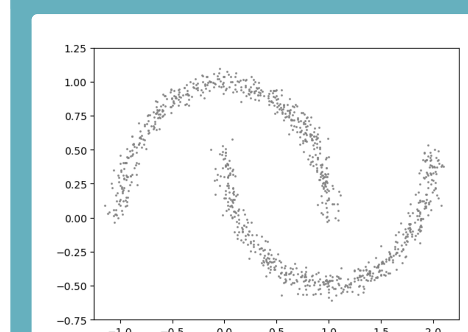

Figure: Training data

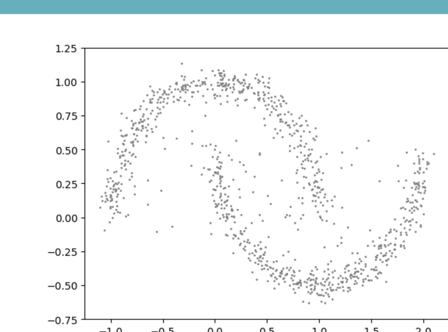

Figure: VAE

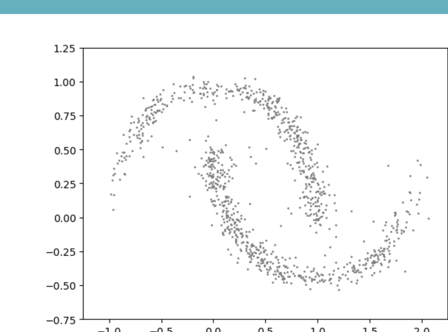

Figure: IWAE

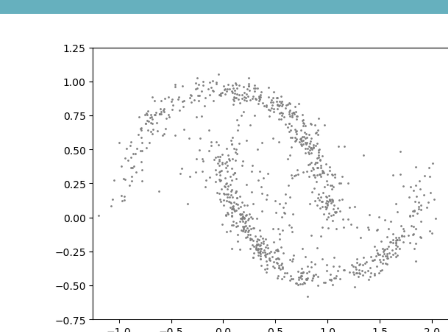

Figure: Householder

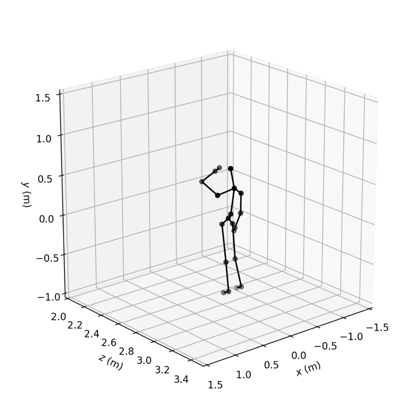

Figure: Training data

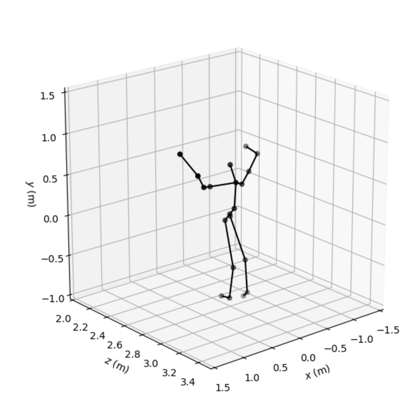

Figure: VAE

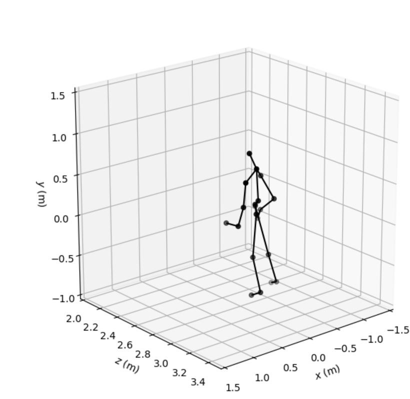

Figure: IWAE

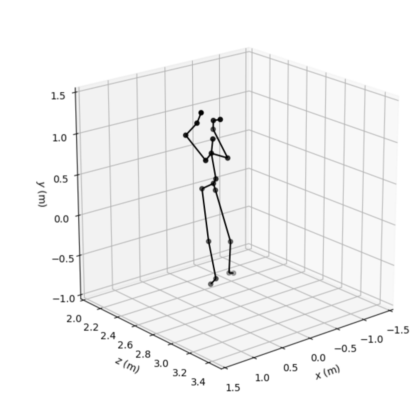

Figure: Householder

Project Status

Completed:

- Implemented basic VAE model.
- Implemented Householder flow model.
- Implemented IWAE.
- Implemented Normalizing (planar) flow model.
- Compared models on pose dataset.

Upcoming:

- Implement ADGM.
- Implement NICE.
- Implement MCMC and VI.
- Compare models on other datasets.

References

[1] Diederik P Kingma and Max Welling.
Auto-encoding variational bayes.
arXiv preprint arXiv:1312.6114, 2013.

[2] Lars Maaløe, Casper Kaae Sønderby, Søren Kaae Sønderby, and Ole Winther.
Auxiliary deep generative models.
arXiv preprint arXiv:1602.05473, 2016.

[3] Jakub M Tomczak and Max Welling.
Improving variational auto-encoders using householder flow.
arXiv preprint arXiv:1611.09630, 2016.

[4] Yuri Burda, Roger Grosse, and Ruslan Salakhutdinov.
Importance weighted autoencoders.
arXiv preprint arXiv:1509.00519, 2015.

[5] Laurent Dinh, David Krueger, and Yoshua Bengio.
Nice: Non-linear independent components estimation.
arXiv preprint arXiv:1410.8516, 2014.

[6] Danilo Jimenez Rezende and Shakir Mohamed.
Variational inference with normalizing flows.
arXiv preprint arXiv:1505.05770, 2015.

[7] Tim Salimans, Diederik P Kingma, Max Welling, et al.
Markov chain monte carlo and variational inference: Bridging the gap.
In *ICML*, volume 37, pages 1218–1226, 2015.

CONTACT

add39@cam.ac.uk