

Generator-Aware Bayesian Variational Autoencoders

July 2017

Abstract

In some scenarios, it is important to have uncertainty in the generator. However, naively using a BNN as the generator will be limited by (mutual info argument). Here we show how to build a GABVAE.

1 Introduction

VAEs [?] overfit in both the recognition and generator networks. The recognition overfitting is solvable. The generator overfitting gets worse with higher latent dimensionality Q and input dimensionality D .

A BVAE is a VAE where we are being Bayesian with the parameters θ of the generator network.

Meaning of z is dependent on θ . Therefore the recognition networks needs to be generator-aware. Here we introduce the Generator-Aware Bayesian Variational Autoencoders (GABVAE).

2 Generator-Aware Bayesian VAE

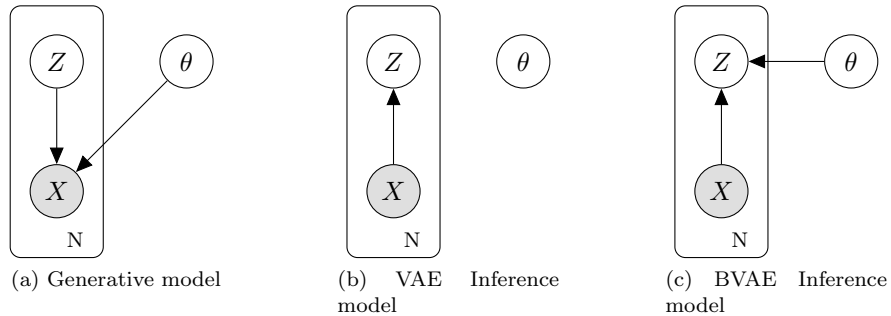


Figure 1: Probabilistic graphical models

2.1 How to summarize the generator?

- Pass the whole network (hypo-network)
- Pass the latent variable (MNF)
- Other summary statistics

Another option is to use HVI [?].

2.2 Mutual Info

Let $\mathbf{x}_i \in \mathbf{X}$ denote the i -th data point of the training data, $\mathbf{z}_i \in \mathbf{Z}$ its corresponding latent representation and θ the parameter variables of the generator network. Assuming that the true posterior factorises as follows,

$$p(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n, \theta | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = p(\theta | \mathbf{X}) p(\mathbf{Z} | \mathbf{X}, \theta) \quad (1)$$

$$= p(\theta | \mathbf{X}) \prod_{i=1}^N p(\mathbf{z}_i | \mathbf{x}_i, \theta) \quad (2)$$

Then, some optimal approximate posterior with the same factorial structure, $q^*(\mathbf{Z}, \theta | \mathbf{X}, \phi_{ga}) = q^*(\theta | \mathbf{X}, \phi_{ga}) \prod_{i=1}^N q^*(\mathbf{z}_i | \mathbf{x}_i, \theta, \phi_{ga})$, parameterised by ϕ_{ga} (“ga” stands for generator-aware) from a rich enough variational family would be able to match the true posterior and set the KL-divergence to zero. More formally,

$$D_{KL} \left(q^*(\theta | \mathbf{X}, \phi_{ga}) \prod_{i=1}^N q^*(\mathbf{z}_i | \mathbf{x}_i, \theta, \phi_{ga}) || p(\mathbf{Z}, \theta | \mathbf{X}) \right) = 0 \quad (3)$$

However, if our approximate posterior does not take into account θ when predicting \mathbf{z} given \mathbf{x} , it is possible that the KL-divergence remain positive even under the optimal approximation, however rich the variational family may be.

$$D_{KL} \left(q^*(\theta | \mathbf{X}, \phi_{ga}) \prod_{i=1}^N q^*(\mathbf{z}_i | \mathbf{x}_i, \phi_{ga}) || p(\mathbf{Z}, \theta | \mathbf{X}) \right) > 0 \quad (4)$$

The loss of information could be measured by the KL-divergence between these two approximations.

$$D_{KL} \left(q^*(\mathbf{Z}, \theta | \mathbf{X}) || q^*(\mathbf{Z} | \mathbf{X}) q^*(\theta | \mathbf{X}) \right) \quad (5)$$

$$= \int_{\theta} \int_{\mathbf{Z}} q^*(\mathbf{Z}, \theta | \mathbf{X}) \log \frac{q^*(\mathbf{Z}, \theta | \mathbf{X})}{q^*(\mathbf{Z} | \mathbf{X}) q^*(\theta | \mathbf{X})} \quad (6)$$

$$= \mathcal{I}(\mathbf{Z}, \theta) \quad (7)$$

where $\mathcal{I}(\mathbf{Z}, \theta)$ is the mutual information between the variables \mathbf{Z} and θ .

3 Related Work

4 Experimental Results

Objective maximized during training:

$$E_{q(\theta, z|x)} \left[\log \left(\frac{p(x, z)p(\theta)}{q(z|x, \theta)q(\theta)} \right) \right] \quad (8)$$

Objective evaluated on the test set:

$$E_{q(\theta)} \left[E_{q(z'|x', \theta)} \left[\log \left(\frac{p(x', z')p(\theta)}{q(z'|x', \theta)q(\theta)} \right) \right] \right] \quad (9)$$

4.1 Choice of Generator-Aware model

4.2 Generator-Aware with non-factorized inference and generator networks

		q(W)	
		BNN	MNF
q(z)	factorized	167, 168	162, 194
	NF	166, 168	161, 185

Table 1: NLL on test set. Left is no GA, right is naive GA.

5 Conclusion

We’ve introduced GABVAEs.

6 Appendix

Objective maximized during training:

$$p(x) = \int_{\theta} \int_z p(x, z, \theta) \quad (10)$$

$$= E_{q(\theta, z|x)} \left[\frac{p(x, z, \theta)}{q(\theta, z|x)} \right] \quad (11)$$

$$= E_{q(\theta, z|x)} \left[\frac{p(x, z)p(\theta)}{q(z|x, \theta)q(\theta)} \right] \quad (12)$$

$$\log(p(x)) = \log \left(E_{q(\theta, z|x)} \left[\frac{p(x, z)p(\theta)}{q(z|x, \theta)q(\theta)} \right] \right) \quad (13)$$

$$\geq E_{q(\theta, z|x)} \left[\log \left(\frac{p(x, z)p(\theta)}{q(z|x, \theta)q(\theta)} \right) \right] \quad (14)$$

Objective evaluated on the test set:

$$p(x'|x) = \int_{\theta} \int_{z'} p(x', z', \theta|x) \quad (15)$$

$$= \int_{\theta} \int_z p(x', z'|\theta, x)p(\theta|x) \quad (16)$$

$$= \int_{\theta} \int_z p(x', z'|\theta)p(\theta|x) \quad (17)$$

$$q(x'|x) = \int_{\theta} \int_z p(x', z'|\theta)q(\theta) \quad (18)$$

$$= E_{q(\theta)} \left[\int_z p(x', z'|\theta) \right] \quad (19)$$

$$= E_{q(\theta)} \left[E_{q(z'|x')} \left[\frac{p(x', z'|\theta)}{q(z'|x', \theta)} \right] \right] \quad (20)$$

$$\log(q(x'|x)) = \log \left(E_{q(\theta)} \left[E_{q(z'|x')} \left[\frac{p(x', z'|\theta)}{q(z'|x', \theta)} \right] \right] \right) \quad (21)$$

$$\geq E_{q(\theta)} \left[E_{q(z'|x')} \left[\log \left(\frac{p(x', z'|\theta)}{q(z'|x', \theta)} \right) \right] \right] \quad (22)$$