Variational Laplace Autoencoders

Yookoon Park, Chris Dongjoo Kim and Gunhee Kim
Vision and Learning Lab
Seoul National University, South Korea
Introduction

- Variational Autoencoders
- Two Challenges of Amortized Variational Inference
- Contributions
Variational Autoencoders (VAEs)

• Generative network $\theta$

$$p_{\theta}(x|z) = \mathcal{N}(g_{\theta}(z), \sigma^2 I), p(z) = \mathcal{N}(0, I)$$

• Inference network $\phi$: amortized inference of $p_{\theta}(z|x)$

$$q_{\phi}(z|x) = \mathcal{N}(\mu_{\phi}(x), \text{diag}(\sigma^2_{\phi}(x)))$$

• Networks jointly trained by maximizing the Evidence Lower Bound (ELBO)

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

$$\leq \log p_{\theta}(x)$$

Two Challenges of Amortized Variational Inference

1. Enhancing the **expressiveness** of $q_\phi(z|x)$
   - The full-factorized assumption is restrictive to capture complex posteriors
   - E.g. normalizing flows (Rezende & Mohamed, 2015; Kingma et al., 2016)

2. Reducing the **amortization error** of $q_\phi(z|x)$
   - The error due to the inaccuracy of the inference network
   - E.g. gradient-based refinements of $q_\phi(z|x)$ (Kim et al, 2018; Marino et al., 2018; Krishnan et al. 2018)

Contributions

• The *Laplace approximation* of the posterior to improve the training of latent deep generative models with:
  1. Enhanced *expressiveness* of full-covariance Gaussian posterior
  2. Reduced *amortization error* due to direct covariance computation from the generative network behavior

• A novel posterior inference exploiting local linearity of ReLU networks
Approach

- Posterior Inference using Local Linear Approximations
- Generalization: Variational Laplace Autoencoders
Observation 1: Probabilistic PCA

- A linear Gaussian model
  (Tipping & Bishop, 1999)

\[
p(z) = \mathcal{N}(0, I)
\]
\[
p_\theta(x|z) = \mathcal{N}(Wz + b, \sigma^2 I)
\]

- The posterior distribution is \textit{exactly}

\[
p_\theta(z|x) = \mathcal{N}\left(\frac{1}{\sigma^2} \Sigma W^T(x - b), \Sigma\right)
\]
where \(\Sigma = \left(\frac{1}{\sigma^2} W^T W + I\right)^{-1}\)

Observation 2: Piece-wise Linear ReLU Networks

- ReLU networks are piece-wise linear (Pascanu et al., 2014; Montufar et al., 2014)
  \[ g_\theta(z) \approx W_z z + b_z \]
- Locally equivalent to probabilistic PCA
  \[ p_\theta(x|z) \approx \mathcal{N}(W_z z + b_z, \sigma^2 I) \]

Toy example. 1-dim ReLU VAE on 2-dim data

---

Posterior Inference using Local Linear Approximations

Linear models give exact posterior distribution

ReLU networks are locally linear

Observation 1

Observation 2

Posterior approximation based on the local linearity
Posterior Inference using Local Linear Approximations

1. Iteratively find the posterior mode $\mu$ where the density is concentrated
   - Solve under the linear assumption $g_\theta(\mu_t) \approx W_t \mu_t + b_t$
     $$
     \mu_{t+1} = \frac{1}{\sigma^2} \left( \frac{1}{\sigma^2} W_t^T W_t + I \right)^{-1} W_t^T (x - b)
     $$
   - Repeat for $T$ steps

2. Posterior approximation using $p_\theta(x|z) \approx \mathcal{N}(W_\mu z + b_\mu, \sigma^2 I)$
   $$
   q(z|x) = \mathcal{N}(\mu, \Sigma), \text{ where } \Sigma = \left( \frac{1}{\sigma^2} W_\mu^T W_\mu + I \right)^{-1}
   $$
Generalization: Variational Laplace Autoencoders

1. Find the posterior mode s.t. $\nabla_z \log p(x, z)|_{z=\mu} = 0$
   - Initialize $\mu_0$ using the inference network
   - Iteratively refine $\mu_t$ (e.g. use gradient-descent)

2. The **Laplace approximation** defines the posterior as:
   $$q(z|x) = \mathcal{N}(\mu, \Sigma), \text{ where } \Sigma^{-1} = \Lambda = -\nabla_z^2 \log p(x, z)|_{z=\mu}$$

3. Evaluate the ELBO using $q(z|x)$ and train the model
Results

- Posterior Covariance
- Log-likelihood Results
Experiments

- Image datasets: MNIST, OMNIGLOT, Fashion MNIST, SVHN, CIFAR10

- Baselines
  - VAE
  - Semi-Amortized (SA) VAE (Kim et al, 2018)
  - VAE + Householder Flows (HF) (Tomczak & Welling, 2016)
  - Variational Laplace Autoencoder (VLAE)

- $T=1, 2, 4, 8$ (number of iterative updates or flows)
Posterior Covariance Matrices

(a) VAE

(b) SA-VAE

(c) VAE + HF

(d) VLAE
Log-likelihood Results on CIFAR10
Thank you

Visit our poster session at Pacific Ballroom #2
Code available at: https://github.com/yookoon/VLAE