Ladder Variational Autoencoders

Maria Kovaleva

MIPT, 2022

November 27, 2022

Classic VAE

Main idea

VAEs simultaneously train a generative model $p_\theta(x, z) = p_\theta(x|z)p_\theta(z)$ for data x using latent variables z, and an inference model $q_\phi(z|x)$ by optimizing a variational lower bound to the likelihood $p_\theta(x) = \int p_\theta(x, z)dz$.

Lower bounds

1. Classic ELBO: $\log p(x) \geq E_{q_\phi(z|x)}\left[\log \frac{p_\theta(x,z)}{q_\phi(z|x)}\right] = \mathcal{L}(\theta, \phi; x) = -KL(q_\phi(z|x)||p_\theta(z)) + E_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right]$
2. IWAE: $\log p(x) \geq E_{q_\phi(z^{(1)}|x)} \dots E_{q_\phi(z^{(K)}|x)}\left[\log \sum_{k=1}^{K} \frac{p_\theta(x,z^{(k)})}{q_\phi(z^{(k)}|x)}\right] \geq \mathcal{L}_K(\theta, \phi; x)$

VAE generative part

In the generative model p_θ, the latent variables z are split into L layers z_i, $i = 1 \dots L$ as follows:

$$p_\theta(z) = p_\theta(z_L) \prod_{i=1}^{L-1} p_\theta(z_i | z_{i+1}) \tag{1}$$

$$p_\theta(z_i | z_{i+1}) = \mathcal{N}\left(z_i | \mu_{p,i}(z_{i+1}), \sigma^2_{p,i}(z_{i+1})\right), \quad p_\theta(z_L) = \mathcal{N}\left(z_L | 0, I\right) \tag{2}$$

$$p_\theta(x | z_1) = \mathcal{N}\left(x | \mu_{p,0}(z_1), \sigma^2_{p,0}(z_1)\right) \text{ or } P_\theta(x | z_1) = \mathcal{B}\left(x | \mu_{p,0}(z_1)\right) \tag{3}$$

Note:

This part will be the same for LVAE

VAE inference part

VAE inference models are parameterized as a bottom-up process. Each stochastic layer is a fully factorized gaussian distribution:

$$q_\phi(z|x) = q_\phi(z_1|x) \prod_{i=2}^{L} q_\phi(z_i|z_{i-1})$$

$$q_\phi(z_1|x) = \mathcal{N}\left(z_1|\mu_{q,1}(x), \sigma^2_{q,1}(x)\right)$$

$$q_\phi(z_i|z_{i-1}) = \mathcal{N}\left(z_i|\mu_{q,i}(z_{i-1}), \sigma^2_{q,i}(z_{i-1})\right), \; i = 2 \dots L.$$

The functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ in the generative and VAE inference models are implemented as:

$$d(y) = \texttt{MLP}(y)$$

$$\mu(y) = \texttt{Linear}(d(y))$$

$$\sigma^2(y) = \texttt{Softplus}(\texttt{Linear}(d(y)))$$

Idea of Ladder VAE

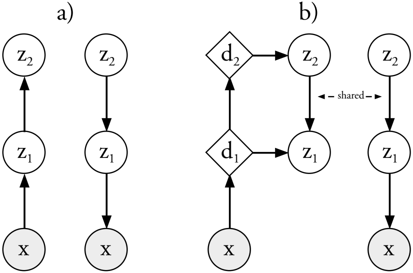

Figure: Inference (or encoder/recognition) and generative (or decoder) models for a) VAE and b) LVAE. Circles are stochastic variables and diamonds are deterministic variables.

LVAE inference part

Upward pass

computes the approximate likelihood contributions (where $d_0 = x$)

$$d_n = \mathtt{MLP}(d_{n-1})$$

$$\hat{\mu}_{q,i} = \mathtt{Linear}(d_i), \hat{\sigma}^2_{q,i} = \mathtt{Softplus}(\mathtt{Linear}(d_i)), i = 1 \dots L$$

Downward pass

recursively computing both the approximate posterior and generative distributions (where $\mu_{q,L} = \hat{\mu}_{q,L}$ and $\sigma^2_{q,L} = \hat{\sigma}^2_{q,L}$)

$$\sigma_{q,i} = \frac{1}{\hat{\sigma}^{-2}_{q,i} + \sigma^{-2}_{p,i}}; \mu_{q,i} = \frac{\hat{\mu}_{q,i}\hat{\sigma}^{-2}_{q,i} + \mu_{p,i}\sigma^{-2}_{p,i}}{\hat{\sigma}^{-2}_{q,i} + \sigma^{-2}_{p,i}}$$

$$q_\phi(z|x) = q_\phi(z_L|x) \prod_{i=1}^{L-1} q_\phi(z_i|z_{i+1}); q_\phi(z_i|\cdot) = \mathcal{N}\left(z_i|\mu_{q,i}, \sigma^2_{q,i}\right),$$

Meaning of LVAE inference part

The inference model is a precision-weighted combination of $\hat{\mu}_q$ and $\hat{\sigma}_q^2$ carrying bottom-up information and μ_p and σ_p^2 from the generative distribution carrying *top-down* prior information.

Together these form the approximate posterior distribution $q_\theta(z|z,x)$ using the same top-down dependency structure both in the inference and generative model.

A line of motivation is that a purely bottom-up inference process as in i.e. VAEs does not correspond well with real perception, where iterative interaction between bottom-up and top-down signals produces the final activity of a unit

Better performance for chains of latent variables

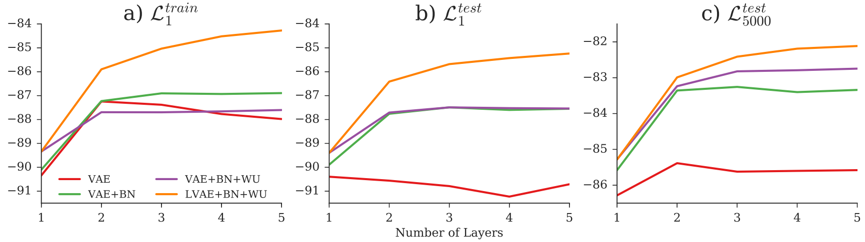

Figure: MNIST log-likelihood values for VAEs and the LVAE model with different number of latent layers, Batch normalization (*BN*) and Warm-up (*WU*). a) Train log-likelihood, b) test log-likelihood and c) test log-likelihood with 5000 importance samples.

Latent representations

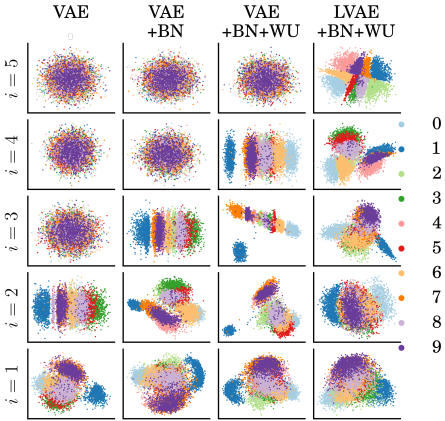

Figure: PCA-plots of samples from $q(z_i|z_{i-1})$ for 5-layer VAE and LVAE models trained on MNIST. Color-coded according to true class label

Literature

[1] Casper Kaae Sønderby et al. *Ladder Variational Autoencoders*. 2016. DOI: 10.48550/ARXIV.1602.02282. URL: https://arxiv.org/abs/1602.02282.