Deep Generative Models

Lecture 5

Roman Isachenko

Moscow Institute of Physics and Technology
Yandex School of Data Analysis

2025, Autumn

Recap of previous lecture

EM-algorithm

- E-step

$$q^*(\mathbf{z}) = \arg\max_q \mathcal{L}_{q,\theta^*}(\mathbf{x}) = \arg\min_q KL(q(\mathbf{z})||p(\mathbf{z}|\mathbf{x},\theta^*));$$

- M-step

$$\theta^* = \arg\max_\theta \mathcal{L}_{q^*,\theta}(\mathbf{x});$$

Amortized Variational Inference

Restrict the family of all possible distributions $q(\mathbf{z})$ to a parametric class $q(\mathbf{z}|\mathbf{x},\phi)$, conditioned on the samples $\mathbf{x}$ with parameters ϕ.

Variational Bayes

- E-step

$$\phi_k = \phi_{k-1} + \eta \cdot \nabla_\phi \mathcal{L}_{\phi,\theta_{k-1}}(\mathbf{x})\big|_{\phi=\phi_{k-1}}$$

- M-step

$$\theta_k = \theta_{k-1} + \eta \cdot \nabla_\theta \mathcal{L}_{\phi_k,\theta}(\mathbf{x})\big|_{\theta=\theta_{k-1}}$$

Recap of previous lecture

$$\mathcal{L}_{\phi,\theta}(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z}|\mathbf{x},\phi)} \log p(\mathbf{x}|\mathbf{z},\theta) - KL(q(\mathbf{z}|\mathbf{x},\phi)||p(\mathbf{z})) \to \max_{\phi,\theta}.$$

M-step: $\nabla_\theta \mathcal{L}_{\phi,\theta}(\mathbf{x})$, Monte Carlo Estimation

$$\nabla_\theta \mathcal{L}_{\phi,\theta}(\mathbf{x}) = \int q(\mathbf{z}|\mathbf{x},\phi) \nabla_\theta \log p(\mathbf{x}|\mathbf{z},\theta) d\mathbf{z} \approx$$

$$\approx \nabla_\theta \log p(\mathbf{x}|\mathbf{z}^*,\theta), \quad \mathbf{z}^* \sim q(\mathbf{z}|\mathbf{x},\phi).$$

E-step: $\nabla_\phi \mathcal{L}_{\phi,\theta}(\mathbf{x})$, Reparameterization Trick

$$\nabla_\phi \mathcal{L}_{\phi,\theta}(\mathbf{x}) = \int p(\epsilon) \nabla_\phi \log p(\mathbf{x}|\mathbf{g}_\phi(\mathbf{x},\epsilon),\theta) d\epsilon - \nabla_\phi \mathrm{KL}$$

$$\approx \nabla_\phi \log p(\mathbf{x}|\mathbf{g}_\phi(\mathbf{x},\epsilon^*),\theta) - \nabla_\phi \mathrm{KL}$$

Variational Assumption

$$p(\epsilon) = \mathcal{N}(0,\mathbf{I}); \quad q(\mathbf{z}|\mathbf{x},\phi) = \mathcal{N}(\boldsymbol{\mu}_\phi(\mathbf{x}), \boldsymbol{\sigma}^2_\phi(\mathbf{x})).$$

$$\mathbf{z} = \mathbf{g}_\phi(\mathbf{x},\epsilon) = \boldsymbol{\sigma}_\phi(\mathbf{x}) \odot \epsilon + \boldsymbol{\mu}_\phi(\mathbf{x}).$$

Recap of previous lecture

Training (EM-algorithm)

- Select a random sample $\mathbf{x}_i$, $i \sim \text{Uniform}\{1, n\}$ (or batch).
- Compute the objective (using the reparameterization trick):

$$\epsilon^* \sim p(\epsilon); \quad \mathbf{z}^* = \mathbf{g}_\phi(\mathbf{x}, \epsilon^*);$$

$$\mathcal{L}_{\phi,\theta}(\mathbf{x}) \approx \log p(\mathbf{x}|\mathbf{z}^*, \theta) - KL(q(\mathbf{z}^*|\mathbf{x}, \phi)||p(\mathbf{z}^*)).$$

- Update parameters via a gradient step using stochastic gradients w.r.t. ϕ and θ through autograd.

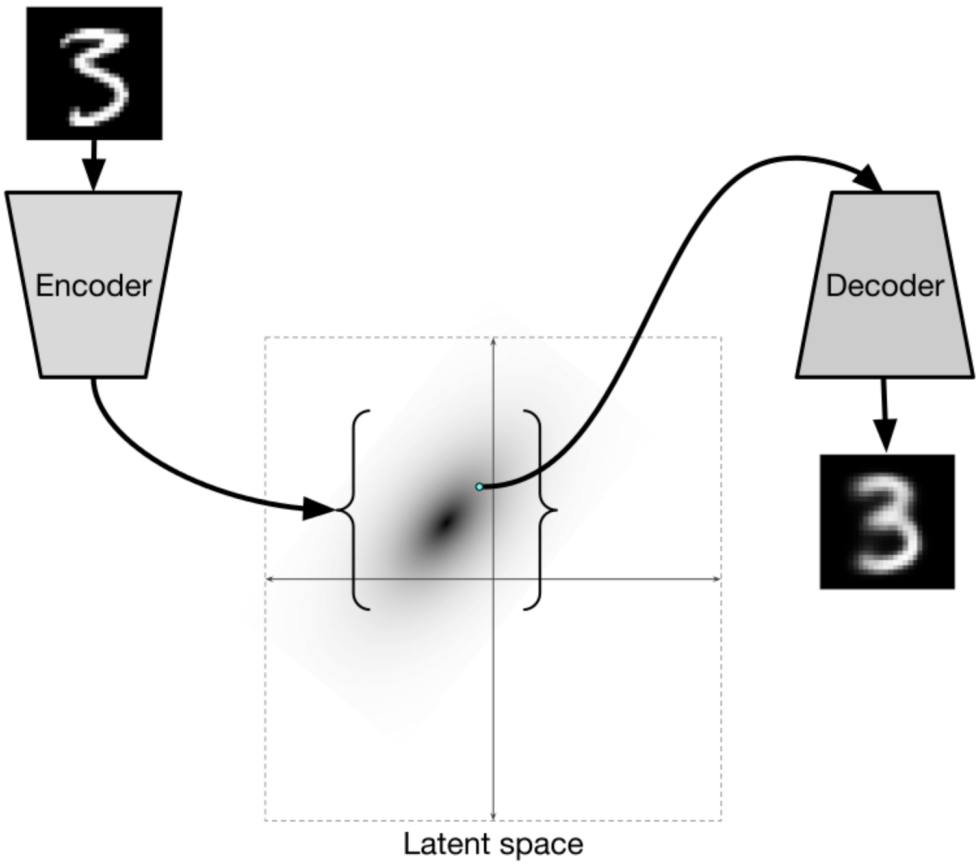
Inference

- Sample $\mathbf{z}^*$ from the prior distribution $p(\mathbf{z})$ ($\mathcal{N}(0, \mathbf{I})$);
- Sample from the decoder $p(\mathbf{x}|\mathbf{z}^*, \theta)$.

Note: The encoder $q(\mathbf{z}|\mathbf{x}, \phi)$ is not needed during generation.

Recap of previous lecture

$$\mathcal{L}_{q,\theta}(\mathbf{x}) = \mathbb{E}_q \log p(\mathbf{x}|\mathbf{z},\theta) - KL(q(\mathbf{z}|\mathbf{x},\phi)||p(\mathbf{z}))$$



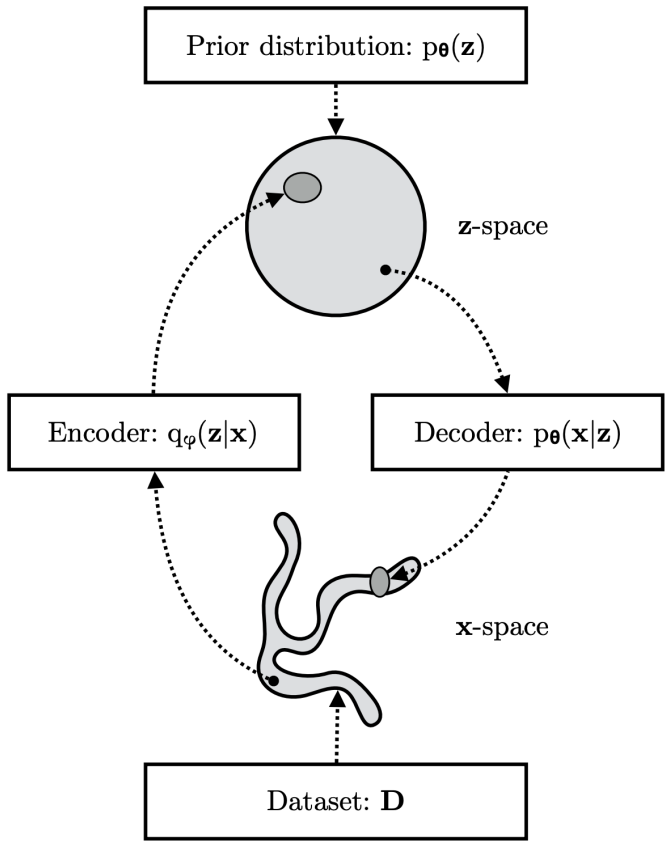


image credit: http://ijdykeman.github.io/ml/2016/12/21/cvae.html
Kingma D. P., Welling M. An introduction to variational autoencoders, 2019

Recap of previous lecture

	VAE	**NF**
Objective	ELBO $\mathcal{L}$	Forward KL/MLE
Encoder	Stochastic $\mathbf{z} \sim q(\mathbf{z}\|\mathbf{x}, \boldsymbol{\phi})$	Deterministic $\mathbf{z} = \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x})$ $q(\mathbf{z}\|\mathbf{x}, \boldsymbol{\theta}) = \delta(\mathbf{z} - \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}))$
Decoder	Stochastic $\mathbf{x} \sim p(\mathbf{x}\|\mathbf{z}, \boldsymbol{\theta})$	Deterministic $\mathbf{x} = \mathbf{g}_{\boldsymbol{\theta}}(\mathbf{z})$ $p(\mathbf{x}\|\mathbf{z}, \boldsymbol{\theta}) = \delta(\mathbf{x} - \mathbf{g}_{\boldsymbol{\theta}}(\mathbf{z}))$
Parameters	$\boldsymbol{\phi}, \boldsymbol{\theta}$	$\boldsymbol{\theta} \equiv \boldsymbol{\phi}$

Theorem

MLE for normalizing flow is equivalent to maximizing the ELBO for a VAE model with deterministic encoder and decoder:

$$p(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}) = \delta(\mathbf{x} - \mathbf{f}_{\boldsymbol{\theta}}^{-1}(\mathbf{z})) = \delta(\mathbf{x} - \mathbf{g}_{\boldsymbol{\theta}}(\mathbf{z}));$$

$$q(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta}) = \delta(\mathbf{z} - \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x})).$$

Nielsen D., et al. SurVAE Flows: Surjections to Bridge the Gap between VAEs and Flows, 2020

Recap of previous lecture

Assumptions

- Let $c \sim \text{Categorical}(\boldsymbol{\pi})$, where
$$\boldsymbol{\pi} = (\pi_1, \dots, \pi_K), \quad \pi_k = P(c = k), \quad \sum_{k=1}^{K} \pi_k = 1.$$
- Suppose the VAE model has a discrete latent variable c with prior $p(c) = \text{Uniform}\{1, \dots, K\}$.

ELBO

$$\mathcal{L}_{\phi,\boldsymbol{\theta}}(\mathbf{x}) = \mathbb{E}_{q(c|\mathbf{x},\phi)} \log p(\mathbf{x}|c, \boldsymbol{\theta}) - KL(q(c|\mathbf{x}, \phi)||p(c)) \to \max_{\phi,\boldsymbol{\theta}}.$$

$$KL(q(c|\mathbf{x}, \phi)||p(c)) = -H(q(c|\mathbf{x}, \phi)) + \log K.$$

- Our encoder should output the discrete distribution $q(c|\mathbf{x}, \phi)$.
- We need an analogue of the reparameterization trick for the discrete distribution $q(c|\mathbf{x}, \phi)$.
- Our decoder $p(\mathbf{x}|c, \boldsymbol{\theta})$ should take the discrete random variable c as input.

Oord A., Vinyals O., Kavukcuoglu K. Neural Discrete Representation Learning, 2017

Outline

1. Vector Quantization: (Discrete VAE Latent Representations)
2. ELBO Surgery
3. Learnable VAE Prior

Outline

1. Vector Quantization: (Discrete VAE Latent Representations)
2. ELBO Surgery
3. Learnable VAE Prior

Vector Quantization

Define the dictionary space $\{\mathbf{e}_k\}_{k=1}^{K}$, where $\mathbf{e}_k \in \mathbb{R}^L$, and K is the dictionary size.

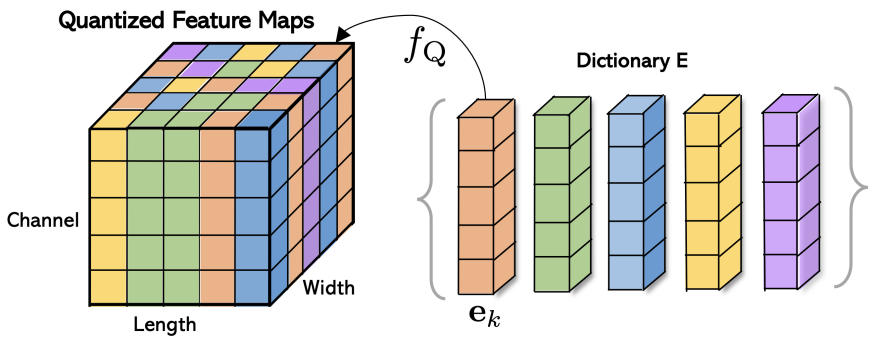
Quantized Representation

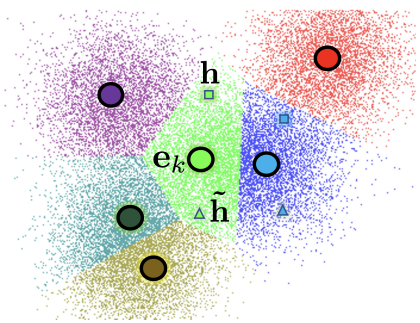
$\mathbf{z}_q \in \mathbb{R}^L$ for $\mathbf{z} \in \mathbb{R}^L$ is defined by a nearest-neighbor look-up using the dictionary space:

$$\mathbf{z}_q = \mathbf{q}(\mathbf{z}) = \mathbf{e}_{k^*}, \quad \text{where } k^* = \arg\min_k \|\mathbf{z} - \mathbf{e}_k\|.$$

Quantization Procedure

If we have a tensor with spatial dimensions, we apply quantization to each of the $W \times H$ spatial locations.





Zhao Y. et al. Feature Quantization Improves GAN Training, 2020

Vector Quantized VAE (VQ-VAE)

- The encoder outputs a continuous representation $\mathbf{z}_e = \mathrm{NN}_{e,\phi}(\mathbf{x}) \in \mathbb{R}^L$.
- Quantization deterministically maps the encoder output $\mathbf{z}_e$ to its quantized representation $\mathbf{z}_q$.
- Use the dictionary elements $\mathbf{e}_c$ in the decoder distribution $p(\mathbf{x}|\mathbf{e}_c, \boldsymbol{\theta})$ (in place of $p(\mathbf{x}|c, \boldsymbol{\theta})$).

Deterministic Variational Posterior

$$q(c = k^*|\mathbf{x}, \boldsymbol{\phi}) = \begin{cases} 1, & \text{for } k^* = \arg\min_k \|\mathbf{z}_e - \mathbf{e}_k\|; \\ 0, & \text{otherwise.} \end{cases}$$

$$KL(q(c|\mathbf{x}, \boldsymbol{\phi})||p(c)) = -\underbrace{H(q(c|\mathbf{x}, \boldsymbol{\phi}))}_{=0} + \log K = \log K.$$

Note: The KL regularizer does not affect the ELBO objective in this case.

Oord A., Vinyals O., Kavukcuoglu K. Neural Discrete Representation Learning, 2017

Vector Quantized VAE (VQ-VAE): Forward

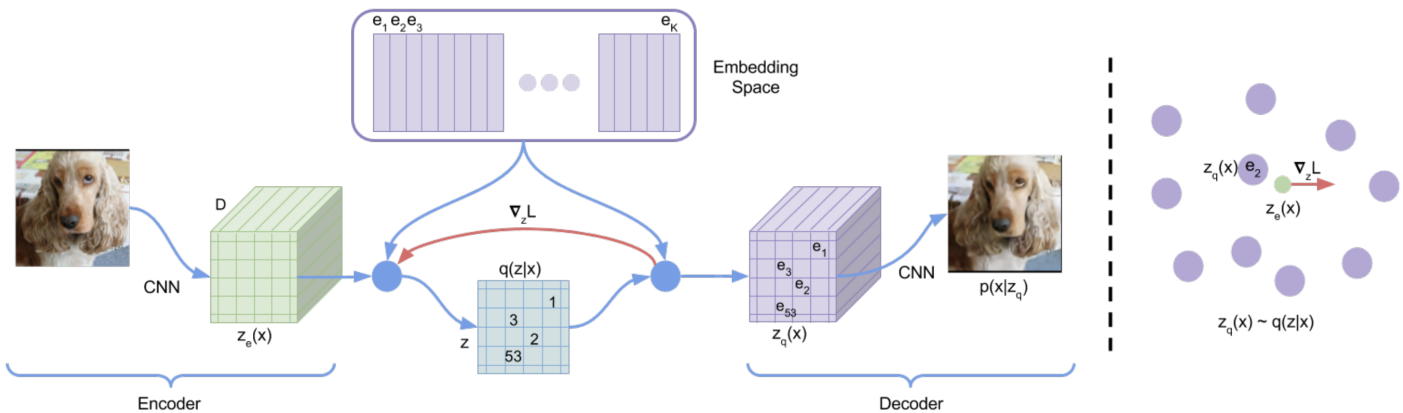
Deterministic Variational Posterior

$$q(c = k^*|\mathbf{x}, \phi) = \begin{cases} 1, & \text{for } k^* = \arg\min_k \|\mathbf{z}_e - \mathbf{e}_k\|; \\ 0, & \text{otherwise.} \end{cases}$$

ELBO

$$\mathcal{L}_{\phi,\theta}(\mathbf{x}) = \mathbb{E}_{q(c|\mathbf{x},\phi)} \log p(\mathbf{x}|\mathbf{e}_c, \theta) - \log K = \log p(\mathbf{x}|\mathbf{z}_q, \theta) - \log K,$$

where $\mathbf{z}_q = \mathbf{e}_{k^*}$, $k^* = \arg\min_k \|\mathbf{z}_e - \mathbf{e}_k\|$.



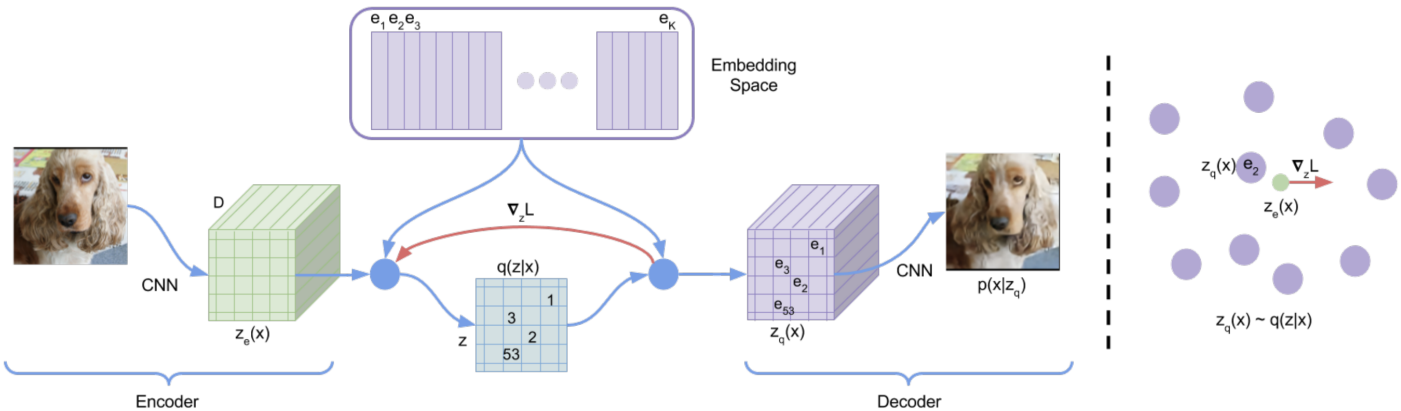
Problem: The arg min operation is not differentiable.

Oord A., Vinyals O., Kavukcuoglu K. Neural Discrete Representation Learning, 2017

Vector Quantized VAE (VQ-VAE): Backward

ELBO

$$\mathcal{L}_{\phi,\theta}(\mathbf{x}) = \log p(\mathbf{x}|\mathbf{z}_q, \theta) - \log K, \quad \mathbf{z}_q = \mathbf{e}_{k^*}, k^* = \arg\min_k \|\mathbf{z}_e - \mathbf{e}_k\|.$$



Straight-Through Gradient Estimation

$$\frac{\partial \log p(\mathbf{x}|\mathbf{z}_q, \theta)}{\partial \phi} = \frac{\partial \log p(\mathbf{x}|\mathbf{z}_q, \theta)}{\partial \mathbf{z}_q} \cdot \frac{\partial \mathbf{z}_q}{\partial \phi} =$$

$$= \frac{\partial \log p(\mathbf{x}|\mathbf{z}_q, \theta)}{\partial \mathbf{z}_q} \cdot \frac{\partial \mathbf{z}_q}{\partial \mathbf{z}_e} \cdot \frac{\partial \mathbf{z}_e}{\partial \phi} \approx \frac{\partial \log p(\mathbf{x}|\mathbf{z}_q, \theta)}{\partial \mathbf{z}_q} \cdot \frac{\partial \mathbf{z}_e}{\partial \phi}$$

Oord A., Vinyals O., Kavukcuoglu K. Neural Discrete Representation Learning, 2017

Vector Quantized VAE-2 (VQ-VAE-2)

Generalization to the spatial dimension: $\mathbf{c} \in \{1, \ldots, K\}^{W \times H}$

$$q(\mathbf{c}|\mathbf{x}, \phi) = \prod_{i=1}^{W} \prod_{j=1}^{H} q(c_{ij}|\mathbf{x}, \phi); \quad p(\mathbf{c}) = \prod_{i=1}^{W} \prod_{j=1}^{H} \text{Uniform}\{1, \ldots, K\}.$$

Sample Diversity

VQ-VAE (Proposed) **BigGAN deep**

Razavi A., Oord A., Vinyals O. Generating Diverse High-Fidelity Images with VQ-VAE-2, 2019

Outline

1. Vector Quantization: (Discrete VAE Latent Representations)
2. ELBO Surgery
3. Learnable VAE Prior

ELBO Surgery

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_{\phi,\theta}(\mathbf{x}_i) = \frac{1}{n}\sum_{i=1}^{n}\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_i,\phi)}\log p(\mathbf{x}_i|\mathbf{z},\theta) - KL(q(\mathbf{z}|\mathbf{x}_i,\phi)||p(\mathbf{z}))\Big].$$

Theorem

$$\frac{1}{n}\sum_{i=1}^{n}KL(q(\mathbf{z}|\mathbf{x}_i,\phi)||p(\mathbf{z})) = KL(q_{\text{agg}}(\mathbf{z}|\phi)||p(\mathbf{z})) + \mathbb{I}_q[\mathbf{x},\mathbf{z}];$$

- $q_{\text{agg}}(\mathbf{z}|\phi) = \frac{1}{n}\sum_{i=1}^{n} q(\mathbf{z}|\mathbf{x}_i,\phi)$ is the **aggregated** variational posterior distribution.
- $\mathbb{I}_q[\mathbf{x},\mathbf{z}]$ is the mutual information between $\mathbf{x}$ and $\mathbf{z}$ under the data distribution $\pi(\mathbf{x})$ and the distribution $q(\mathbf{z}|\mathbf{x},\phi)$.
- The first term pushes $q_{\text{agg}}(\mathbf{z}|\phi)$ towards the prior $p(\mathbf{z})$.
- The second term reduces the amount of information about $\mathbf{x}$ encoded in $\mathbf{z}$.

Hoffman M. D., Johnson M. J. ELBO surgery: yet another way to carve up the variational evidence lower bound, 2016

ELBO Surgery

$$\frac{1}{n}\sum_{i=1}^{n} KL(q(\mathbf{z}|\mathbf{x}_i,\phi)||p(\mathbf{z})) = KL(q_{\text{agg}}(\mathbf{z}|\phi)||p(\mathbf{z})) + \mathbb{I}_q[\mathbf{x},\mathbf{z}].$$

Proof

$$\frac{1}{n}\sum_{i=1}^{n} KL(q(\mathbf{z}|\mathbf{x}_i,\phi)||p(\mathbf{z})) = \frac{1}{n}\sum_{i=1}^{n}\int q(\mathbf{z}|\mathbf{x}_i,\phi)\log\frac{q(\mathbf{z}|\mathbf{x}_i,\phi)}{p(\mathbf{z})}d\mathbf{z} =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int q(\mathbf{z}|\mathbf{x}_i,\phi)\log\frac{q_{\text{agg}}(\mathbf{z}|\phi)q(\mathbf{z}|\mathbf{x}_i,\phi)}{p(\mathbf{z})q_{\text{agg}}(\mathbf{z}|\phi)}d\mathbf{z} =$$

$$= \int \frac{1}{n}\sum_{i=1}^{n} q(\mathbf{z}|\mathbf{x}_i,\phi)\log\frac{q_{\text{agg}}(\mathbf{z}|\phi)}{p(\mathbf{z})}d\mathbf{z} + \frac{1}{n}\sum_{i=1}^{n}\int q(\mathbf{z}|\mathbf{x}_i,\phi)\log\frac{q(\mathbf{z}|\mathbf{x}_i,\phi)}{q_{\text{agg}}(\mathbf{z}|\phi)}d\mathbf{z} =$$

$$= KL(q_{\text{agg}}(\mathbf{z}|\phi)||p(\mathbf{z})) + \frac{1}{n}\sum_{i=1}^{n} KL(q(\mathbf{z}|\mathbf{x}_i,\phi)||q_{\text{agg}}(\mathbf{z}|\phi))$$

$$\mathbb{I}_q[\mathbf{x},\mathbf{z}] = \frac{1}{n}\sum_{i=1}^{n} KL(q(\mathbf{z}|\mathbf{x}_i,\phi)||q_{\text{agg}}(\mathbf{z}|\phi)).$$

Hoffman M. D., Johnson M. J. ELBO surgery: yet another way to carve up the variational evidence lower bound, 2016

ELBO Surgery

ELBO Revisiting

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_{\phi,\theta}(\mathbf{x}_i) = \frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_i,\phi)}\log p(\mathbf{x}_i|\mathbf{z},\theta) - KL(q(\mathbf{z}|\mathbf{x}_i,\phi)||p(\mathbf{z}))\right] =$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_i,\phi)}\log p(\mathbf{x}_i|\mathbf{z},\theta)}_{\text{Reconstruction Loss}} - \underbrace{\mathbb{I}_q[\mathbf{x},\mathbf{z}]}_{\text{MI}} - \underbrace{KL(q_{\text{agg}}(\mathbf{z}|\phi)||p(\mathbf{z}))}_{\text{Marginal KL}}$$

The prior distribution $p(\mathbf{z})$ appears only in the last term.

Optimal VAE Prior

$$KL(q_{\text{agg}}(\mathbf{z}|\phi)||p(\mathbf{z})) = 0 \quad \Leftrightarrow \quad p(\mathbf{z}) = q_{\text{agg}}(\mathbf{z}|\phi) = \frac{1}{n}\sum_{i=1}^{n} q(\mathbf{z}|\mathbf{x}_i,\phi).$$

The optimal prior $p(\mathbf{z})$ is the aggregated variational posterior distribution $q_{\text{agg}}(\mathbf{z}|\phi)$!

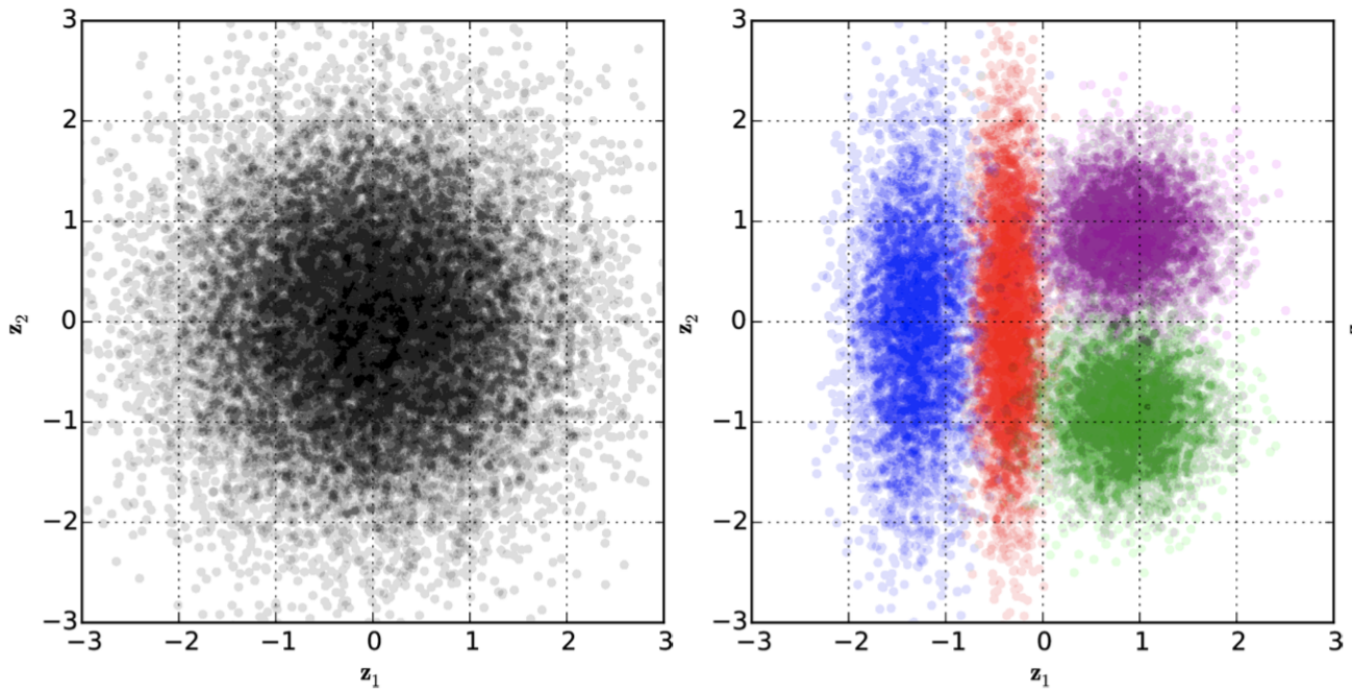
Hoffman M. D., Johnson M. J. ELBO surgery: yet another way to carve up the variational evidence lower bound, 2016

Variational Posterior

ELBO Decomposition

$$\log p(\mathbf{x}|\boldsymbol{\theta}) = \mathcal{L}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{x}) + KL(q(\mathbf{z}|\mathbf{x},\boldsymbol{\phi})||p(\mathbf{z}|\mathbf{x},\boldsymbol{\theta})).$$

- $q(\mathbf{z}|\mathbf{x},\boldsymbol{\phi}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\mathbf{x}), \boldsymbol{\sigma}^2_{\boldsymbol{\phi}}(\mathbf{x}))$ is a unimodal distribution.
- It is widely believed that the **mismatch between $p(\mathbf{z})$ and $q_{\text{agg}}(\mathbf{z}|\boldsymbol{\phi})$ is the main cause of VAE blurry images**.



(a) Prior distribution (b) Posteriors in standard VAE

Rezende D. J., Mohamed S. Variational Inference with Normalizing Flows, 2015

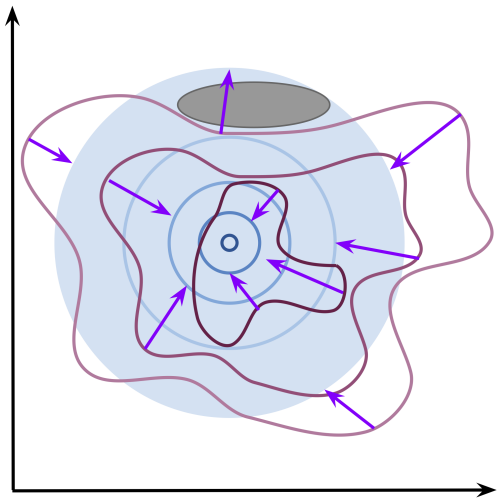
Outline

1. Vector Quantization: (Discrete VAE Latent Representations)
2. ELBO Surgery
3. Learnable VAE Prior

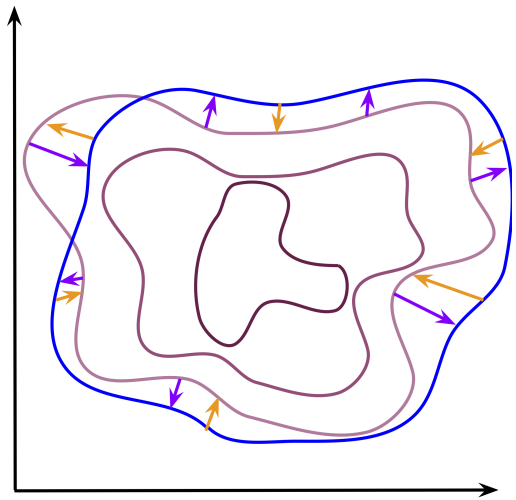
Optimal VAE Prior

- Standard Gaussian $p(\mathbf{z}) = \mathcal{N}(0, \mathbf{I}) \Rightarrow$ over-regularization;
- $p(\mathbf{z}) = q_{\text{agg}}(\mathbf{z}|\phi) = \frac{1}{n}\sum_{i=1}^{n} q(\mathbf{z}|\mathbf{x}_i, \phi) \Rightarrow$ overfitting and computationally expensive.

Non-Learnable Prior $p(\mathbf{z})$



Learnable Prior $p(\mathbf{z}|\boldsymbol{\lambda})$



ELBO Revisiting

$$\frac{1}{n}\sum_{i=1}^{n} \mathcal{L}_{\phi,\theta}(\mathbf{x}_i) = \text{RL} - \text{MI} - KL(q_{\text{agg}}(\mathbf{z}|\phi)||p(\mathbf{z}|\boldsymbol{\lambda}))$$

This is the forward KL divergence with respect to $p(\mathbf{z}|\boldsymbol{\lambda})$.

---

image credit: https://jmtomczak.github.io/blog/7/7_priors.html

NF-Based VAE Prior

NF Model in Latent Space

$$\log p(\mathbf{z}|\boldsymbol{\lambda}) = \log p(\mathbf{z}^*) + \log \left| \det \left(\frac{d\mathbf{z}^*}{d\mathbf{z}} \right) \right| = \log p(\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{z})) + \log \left| \det(\mathbf{J}_{\mathbf{f}}) \right|$$

$$\mathbf{z} = \mathbf{g}_{\boldsymbol{\lambda}}(\mathbf{z}^*) = \mathbf{f}_{\boldsymbol{\lambda}}^{-1}(\mathbf{z}^*)$$

- RealNVP with coupling layers.
- Autoregressive NF (fast $\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{z})$, slow $\mathbf{g}_{\boldsymbol{\lambda}}(\mathbf{z}^*)$).

ELBO with NF-Based VAE Prior

$$\mathcal{L}_{\phi,\theta}(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z}|\mathbf{x},\phi)} \log p(\mathbf{x}|\mathbf{z},\theta) - KL(q(\mathbf{z}|\mathbf{x},\phi)||p(\mathbf{z})) =$$
$$= \mathbb{E}_{q(\mathbf{z}|\mathbf{x},\phi)} \left[\log p(\mathbf{x}|\mathbf{z},\theta) + \log p(\mathbf{z}|\boldsymbol{\lambda}) - \log q(\mathbf{z}|\mathbf{x},\phi) \right]$$
$$= \mathbb{E}_{q(\mathbf{z}|\mathbf{x},\phi)} \Big[\log p(\mathbf{x}|\mathbf{z},\theta) + \underbrace{\left(\log p(\mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{z})) + \log \left| \det(\mathbf{J}_{\mathbf{f}}) \right| \right)}_{\text{NF-based prior}} - \log q(\mathbf{z}|\mathbf{x},\phi) \Big]$$

Chen X. et al. Variational Lossy Autoencoder, 2016

Summary

- Vector quantization is a technique for constructing a VAE with discrete latent space and deterministic variational posterior.
- The straight-through gradient estimator ignores the quantization operation during backpropagation.
- ELBO surgery provides insights into the role of the prior in VAEs. The optimal prior is the aggregated variational posterior distribution.
- It is widely believed that the mismatch between $p(\mathbf{z})$ and $q_{\text{agg}}(\mathbf{z}|\phi)$ is the primary reason for the blurry images produced by VAEs.
- Normalizing flow-based priors can be incorporated into VAEs, including autoregressive flows.