Deep Generative Models

Lecture 13

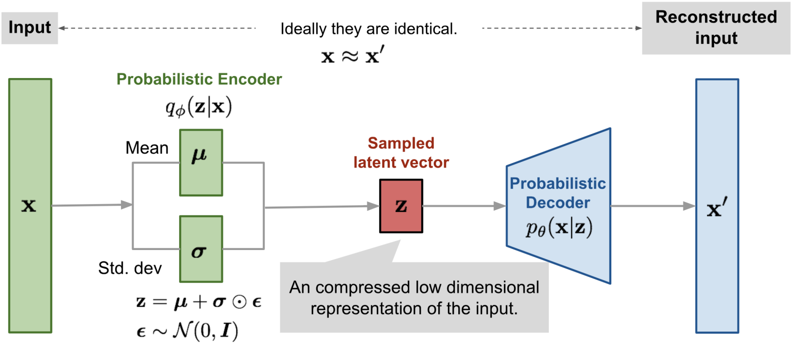
Roman Isachenko

Ozon Masters

Spring, 2022

Outline

Discrete VAE



- Previous VAE models had **continuous** latent variables $\mathbf{z}$.
- **Discrete** representations $\mathbf{z}$ are potentially a more natural fit for many of the modalities.
- Powerful autoregressive models (like PixelCNN) have been developed for modelling distributions over discrete variables.

Discrete VAE

If $\mathbf{z}$ is a discrete random variable we cannot differentiate through it.

Gumbel-Max trick

Let $G_k \sim \text{Gumbel}$ for $k = 1, \dots, K$, i.e. $G = -\log(\log u)$, $u \sim \text{Uniform}[0, 1]$. Then a discrete random variable

$$z = \arg\max_k (\log \pi_k + G_k), \quad \sum_k \pi_k = 1$$

has a categorical distribution $z \sim \text{Categorical}(\boldsymbol{\pi})$ $(P(z = k) = \pi_k)$.

Problem: We still have non-differentiable $\arg\max$ operation.

Gumbel-Softmax relaxation

$$z_k = \frac{\exp((\log \pi_k + G_k)/\tau)}{\sum_{j=1}^K \exp((\log \pi_j + G_j)/\tau)}, \quad k = 1, \dots, K.$$

Here τ is a temperature parameter.

Maddison C. J., Mnih A., Teh Y. W. The Concrete distribution: A continuous relaxation of discrete random variables, 2016
Jang E., Gu S., Poole B. Categorical reparameterization with Gumbel-Softmax, 2016

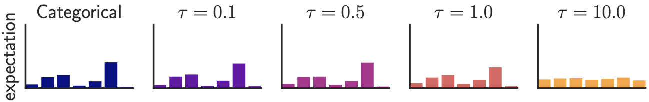
Discrete VAE

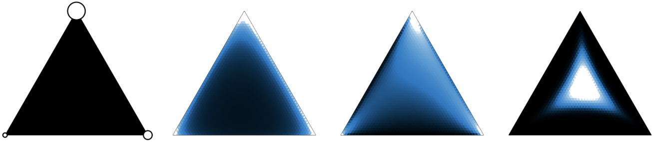
Gumbel-Softmax relaxation

Concrete distribution = continuous + discrete

$$z_k = \frac{\exp((\log \pi_k + G_k)/\tau)}{\sum_{j=1}^{K} \exp((\log \pi_j + G_j)/\tau)}, \quad k = 1, \dots, K.$$

Here τ is a temperature parameter. Now we have differentiable operation.





(a) $\lambda = 0$ (b) $\lambda = 1/2$ (c) $\lambda = 1$ (d) $\lambda = 2$

Maddison C. J., Mnih A., Teh Y. W. The Concrete distribution: A continuous relaxation of discrete random variables, 2016

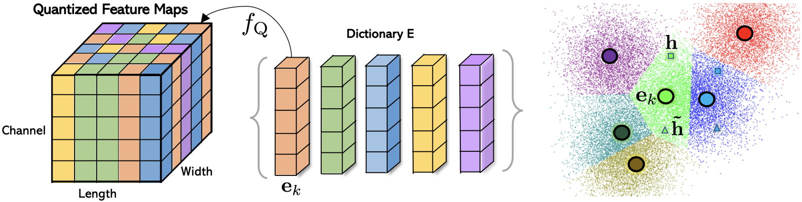
Jang E., Gu S., Poole B. Categorical reparameterization with Gumbel-Softmax, 2016

Vector Quantized VAE

- Define dictionary space $\{\mathbf{e}_k\}_{k=1}^{K}$, where $\mathbf{e}_k \in \mathbb{R}^C$, K is the size of the dictionary.
- Let $\mathbf{z} = \mathrm{NN_e}(\mathbf{x}) \in \mathbb{R}^{W \times H \times C}$ be an encoder output.
- Quantized representation $\mathbf{z}_q \in \mathbb{R}^{W \times H \times C}$ is defined by a nearest neighbour look-up using the shared dictionary space for each of $W \times H$ spatial locations

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

Quantization procedure



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

Vector Quantized VAE

Define VAE latent variable $\hat{\mathbf{z}} \in \mathbb{R}^{W \times H}$ with prior distribution $p(\hat{\mathbf{z}}) = \text{Uniform}\{1, \dots, K\}$ and variational posterior distribution

$$q(\hat{\mathbf{z}}|\mathbf{x}) = \prod_{i=1}^{W} \prod_{j=1}^{H} q(\hat{z}_{ij}|\mathbf{x})$$

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

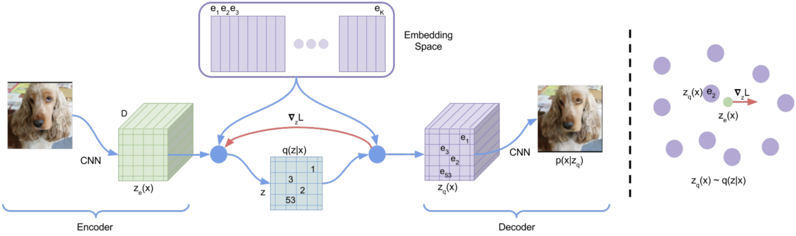
ELBO objective

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

- VAE proposal distribution $q(\hat{\mathbf{z}}|\mathbf{x})$ is deterministic.
- $KL(q(\hat{z}|\mathbf{x})||p(\hat{z}))$ term in ELBO is constant (equals to $\log K$).

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

Vector Quantized VAE



Objective

$$\log p(\mathbf{x}|\mathbf{z}_q) + \|\mathrm{sg}(\mathbf{z}_e) - \mathbf{z}_q\| + \beta\|\mathbf{z}_e - \mathrm{sg}(\mathbf{z}_q)\|$$

- First term is ELBO part.
- Quantization operation is not differentiable.
- Straight-through gradient estimation is used to backpropagate the quantization operation.

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

Vector Quantized VAE-2

Samples 1024x1024

Samples diversity

VQ-VAE (Proposed) BigGAN deep

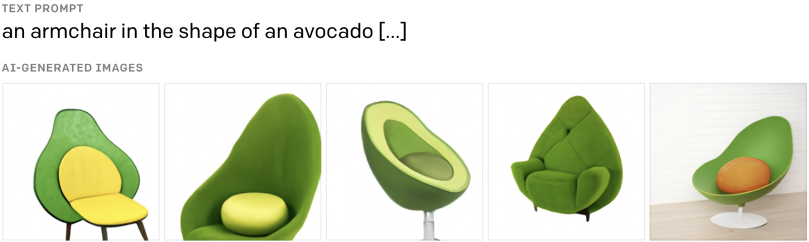
Razavi A., Oord A., Vinyals O. Generating Diverse High-Fidelity Images with VQ-VAE-2, 2019

DALL-E

Deterministic VQ-VAE posterior

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

- It is possible to use Gumbel-Softmax trick to relax this distribution to continuous one.
- Since latent space is discrete we could train autoregressive transformers in it.
- It is a natural way to incorporate text and image spaces.



Ramesh A. et al. Zero-shot text-to-image generation, 2021

Summary

- Gumbel-Softmax and Quantization are the two ways to create VAE with discrete latent space.

- It becomes more and more popular to use discrete latent spaces in the fields of image/video/music generaton.