Finite Scalar Quantization: VQ-VAE made simple

Presented by: Achint Kumar

Generative AI Reading Club

October 4, 2023

Desiderata

Finite Scalar Quantization (FSQ) vs. VQ-VAE

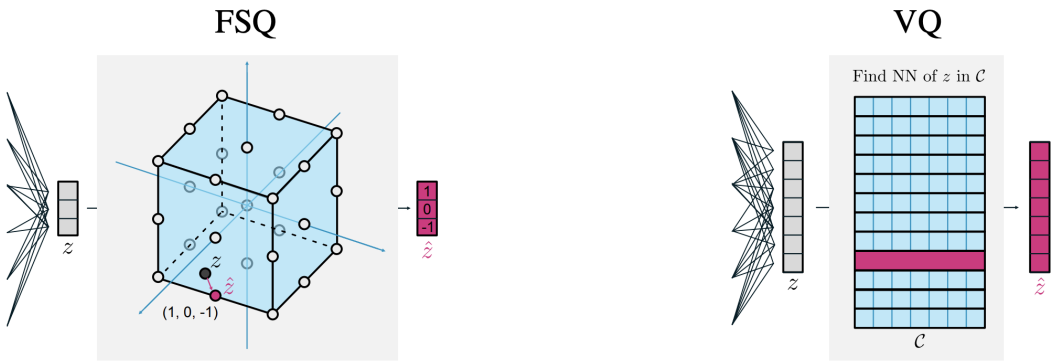

Figure 1: *FSQ (left):* the final encoder layer projects to d dimensions ($d = 3$ shown). We bound each dimension of the encoder output z to L values ($L = 3$ shown), and then round to integers, resulting in the quantized $\hat{z}$, the nearest point in this hypercube. *VQ (right)*: The final encoder layer projects to d dimensions ($d = 7$ shown, as d is typically much larger for VQ). The resulting vector z is replaced with the closest vector from the codebook, $\hat{z}$, by nearest neighbor lookup.

	VQ	FSQ
Quantization	$\arg\min_{c \in \mathcal{C}} \|\|z - c\|\|$	$\text{round}(f(z))$
Gradients	STE	STE
Aux. Losses	Commitment, codebook, entropy loss	-
Tricks	EMA on codebook, codebook splitting projections, ...	-
Parameters	Codebook	-

Finite Scalar Quantization

$$\hat{z} = \text{round}\left(f(z)\right)$$

where

$$f(z) = \lfloor z \rfloor \tanh(z)$$

To calculate gradient, we again use straight-through estimator (STE):

$$\text{round\_ste}: \; x \mapsto x + sg(\text{round}(x) - x)$$