Assignment 5

Ehsan Merrikhi 400101967
github repository

Deep Learning

Dr. Fatemizadeh

January 25, 2025

Deep Learning

Assignment 5

Ehsan Merrikhi 400101967
github repository

Contents

1. Cost Function in VQ-VAE

The cost function for VQ-VAE is shown in equation (2). Explain what each of the three terms in this cost function signifies.

$$\mathcal{L} = \underbrace{\log(p(x|z_q(x)))}_{(1)} + \underbrace{\|z_e(x) \cdot \text{detach}() - e\|}_{(2)} + \beta \underbrace{\|z_e(x) - e \cdot \text{detach}()\|}_{(3)} \tag{2}$$

Soloution

1. The first term $\log(p(x|z_q(x)))$ measures the reconstruction loss.
2. The second term $\|z_e(x) \cdot \text{detach}() - e\|$ enforces the encoder's output to stay close to the quantized representation e.
3. The third term $\beta\|z_e(x) - e \cdot \text{detach}()\|$ ensures that the quantized representation e follows the encoder's output.

Additional Explanation About the `detach()` Function

The article states that during the learning process, gradients after the codebook are directly copied to before the codebook, and there is no need to compute the derivative of the process (function) of the codebook itself. This gradient copying process can be expressed with the following equation:

$$B = A + (f(A) - A).\text{detach}() \tag{3}$$

The `detach()` method ensures that the specified variable is treated as a constant when computing derivatives, meaning it does not affect the backpropagation process. In the equation above, A is the input before the codebook, B is the output of the codebook, and the function f(.) represents the codebook process, which is also referred to as the Quantization Vector.

2. Quantization and Codebook

How are vectors quantized in this model? Briefly explain the codebook and how it is learned. Also, explain what the codebook vectors represent intuitively.

Soloution

- In this model, vectors are quantized by mapping the encoder output $z_e(x)$ to the closest vector in the codebook e_i based on a distance metric (e.g., Euclidean distance). This process replaces $z_e(x)$ with e_i from the codebook.
- Concretely, the codebook vectors are represented as a set of K learned embeddings, where each vector corresponds to a distinct quantization point. The codebook is learned during training through a combination of the encoder and decoder losses, as well as specific terms ensuring $z_e(x)$ is close to the chosen e_i.
- Codebook vectors intuitively represent discrete prototypes or cluster centers in the latent space, capturing key features or patterns from the input data for efficient encoding and reconstruction.