State-of-the-Art Loss Function for Training Diffusion Models

June 6, 2024

1 Introduction

Diffusion models have emerged as powerful generative models capable of producing high-quality samples. These models, such as Denoising Diffusion Probabilistic Models (DDPMs), are trained to reverse a diffusion process that gradually adds noise to the data. The state-of-the-art (SOTA) loss function used for training these models is typically the **variational bound loss**, also known as the **evidence lower bound (ELBO)**. This report details the components and formulation of this loss function.

2 Variational Bound Loss

The variational bound loss for diffusion models consists of several terms that capture different aspects of the model's performance. The primary focus is on the difference between the predicted noise and the actual noise in the data. The main components of this loss function are the *reconstruction loss* and the *prior matching loss.*

2.1 Reconstruction Loss

The reconstruction loss is implemented as the mean squared error (MSE) between the predicted noise $\epsilon_\theta(x_t, t)$ and the true noise ϵ. This term ensures that the model accurately predicts the noise added to the data at each time step. Formally, the reconstruction loss is defined as:

$$L_{\text{recon}} = \mathbb{E}_{t,x_0,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\right] \tag{1}$$

Here, x_t represents the noised version of the data sample x_0 at time step t, and ϵ is the actual noise added to the sample.

2.2 Prior Matching Loss

The prior matching loss ensures that the final distribution of the generated samples matches the prior distribution, typically a standard Gaussian distribution. This is crucial for the model to generate realistic samples that conform to the expected distribution.

2.3 KL Divergence Term

Another critical component is the Kullback-Leibler (KL) divergence term, which measures the difference between the variational posterior and the true posterior. This term acts as a regularizer, ensuring that the approximate posterior does not diverge significantly from the prior distribution. The KL divergence term is defined as:

$$L_{\mathrm{KL}} = D_{\mathrm{KL}}(q(x_{t-1}|x_t)\|p(x_{t-1}|x_t)) \tag{2}$$

3 Total Loss Function

The total loss for training a diffusion model is a weighted combination of the reconstruction loss and the KL divergence term. The commonly used formulation is:

$$L_{\mathrm{total}} = L_{\mathrm{recon}} + \beta L_{\mathrm{KL}} \tag{3}$$

where β is a weighting factor that balances the importance of the KL divergence term relative to the reconstruction loss.

4 References

- Understanding the Variational Lower Bound
- Loss function in Diffusion models