Diffusion probabilistic model in GAN

Qi Wang

December 1, 2021

1 Pre-requisites

Monte Carlo simulatioin and Markov Chain

Variational bound

Annealed Importance Sampling

This uses a Markov chain to slowly convert one distribution to another to compute a ratio of normalization constants.

Langevin dynamics

These are the stochatic realization of the Fokker-Planck equation, which show how to define a Gaussian diffusion process who has many target distributions at its equilibrium.

Kolmogorov equations

It shows that forward and backward diffusion process can be drscribed using the same functional form. The forward Kolmogorov equation corresponds to Fokker-Planck equation, while the backward one describes the time-reversal of this diffusion process, but requires knowing gradients of the density function as a function of time.

2 Related works

There are a number of techniques for training probabilistic model for generative purpose.Reweighted wake-sleep algorithm develops extension and improves learning rules to original wake-sleep algorithm. Generative stochastic networks train a Markov kernel to match its equilibrium distributions to data. Neural autoregressive distribution estimator decomposes a joint distribution into a sequence of tractable conditional distributions. And adversrarial networks train a generative model against a classifier which attempts to distinguish generated samples from real data distributions.

3 Algorithm

The goal is to define a forward(or backward) diffusion process which converts any data distribution into a simple, tractable, distribution, and then learn a finite-time reversal of this diffusion process(which defines the generative model).

Forward process

This is the process where original data get diffused into another distribution(e.g. from image to noise).The original data distribution to be noted as $q(x^0)$, and eventually it's converted to a distribution $\pi(y)$, by repeatedly applying Markov diffusion kernel $T_\pi(y|y';\beta)$, which could be either a Gaussian or a binomial diffusion with β being diffusion rate.

$$\pi(y) = \int dy' T_\pi(y|y';\beta)\pi(y') \tag{1}$$

$$q(x^t|x^{(t-1)}) = T_\pi(x^t|x^(t-1);\beta_t) \tag{2}$$

Forward process starts at data distribution and perform T steps of diffusion:

$$q(x^{(0...T)}) = q(x^0)\prod_{t=1}^{T} q(x^t|x^{(t-1)}) \tag{3}$$

$q(x^t|x^{(t-1)})$ is either Gaussian or binomial diffusion.

Reverse process

This is the process where the model learns to generate.

$$p(x^T) = \pi(x^{T]}) \tag{4}$$

$$p(x^{(0...T)}) = p(x^T)\prod_{t=1}^{T} p(x^{(t-1)}|x^t) \tag{5}$$

Since $q(x^t|x^{(t-1)})$ is Gaussian(binomial) distribution, if β_t is small enough to be neglected, $q(x^{(t-1)}|x^t)$ will also be Gaussian(binomial) distribution.

During the learning, only the mean and the convariance for Gaussian kernel need be estimated.

Model probability

The probability the genertive model assigns to the data is:

$$sp(x^0) = \int dx^{(1...T)} p(x^{(0...T)}) \tag{6}$$

and by rewriting the Eq. (6) into:

$$\begin{aligned} p(x^{(0)}) &= \int dx^{(1\ldots T)} p(x^{(0\ldots T)}) \frac{q(x^{(1\ldots T)}|x^0)}{q(x^{(1\ldots T)}|x^0)} \\ &= \int dx^{(1\ldots T)} q(x^{(1\ldots T)}) \frac{p(x^{(0\ldots T)})}{q(x^{(1\ldots T)}|x^0)} \\ &= \int dx^{(1\ldots T)} q(x^{(1\ldots T)}|x^{(0)}) * p(x^{(T)}) \prod_{t=1}^{T} \frac{p(x^{(t-1)}|x^{(t)})}{q(x^{(t)}|x^{(t-1)})} \end{aligned} \tag{7}$$