TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2022

Diffusion Models

Modeling Densities on R^d

Consider a model density $p_\Phi(y)$ on $y \in R^d$ (for example sound waves or images).

Ideally we want to be able to compute $p_\Phi(y)$, the denisty for any given y, and to also sample y from $p_\Phi(y)$.

Continuous Softmax

We consider the case where a density is defined by a softmax integral,

$$
\begin{aligned}
p_{\text{score}}(y) &= \operatorname*{softmax}_{y}\; s(y) \\
&= \frac{1}{Z} e^{\text{score}(y)} \\
Z &= \int dy\; e^{\text{score}(y)}
\end{aligned}
$$

Here score(y) is a parameterized model computing a score and defining a probability density on R^d.

if y is discrete, but from an exponentially large space (such as sentences) we can use MCMC sampling (the Metropolis algorithm or Gibbs sampling).

In the continuous case the analogue of MCMC sampling is Langevin dynamics.

next

In principe we can sample from $P_{\mathrm{score}}(y)$
Suppose that we want to sample y from this distribution.
In theory this can be done with Langaevin dynamics.

$$\Phi(t+\Delta t) \;\approx\; \Phi(t) - g(\Phi)\Delta t + \epsilon\sqrt{\Delta t}$$
$$\epsilon \;\sim\; \mathcal{N}(0, \eta\sigma^2)$$

We can take this last equation to hold in the limit of arbitrarily small Δt in which case we get a continuous time stochastic process. This process can be written as

$$d\Phi = -g(\Phi)dt + \epsilon\sqrt{dt} \qquad \epsilon \sim \mathcal{N}(0, \eta\sigma^2)$$

The Stationary Distribution with Constant Gradiant Noise

We consider the one dimensional case — a single parameter x — and a probability density $p(x)$.

We will assume the stationary distribution is limited to a region where the gradient noise is effectively constant.

The gradient flow is equal to $-p(x)g$.

The diffusion flow is $-\frac{1}{2}\,\eta\sigma^2\,dp(x)/dx$ (see the appendix).

For a stationary distribution the sum of the two flows is zero giving.

$$\frac{1}{2}\eta\sigma^2\frac{dp}{dx} = -p\frac{d\mathcal{L}}{dx}$$

The 1-D Stationary Distribution

$$
\begin{aligned}
\frac{1}{2}\eta^2\sigma^2\frac{dp}{dx} &= -\eta p\frac{d\mathcal{L}}{dx} \\
\frac{dp}{p} &= \frac{-2d\mathcal{L}}{\eta\sigma^2} \\
\ln p &= \frac{-2\mathcal{L}}{\eta\sigma^2} + C \\
p(x) &= \frac{1}{Z}\exp\left(\frac{-2\mathcal{L}(x)}{\eta\sigma^2}\right)
\end{aligned}
$$

We get a Gibbs distribution with η as temperature!

END