Deep Learning - Foundations and Concepts

Chapter 20. Diffusion Models

nonlineark@github

April 19, 2025

Outline

Forward encoder

Suppose we take an image from the training set, which we will denote by x, and blend it with Gaussian noise independently for each pixel to give a noise-corrupted image z_1 defined by:

$$
\begin{aligned}
z_1 &= \sqrt{1-\beta_1}x + \sqrt{\beta_1}\epsilon_1 \qquad \epsilon_1 \sim \mathcal{N}(\epsilon_1; 0, I) \\
q(z_1|x) &= \mathcal{N}(z_1; \sqrt{1-\beta_1}x, \beta_1 I)
\end{aligned}
$$

where $\beta_1 < 1$ is the variance of the noise distribution.

Forward encoder

We then repeat the process with additional independent Gaussian noise steps to give a sequence of increasingly noisy images $z_1, \dots, z_T$:

$$
\begin{aligned}
z_t &= \sqrt{1-\beta_t} z_{t-1} + \sqrt{\beta_t} \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(\epsilon_t; 0, I) \\
q(z_t | z_{t-1}) &= \mathcal{N}(z_t; \sqrt{1-\beta_t} z_{t-1}, \beta_t I)
\end{aligned}
$$

The values of the variance parameters $\beta_t \in (0, 1)$ are set by hand and are typically chosen such that the variance values increase through the chain according to a prescribed schedule such that $\beta_1 < \cdots < \beta_T$.

Diffusion kernel

Using induction, it's straightforward to verify that:

$$z_t = \sqrt{\alpha_t} x + \sqrt{1 - \alpha_t} \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(\epsilon_t; 0, I)$$
$$q(z_t | x) = \mathcal{N}(z_t; \sqrt{\alpha_t} x, (1 - \alpha_t) I)$$

where we have defined:

$$\alpha_t = \prod_{\tau=1}^{t} (1 - \beta_\tau)$$

We call $q(z_t|x)$ the diffusion kernel. After many steps the image becomes indistinguishable from Gaussian noise, and in the limit $T \to \infty$ we have:

$$q(z_T | x) = \mathcal{N}(z_T; 0, I)$$

Conditional distribution

Our goal is to learn to undo the noise process, and so it is natural to consider the reverse of the conditional distribution $q(z_t|z_{t-1})$:

$$q(z_{t-1}|z_t) = \frac{q(z_t|z_{t-1})q(z_{t-1})}{q(z_t)}$$

But $q(z_{t-1})$ is difficult to calculate:

- Evaluation of the integral $\int q(z_{t-1}|x)p(x)\mathrm{d}x$ is intractable, because we must integrate over the unknown data density $p(x)$.
- If we approximate the integration using samples from the training data set, we obtain a complicated distribution expressed as a mixture of Gaussians.

Conditional distribution

Instead, we consider the conditional version of the reverse distribution, conditioned on the data vector x, defined by $q(z_{t-1}|z_t, x)$:

$$
\begin{aligned}
q(z_{t-1}|z_t, x) &= \frac{q(z_t|z_{t-1}, x)q(z_{t-1}|x)}{q(z_t|x)} = \frac{q(z_t|z_{t-1})q(z_{t-1}|x)}{q(z_t|x)} \\
&= \mathcal{N}(z_{t-1}; m_t(x, z_t), \sigma_t^2 I) \\
m_t(x, z_t) &= \frac{(1-\alpha_{t-1})\sqrt{1-\beta_t}z_t + \sqrt{\alpha_{t-1}}\beta_t x}{1-\alpha_t} \\
\sigma_t^2 &= \frac{\beta_t(1-\alpha_{t-1})}{1-\alpha_t}
\end{aligned}
$$