Project Notes
Bayesian Inference FTN0548

Usama Zafar, Li Ju

March 14, 2024

1 Forward $X_0 \to X_T$

We first define the conditional distributions for X_t on x_{t-1} at time step t:

$$X_t|x_{t-1} \sim \mathcal{N}(X_t; \sqrt{a_t}x_{t-1}, (1-a_t)\mathbf{I})$$

where $0 < a_t < 1$ is a predefined scalar. The intuition behind this is to remain a_t information from x_{t-1} and add $1 - a_t$ noise to the image, to gradually noisify the image.

Thanks to the properties of Gaussian, we have

$$X_T|x_0 \sim \mathcal{N}(X_T; \sqrt{\overline{a}_T}x_0, (1-\overline{a}_T)\mathbf{I})$$

where $\overline{a}_T = \prod_{t=1}^T a_t$.

By carefully designing the sequence of a_t such that $\overline{a}_T \approx 0$, we have $X_T \sim \mathcal{N}(X_T; 0, \mathbf{I})$ approximately.

2 Reverse $X_T \to X_0$

We are interested in the reverse process $p(X_0|X_{1:T})$. However, without knowing $p(X_0)$, $p(X_0|X_{1:T})$ is intractable. But we can approximate it with another parameterised distribution q_θ which can be factorised as

$$q_\theta(X_{0:T}) = q_\theta(X_T)\prod_{t=1}^T q_\theta(X_{t-1}|x_t)$$

2.1 Derivation of Variational Lower Bound

We want to minimise the KL divergence between two joint distributions $\mathrm{KL}\left(p(X_{0:T})\|q_\theta(X_{0:T})\right)$.

$$\begin{aligned}
\mathrm{KL}\left(p(X_{0:T})\|q_\theta(X_{0:T})\right) &= \mathbb{E}_p\left[\ln\frac{p(X_{0:T})}{q_\theta(X_{0:T})}\right] \\
&= \mathbb{E}_p\left[\ln\frac{p(X_{1:T}|x_0)}{q_\theta(X_{1:T}|x_0)} - \ln q_\theta(X_0) + \ln p(X_0)\right] \\
&= \mathbb{E}_{p(X_0)}\left[\underbrace{\mathbb{E}_{p(X_{1:T}|x_0)}\left[\ln\frac{p(X_{1:T}|x_0)}{q_\theta(X_{1:T}|x_0)}\right] - \ln q_\theta(X_0)}_{\ell(x_0,\theta)} + \ln p(X_0)\right]
\end{aligned}$$

The loss expectation w.r.t $p(X_0)$ can be estimated with Monte Carlo methods with the finite samples. Also since $p(X_0)$ is a constant w.r.t θ, we can simply ignore it and focus on term ℓ.

Further we break ℓ into discrete time steps

$$\begin{aligned}
&\mathbb{E}_{p(X_{1:T}|x_0)}\left[\ln \frac{p(X_{1:T}|x_0)}{q_\theta(X_{1:T}|x_0)}\right] - \ln q_\theta(X_0) \\
&= \mathbb{E}_{p(X_{1:T}|x_0)}\left[\ln \frac{p(X_{1:T}|x_0)}{q_\theta(X_{0:T})}\right] \\
&= \mathbb{E}_{p(X_{1:T}|x_0)}\left[\sum_{t=2}^{T} \ln \frac{p(X_t|x_{t-1})}{q_\theta(X_{t-1}|x_t)} + \ln \frac{p(X_1|x_0)}{q_\theta(X_0|x_1)} - \ln q_\theta(X_T)\right] \\
&= \mathbb{E}_{p(X_{1:T}|x_0)}\left[\sum_{t=2}^{T} \ln \frac{p(X_{t-1}|x_t, x_0)p(X_t|x_0)}{p(X_{t-1}|x_0)q_\theta(X_{t-1}|x_t)} + \ln \frac{p(X_1|x_0)}{q_\theta(X_0|x_1)} - \ln q_\theta(X_T)\right] \\
&= \mathbb{E}_{p(X_{1:T}|x_0)}\left[\sum_{t=2}^{T} \ln \frac{p(X_{t-1}|x_t, x_0)}{q_\theta(X_{t-1}|x_t)} + \ln \frac{p(X_T|x_0)}{p(X_1|x_0)} + \ln \frac{p(X_1|x_0)}{q_\theta(X_0|x_1)} - \ln q_\theta(X_T)\right] \\
&= \mathbb{E}_{p(X_{1:T}|x_0)}\left[\sum_{t=2}^{T} \ln \frac{p(X_{t-1}|x_t, x_0)}{q_\theta(X_{t-1}|x_t)} + \ln \frac{p(X_T|x_0)}{q_\theta(X_T)} - \ln q_\theta(X_0|x_1)\right] \\
&= \underbrace{-\mathbb{E}_{p(X_1|x_0)}\left[\ln q_\theta(X_0|x_1)\right]}_{\ell_0} + \underbrace{\sum_{t=2}^{T} \mathbb{E}_{p(X_t|x_0)}\left[\mathrm{KL}(p(X_{t-1}|x_t, x_0) \| q_\theta(X_{t-1}|x_t))\right]}_{\ell_t} + \underbrace{\mathrm{KL}(p(X_T|x_0) \| q_\theta(X_T))}_{\ell_T}
\end{aligned}$$