Diffusion Models (ELBO)

Dr. Alireza Aghamohammadi

Evidence Lower Bound (ELBO)

- Since the exact likelihood $\log P_r(x \mid \phi)$ is generally intractable, we can use a similar approach to that used with Variational Autoencoders (VAEs) by maximizing a lower bound on the log-likelihood, called the Evidence Lower Bound (ELBO).
- For any choice of distribution $q(z)$, the following decomposition holds:

$$\log P_r(x \mid \phi) = \mathsf{ELBO}(x, \phi) + KL\left[q(z) \parallel P_r(z \mid x, \phi)\right],$$

where

$$\mathsf{ELBO}(x, \phi) = \int q(z) \log\left[\frac{P_r(x, z \mid \phi)}{q(z)}\right] \mathrm{d}z.$$

- Because $\log P_r(x \mid \phi) \geq \mathsf{ELBO}(x, \phi)$, we can train the model by maximizing the ELBO as a lower bound on the log-likelihood.
- We are free to choose $q(z)$, and here we define it as the joint distribution of the latent variables conditioned on the observed data x:

$$q(z_1, z_2, \ldots, z_T \mid x) = q(z_1 \mid x) \prod_{t=2}^{T} q(z_t \mid z_{t-1}).$$

- We start by reformulating the log term in the ELBO to derive a form we will optimize:

$$
\begin{aligned}
\log\left[\frac{P_r(x, z_1, z_2, \ldots, z_T \mid \phi)}{q(z_1, z_2, \ldots, z_T \mid x)}\right] &= \log\left[\frac{P_r(z_T) \cdot \prod_{t=2}^{T} P_r(z_{t-1} \mid z_t, \phi) \cdot P_r(x \mid z_1, \phi)}{q(z_1 \mid x) \prod_{t=2}^{T} q(z_t \mid z_{t-1})}\right] \\
&= \log\left[P_r(z_T)\right] + \log\left[\frac{\prod_{t=2}^{T} P_r(z_{t-1} \mid z_t, \phi)}{\prod_{t=2}^{T} q(z_t \mid z_{t-1})}\right] + \log\left[\frac{P_r(x \mid z_1, \phi)}{q(z_1 \mid x)}\right].
\end{aligned}
$$

- We then expand the denominator in the second term using conditional probabilities:

$$
q(z_t \mid z_{t-1}) = q(z_t \mid z_{t-1}, x) = \frac{q(z_{t-1} \mid z_t, x)\, q(z_t \mid x)}{q(z_{t-1} \mid x)}.
$$

❖ Substituting the expanded terms, we get:

$$
\begin{aligned}
&\log\left[\frac{P_r\left(x, z_1, z_2, \ldots, z_T \mid \phi\right)}{q\left(z_1, z_2, \ldots, z_T \mid x\right)}\right] \\
&= \log\left[P_r\left(z_T\right)\right] + \log\left[\frac{\prod_{t=2}^{T} P_r(z_{t-1} \mid z_t, \phi)\ q\left(z_{t-1} \mid x\right)}{\prod_{t=2}^{T} q\left(z_{t-1} \mid z_t, x\right)\ \ q\left(z_t \mid x\right)}\right] + \log\left[\frac{P_r\left(x \mid z_1, \phi\right)}{q\left(z_1 \mid x\right)}\right] \\
&= \log\left[\frac{P_r\left(z_T\right)}{q\left(z_T \mid x\right)}\right] + \log\left[\frac{\prod_{t=2}^{T} P_r(z_{t-1} \mid z_t, \phi)}{\prod_{t=2}^{T} q\left(z_{t-1} \mid z_t, x\right)}\right] + \log\left[P_r\left(x \mid z_1, \phi\right)\right] \\
&\approx \log\left[P_r\left(x \mid z_1, \phi\right)\right] + \sum_{t=2}^{T} \log\left[\frac{P_r\left(z_{t-1} \mid z_t, \phi\right)}{q\left(z_{t-1} \mid z_t, x\right)}\right]
\end{aligned}
$$

The simplified Evidence Lower Bound (ELBO) is given by:

$$\begin{aligned}
\mathsf{ELBO}[x,\phi] &\approx \int \cdots \int q(z_1, z_2, \ldots, z_T \mid x) \left(\log\left[P_r(x \mid z_1, \phi)\right] + \sum_{t=2}^{T} \log\left[\frac{P_r(z_{t-1} \mid z_t, \phi)}{q(z_{t-1} \mid z_t, x)}\right] \right) \mathrm{d}z_1 \ldots \mathrm{d}z_T \\
&= \mathbb{E}_{q(z_1 \mid x)}\left[\log P_r(x \mid z_1, \phi)\right] - \sum_{t=2}^{T} \mathbb{E}_{q(z_t \mid x)}\left[D_{KL}\left(q(z_{t-1} \mid z_t, x) \,\|\, P_r(z_{t-1} \mid z_t, \phi)\right)\right]
\end{aligned}$$

- ❖ The first term maximizes the probability of the observed data sample x.
- ❖ The second term consists of Kullback-Leibler (KL) divergences between pairs of Gaussian distributions, which can be computed in closed form:

$$D_{KL}\left(q(z_{t-1} \mid z_t, x) \,\|\, P_r(z_{t-1} \mid z_t, \phi)\right) = \frac{1}{2\beta_t} \left\| m_t(x, z_t) - f(z_t, \phi, t) \right\|^2 + \mathsf{const}$$

where

$$m_t(x, z_t) = \frac{(1-\alpha_{t-1})\sqrt{1-\beta_t}\, z_t + \sqrt{\alpha_{t-1}}\beta_t\, x}{1-\alpha_t}$$