Reformulating Marginal Log-likelihood

With the assumption of the latent variable z and choice of inference model $q_\phi(z|x)$, now analyzing the log-marginal likelihood. From (1), we have single datapoint marginal likelihood:

$$\begin{aligned} p_\theta(x) &= \int p_\theta(x,z)dz \\ &= \int p_\theta(z|x)p_\theta(x)dz \\ &= \int p_\theta(x)p_\theta(z|x)dz \\ &\approx \int p_\theta(x)q_\phi(z|x)dz \\ &= \mathbf{E}_{q_\phi(z|x)}[p_\theta(x)] \end{aligned}$$

From Jenson's inequality $log(\mathbf{E}[x]) \geq \mathbf{E}[log(x)]$. Hence we have:

$$log(p_\theta(x)) \geq \mathbf{E}_{q_\phi(z|x)}[log(p_\theta(x))] \tag{4}$$

Optimizing VAE: Evidence Lower Bound (ELBO)

Hence we have:

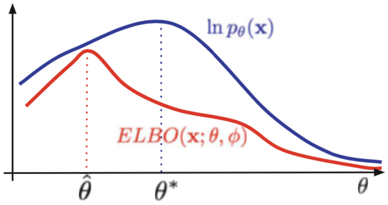
$$
\begin{aligned}
log(p_\theta(x)) &\geq \mathbf{E}_{q_\phi(z|x)}[log(p_\theta(x))] \\
&= \mathbf{E}_{q_\phi(z|x)}\left[log\left[\frac{p_\theta(x,z)}{p_\theta(z|x)}\right]\right] \\
&= \mathbf{E}_{q_\phi(z|x)}\left[log\left[\frac{p_\theta(x,z)}{q_\phi(z|x)}\frac{q_\phi(z|x)}{p_\theta(z|x)}\right]\right] \\
&= \underbrace{\mathbf{E}_{q_\phi(z|x)}\left[log\left[\frac{p_\theta(x,z)}{q_\phi(z|x)}\right]\right]}_{\substack{=\mathcal{L}_{\theta,\phi}(x) \\ ELBO}} + \underbrace{\mathbf{E}_{q_\phi(z|x)}\left[log\left[\frac{q_\phi(z|x)}{p_\theta(z|x)}\right]\right]}_{=D_{KL}(q_\phi(z|x)||p_\theta(z|x)\ \geq\ 0}
\end{aligned}
$$

Hence maximizing ELBO will approximately maximize log-marginal likelihood of data

Term II: $D_{KL}(q_\phi(z|x)||p_\theta(z|x))$

Here the second term (on previous slide) is the KL divergence between $q_\phi(z|x)$ and $p_\theta(z|x)$ which is **intractable**. It quantifies 2 distances:

- By definition, discrepancy between the approximate posterior $q_\phi(z|x)$ and the true posterior $p_\theta(z|x)$.
- The gap between $\mathcal{L}_{\theta,\phi}(x)$ (ELBO) and the marginal log-likelihood $log p_\theta(x)$; this is also called the tightness of the bound. The better $q_\phi(z|x)$ approximates the true (posterior) distribution $p_\theta(z|x)$, in terms of the KL divergence, the smaller the gap



Term I: ELBO

ELBO or Evidence Lower Bound is given by the equation:

$$\mathcal{L}_{\theta,\phi}(x) = \mathbf{E}_{q_\phi(z|x)}\left[log\left[\frac{p_\theta(x,z)}{q_\phi(z|x)}\right]\right] = \mathbf{E}_{q_\phi(z|x)}[log(p_\theta(x,z)) - log(q_\phi(z|x))] \tag{5}$$

eq. (5) is also referred to as **variational lower bound** as it lower bounds the marginal log-likelood of x.

Optimization problem for VAE: ELBO

Coming back to the marginal likelihood:

$$log(p_\theta(x)) \geq \mathcal{L}_{\theta,\phi}(x) + D_{KL}(q_\phi(z|x)||p_\theta(z|x)) \tag{6}$$

we have:

$$\mathcal{L}_{\theta,\phi}(x) \leq log(p_\theta(x)) - D_{KL}(q_\phi(z|x)||p_\theta(z|x)) \tag{7}$$

Hence by targeting the optimization problem:

$$\boxed{\max_{\phi,\theta} \mathcal{L}_{\theta,\phi}(x)} \tag{8}$$

we:

- Approximately maximize the marginal likelihood $p_\theta(x)$. This means that our generative model will become better.
- Minimize the KL divergence of the approximation $q_\phi(z|x)$ from the true posterior $p_\theta(z|x)$, so $q_\phi(z|x)$ becomes better.

Variational Autoencoders (VAE) continued 03/07/2023

- probablistic PCA
  - $P(z) = N(0, I)$
  - $P_\theta(x|z) = N(W^2 + \mu, \sigma^2 I)$
  - $P_\theta(x) = \int P(x|z)P(z)dz = N(\mu, WW^T + \sigma^2 I)$
- VAE
  - $P(z) = N(0, I)$, assumption of prior
  - $P_\theta(x|z)$ = Neural Network
  - $P_\theta(x) = \int P_\theta(x|z) P(z) dz$ ← intractable
  - Since we cannot compute $P_\theta(x)$, $P_\theta(z|x)$ is also intractable.
  - ∴ let's approximate $P_\theta(z|x)$ with $q_\phi(z|x)$, a neural network

    ↑ update distribution of prior $P(z)$
- reformulating marginal log likelihood
  - $P_\theta(x) = \int P_\theta(x,z)dz$

    $= \int P_\theta(z|x)p_\theta(x)dz$

    $= \int p_\theta(x)p_\theta(z|x)dz$

    $\approx \int P_\theta(x)q_\phi(z|x)dz$

    $= E_{q_\phi(z|x)}[p_\theta(x)]$

    (E: expected value; $p_\theta(x)$: random variable)
  - expected value
    - $E[x] = \int x\, p(x)dx$ (x → $P_\theta(x)$, p(x) → $q_\phi(z|x)$)
  - Jenson's inequality $\log(E[x]) \geq E[\log(x)]$

    ∴ $\log(p_\theta(x)) \geq E_{q_\phi(z|x)}[\log(P_\theta(x))]$

    ↖ log likelihood!

- $\log(p_\theta(x)) \geq E_{q_\phi(z|x)}\left[\log\left(\frac{p_\theta(x|z)}{p_\theta(z|x)}\right)\right]$

$\log(p_\theta(x)) \geq E_{q_\phi(z|x)}\left[\log\left(\frac{p_\theta(x,z)}{q_\phi(z|x)} \frac{q_\phi(z|x)}{p_\theta(z|x)}\right)\right]$

$$\log(p_\theta(x)) \geq \underbrace{E_{q_\phi(z|x)}\left(\log\left(\frac{p_\theta(x,z)}{q_\phi(z|x)}\right)\right)}_{\mathcal{L}_{\theta,\phi}(x)} + \underbrace{E_{q_\phi(z|x)}\left(\log\left(\frac{q_\phi(z|x)}{p_\theta(z|x)}\right)\right)}_{D_{KL}(q_\phi(z|x) \| p_\theta(z|x)) \geq 0}$$

evidence lower bound → ELBO

- the second term determines how different the ELBO and marginal log-likelihood $\log(p_\theta(x))$
- ELBO (variational lower bound)
  - $E_{q_\phi(z|x)}(\log(p_\theta(x,z)) - \log(q_\phi(z|x)))$
- max ELBO
  - $\max_{\theta,\phi} \mathcal{L}_{\theta,\phi}(x)$

$$\mathcal{L}_{\theta,\phi}(x) = \log(p_\theta(x)) - D_{KL}(q_\phi(z|x) \| p_\theta(z|x))$$