KL Divergence

Show that

$$KL(p(x)\|q(x)) = \sum_x p(x)\log\frac{p(x)}{q(x)}$$

is always non-negative.

Solution

To show this property of KL divergence, we can start by manipulating this expression:

$$KL(p(x)\|q(x)) = \sum_x p(x)\log\frac{p(x)}{q(x)}$$

This is essentially a expectation term, which can be written as:

$$KL(p(x)\|q(x)) = \mathbb{E}_{p(x)}[\log\frac{p(x)}{q(x)}]$$

We can apply Jensen's inequality for a convex function:

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$$

For a concave function $f(x) = \log(x)$, Jensen's inequality would be flipped:

$$\mathbb{E}[f(X)] \leq f(\mathbb{E}[X])$$

Applying to the original function, we have:
Thus, combining the terms:

$$\mathbb{E}_{p(x)}[\log\frac{p(x)}{q(x)}] \leq \log\mathbb{E}_{p(x)}[\frac{p(x)}{q(x)}]$$

Notice that we can rewrite this in:

$$\log\mathbb{E}_{p(x)}[\frac{p(x)}{q(x)}] = -\log\mathbb{E}_{p(x)}[\frac{q(x)}{p(x)}]$$

We can rewrite this in summation form again:

$$\sum_x p(x)\log\frac{p(x)}{q(x)} \leq -\log\sum_x p(x)\frac{q(x)}{p(x)}$$

$$\sum_x p(x)\log\frac{p(x)}{q(x)} \geq \log\sum_x p(x)\frac{q(x)}{p(x)} = \log\sum_x q(x) = \log(1) = 0$$

Thus we can deduce that:

$$KL(p(x)\|q(x)) \geq 0$$

ELBO + KL Proof

Given a dataset $D = \{x\}$, for each individual data point x, the ELBO on the marginal likelihood of x is given by

$$p(x|\theta) \geq \mathcal{L}(q, \theta; x)$$

where

$$\mathcal{L}(q, \theta; x) = \mathbb{E}_{q(z|x)}\left[\log \frac{p(x, z|\theta)}{q(z|x)}\right].$$

Show that

$$\log p(x|\theta) = \mathbb{E}_{q(z|x)}\left[\log \frac{p(x, z|\theta)}{q(z|x)}\right] + KL\big(q(z|x) \| p(z|x, \theta)\big).$$

Solution

We know that the log marginal likelihood of the data x given parameter θ is the following. Based on our assumption of a hidden z distribution that maps $x \to z$ we can frame it as a expectation under the hidden $q(z)$ distribution:

$$\log p(x|\theta) = \mathbb{E}_{q(z)}[\log p(x|\theta)]$$

We can use the Bayesian rule to expand $p(x|\theta)$ using the joint distribution $p(x, z|\theta)$ and the posterior $p(z|x)$:

$$\log p(x|\theta) = \mathbb{E}_{q(z)}\left[\log \frac{p(x,z|\theta)}{p(z|x,\theta)}\right]$$

Introduce the variational distribution $q(z)$ by multiplying and dividing the same $q(z)$:

$$\log p(x|\theta) = \mathbb{E}_{q(z)}\left[\log\left(\frac{p(x,z|\theta)}{q(z)} \cdot \frac{q(z)}{p(z|x,\theta)}\right)\right]$$

Simplify the logarithm into the expression:

$$\log p(x|\theta) = \mathbb{E}_{q(z)}\left[\log \frac{p(x,z|\theta)}{q(z)}\right] + \mathbb{E}_{q(z)}\left[\log \frac{q(z)}{p(z|x,\theta)}\right]$$

The first term is the Evidence Lower Bound (ELBO) as defined in the question above:

$$\text{ELBO} = \mathbb{E}_{q(z)}\left[\log \frac{p(x,z|\theta)}{q(z)}\right] = \mathcal{L}(q, \theta; x)$$

The second term is the KL divergence (from definition of KL divergence) between $q(z|x)$ and the true posterior $p(z|x)$:

$$\mathbb{E}_{q(z)}\left[\log \frac{q(z)}{p(z|x)}\right] = KL(q(z)\|p(z|x,\theta))$$

Combine the terms:

$$\log p(x|\theta) = \text{ELBO} + KL(q(z)\|p(z|x,\theta))$$

Rearranging, the ELBO is:

$$\text{ELBO} = \mathbb{E}_{q(z)}\left[\log \frac{p(x,z|\theta)}{q(z)}\right] = \log p(x|\theta) - KL(q(z)\|p(z|x,\theta))$$

Thus, we can conclude that:

$$\log p(x|\theta) = \mathbb{E}_{q(z)}\left[\log \frac{p(x,z|\theta)}{q(z)}\right] + KL\big(q(z)\|p(z|x,\theta)\big)$$