Understanding the Evidence Lower Bound

Yurong Chen[*]
Hunan University
chenyurong1998@outlook.com

1 Introduction

The evidence lower bound (ELBO) is an important quantity that lies at the core of a number of algorithms in probabilistic inference, such as variational inference [1]. The variational inference can be formulated as computing the posterior distribution $p(Z|X;\theta)$, given some fixed value for θ.

The definition of ELBO:

$$\text{evidence} := \log p(x;\theta).$$

If we have chosen the right model p and its parameters θ, then we would expect that the marginal probability of our observed data x, would be high. That is, this quantity is evidence that we have chosen the right model for data.

If we happen to also known (or posit) that Z follows some distribution denoted by $q(Z)$ (and that $p(x,z;\theta) := p(x|z;\theta)q(z)$), that the evidence lower bound is:

$$\begin{aligned}
\log p(x;\theta) &= \log \int p(x,z;\theta)dz \\
&= \log \int p(x,z;\theta)\frac{q(z)}{q(z)}dz \\
&= \log \mathbb{E}_{q(Z)}[\frac{p(x,Z)}{q(Z)}] \\
&\geq \mathbb{E}_{q(Z)}[\log \frac{p(x,Z;\theta)}{q(Z)}].
\end{aligned}$$

The final inequality follow from Jensen's inequality. Thus, the ELBO is simply the right-hand side of the above equation:

$$\text{ELBO} := \mathbb{E}_{q(Z)}[\log \frac{p(x,Z;\theta)}{q(Z)}].$$

The gap between the evidence and the ELBO turns to the Kullback Leibler divergence between $p(z|x;\theta)$ and $q(z)$:

$$\begin{aligned}
\text{evidence} - \text{ELBO} &= \log p(x;\theta) - \mathbb{E}_{q(Z)}[\log p(x,Z;\theta)] \\
&= \mathbb{E}_{q(Z)}[\log p(x;\theta)] - \mathbb{E}_{q(Z)}[\log \frac{p(x,Z;\theta)}{p(x;\theta)}] \\
&= \mathbb{E}_{q(Z)}[\log[\frac{q(Z)}{p(Z|x,\theta)}] \\
&:= KL(q(z)||p(z|x;\theta)).
\end{aligned}$$

References

[1] https://mbernste.github.io/posts/elbo/

[*] https://yurongchen1998.github.io