EM and ELBO

Shouchen zhou
zhoushch@shanghaitech.edu.cn

May 23, 2025

If you have any questions, feel free to contact me(above email).

The EM algorithm maximizes a lower bound of the marginal likelihood $\Pr(\mathcal{D};\theta)$.

It is correct that there could have a "log" before the term "marginal likelihood $\Pr(\mathcal{D};\theta)$".

But it is incorrect that there should be a "expected" before the term "marginal likelihood $\Pr(D;\theta)$".

Notice: Please classify the difference between a random variable X, and its value x. It is customary to use uppercase letters to represent random variables and lowercase letters to represent values.

$$
\begin{aligned}
\log p(X) &= \log \frac{p(X,Z)}{\frac{p(X,Z)}{p(X)}} \\
&= \log \frac{p(X,Z)}{q(Z)} \frac{q(Z)}{p(Z \mid X)} \\
&= \underbrace{\int q(z) \mathrm{d}z}_{\text{this equals to 1}} \left[\log \frac{p(X,Z)}{q(Z)} + \log \frac{q(Z)}{p(Z \mid X)}\right] \\
&= \int q(z) \log \frac{p(X,z)}{q(z)} \mathrm{d}z + \int q(z) \log \frac{q(z)}{p(z \mid X)} \mathrm{d}z \\
&= \underbrace{\int q(z) \log \frac{p(X,z)}{q(z)} \mathrm{d}z}_{\text{Evidence Lower Bound(ELBO)}} + \mathrm{KL}(q(Z) \| p(Z \mid X))
\end{aligned}
$$

In the last line, since KL divergence is non-negative, so

$$\int q(z) \log \frac{p(X,z)}{q(z)} \mathrm{d}z$$

is a lower bound of $\log p(X), p(X)$ is called evidence and

$$\int q(z) \log \frac{p(X,z)}{q(z)} \mathrm{d}z$$

is call evidence lower bound(ELBO).

We can further write the ELBO as:

$$
\begin{aligned}
\int q(z) \log \frac{p(X,z)}{q(z)} \mathrm{d}z &= \int q(z) \log p(X,z) \mathrm{d}z - \int q(z) \log q(z) \mathrm{d}z \\
&= \int q(z) \log p(X,z) \mathrm{d}z + \underbrace{H(q(Z))}_{\text{entropy of } q(Z)} \\
&= \mathbb{E}_{q(Z)}[\log p(X,Z)] + H(q(Z))
\end{aligned}
$$

And in the M step of EM, we approximately maximize $\mathbb{E}_{q(Z)}[\log p(X,Z)]$ as maximizing the ELBO, which is the lower bound of the log marginal likelihood $\log p(X)$.