EM and ELBO

Jiachun Jin
jinjch@shanghaitech.edu.cn

> ### The claim
>
> The EM algorithm maximizes a lower bound of the marginal likelihood P(D; θ)

> ### My opinion
>
> If the student claims that there should be a "log" before the term "marginal likelihood P(D; θ)", then this claim is correct.
>
> But the claim that there should be a "expected" before the term "marginal likelihood P(D; θ)" is wrong

$$
\begin{aligned}
\log p(X) &= \log \frac{p(X,Z)}{p(X,Z)/p(X)} && (1)\\
&= \log \frac{p(X,Z)}{q(Z)} \frac{q(Z)}{p(Z|X)} && (2)\\
&= \underbrace{\int q(Z) dZ}_{\text{this equals to 1}} \left[\log \frac{p(X,Z)}{q(Z)} + \log \frac{q(Z)}{p(Z|X)}\right] && (3)\\
&= \int q(Z) \log \frac{p(X,Z)}{q(Z)} dZ + \int q(Z) \log \frac{q(Z)}{p(Z|X)} dZ && (4)\\
&= \underbrace{\int q(Z) \log \frac{p(X,Z)}{q(Z)} dZ}_{\text{Evidence Lower Bound(ELBO)}} + KL(q(Z)||p(Z|X)) && (5)
\end{aligned}
$$

In (5), since KL divergence is non-negative, so $\int q(Z) \log \frac{p(X,Z)}{q(Z)} dZ$ is a lower bound of $\log p(X)$, $p(X)$ is called evidence and $\int q(Z) \log \frac{p(X,Z)}{q(Z)} dZ$ is call evidence lower bound(ELBO).

We can further write the ELBO as:

$$
\begin{aligned}
\int q(Z) \log \frac{p(X,Z)}{q(Z)} dZ &= \int q(Z) \log p(X,Z) - \int q(Z) \log q(Z) dZ && (6) \\
&= \int q(Z) \log p(X,Z) dZ + \underbrace{H(q(Z))}_{\text{entropy of q(Z)}} && (7) \\
&= \mathbb{E}_{q(Z)} \left[\log p(X,Z) \right] + H(q(Z)) && (8)
\end{aligned}
$$

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