Variational Inference (VI)

1 VARIATIONAL INFERENCE (VI)

1.1 Mean Field Theory Based VI

Have observed data X, latent variable Z, and complete data (X, Z).

$$\begin{aligned} \log p(X) &= \log p(X,Z) - \log p(Z|X) \\ &= \log \frac{p(X,Z)}{q(Z)} - \log \frac{p(Z|X)}{q(Z)} \\ \text{left: } & \int_Z q(Z) \log p(X) dZ = \log p(X) \\ \text{right: } & \int_Z q(Z) \log \frac{p(X,Z)}{q(Z)} dZ - \int_Z q(Z) \log \frac{p(Z|X)}{q(Z)} dZ \\ \therefore \quad \log p(X) &= \text{ELBO} + \underbrace{KL(q||p)}_{\geq 0} \end{aligned} \tag{1}$$

In reality, the posterior $P(Z|X)$ is impossible to compute, therefore the VI probabilistic model specifies the joint distribution $P(X, Z)$, and the goad of VI algorithm is to find an approximation for the posterior distribution $p(Z|X)$ as well as for the model evidence $P(X)$.

As shwon in Eq. 1, $\log p(X)$ is the function of X, which is not related to $q(Z)$. Therefore, when the X is given, $\log p(X)$ is a constant. While, our purpose to find a $q(Z)$ to approximate the posterior $p(Z|X)$, which leads the KL divergence to 0. Therefore, the question is equivalent to find the maximal value of ELBO. We can maximize the lower bound (ELBO) by optimization with respect to the distribution $q(z)$, which is equivalent to minimizing the KL divergence.

Therefore, according the Mean Field Theory,

$$q(Z) = \prod_{i=1}^{M} q_i(Z_i) \tag{2}$$

Therefore,

$$
\begin{aligned}
\text{ELBO} &= \int_Z q(Z)\log p(X,Z)dz - \int_Z q(Z)\log q(Z)dz \\
\int_Z q(Z)\log p(X,Z)dz &= \int_Z \prod_{i=1}^{M} q_i(Z_i)\log p(X,Z)dZ \\
&= \int_{Z_j} q_j(Z_j)\left(\int_{Z_i, i\neq j}\prod_{i\neq j}^{M} q_i(Z_i)\log p(X,Z)dZ_1\underbrace{\cdots}_{i\neq j}dZ_M\right)dZ_j \\
&= \int_{Z_j} q_j(Z_j)\left(\mathbb{E}_{\prod_{i\neq j}^{M} q_i(Z_i)}\left(\log p(X,Z)\right)\right)dZ_j \\
\text{Let } & \mathbb{E}_{\prod_{i\neq j}^{M} q_i(Z_i)}\left(\log p(X,Z)\right) = \log p(\hat{X},Z) \\
\therefore \quad &= \int_{Z_j} q_j(Z_j)\log p(\hat{X},Z)dZ_j \\
\int_Z q(Z)\log q(Z)dZ &= \int_Z \prod_{i=1}^{M} q_i(Z_i)\sum_{i=1}^{M}\log q_i(Z_i)dZ \\
&= \int_Z \prod_{i=1}^{M} q_i(Z_i)\left(\log q_1(Z_1)+\log q_2(Z_2)+\cdots+\log q_M(Z_M)\right)dZ \\
&= \int_Z \prod_{i=1}^{M} q_i(Z_i)\log q_1(Z_1)dZ + \cdots + \int_Z \prod_{i=1}^{M} q_i(Z_i)\log q_M(Z_M)dZ \\
\because \quad \int_Z \prod_{i=1}^{M} q_i(Z_i)\log q_1(Z_1)dZ &= \int_{Z_1,Z_2,\cdots,Z_M} q_1(Z_1)\log q_1(Z_1)\prod_{j=2}^{M} q_j(Z_j)dZ_1dZ_2\cdots dZ_M \\
&= \int_{Z_1} q_i(Z_1)\log q_1(Z_1)dZ_1\int_{Z_2,Z_3,\cdots,Z_M}\prod_{j=2}^{M} q_j(Z_j)dZ_2dZ_3\cdots dZ_M \\
\because \quad \int_{Z_2,Z_3,\cdots,Z_M}\prod_{j=2}^{M} q_j(Z_j)dZ_2dZ_3\cdots dZ_M &= \underbrace{\int_{Z_2} q_2(Z_2)dZ_2}_{=1}\cdots\underbrace{\int_{Z_M} q_M(Z_M)dZ_M}_{=1} \\
\therefore \quad \int_{Z_2,Z_3,\cdots,Z_M}\prod_{j=2}^{M} q_j(Z_j)dZ_2dZ_3\cdots dZ_M &= 1 \\
\therefore \quad \int_Z \prod_{i=1}^{M} q_i(Z_i)\log q_1(Z_1)dZ &= \int_{Z_1} q(Z_1)\log q_1(Z_1)dZ_1 \\
\therefore \quad \int_Z q(Z)\log q(Z)dZ &= \sum_{i=1}^{M}\int_{Z_i} q(Z_i)\log q(Z_i)dZ_i = \int_{Z_j} q_j(Z)j)\log q_j(Z_j)dZ_j + \text{Const}
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
\text{ELBO} &= \int_Z q(Z)\log p(X,Z)dz - \int_Z q(Z)\log q(Z)dz \\
&= \int_{Z_j} q_j(Z_j)\log p(\hat{X},Z)dZ_j - \int_{Z_j} q_j(Z_j)\log q_j(Z_j)dZ_j \\
&= \int_{Z_j} q_j(Z_j)\log\left(\frac{p(\hat{X},Z_j)}{q_j(Z_j)}\right)dZ_j \\
&= -KL(q_j||p(\hat{X},Z_j)), \quad \text{where } p(\hat{X},Z_j) = \mathbb{E}_{\prod_{i\neq j}^{M} q_i(Z_i)}\left(\log p(X,Z)\right)
\end{aligned}
\tag{4}
$$