

ELBO+Jensen 的角度推导

E-Step:

$$Q(\theta, \theta^{(i)}) = E_{Z|Y, \theta^{(i)}}[\log P(Y, Z|\theta)] = \int P(Z|Y, \theta) \log P(Y, Z|\theta^{(i)}) dZ$$

【利用当前时刻所估算参数，计算 Z 的分布，然后求期望】

M-Step

$$\theta^{(i+1)} = \arg \max_{\theta} E_{Z|Y, \theta^{(i)}}[\log P(Y, Z|\theta)]$$

【最大化期望】

【可以证明，迭代参数会让似然函数 $\log P(Y|\theta)$ 越来越大】

推导

从最大化 $\log P(Y|\theta)$ 开始，除分布 $Q(Z)$ 再乘以 $Q(Z)$

$$\log P(Y|\theta) = \log \sum_Z P(Y, Z|\theta) = \log \sum_Z \frac{P(Y, Z|\theta)}{Q(Z)} Q(Z) = \log E_{Q(Z)} \left[\frac{P(Y, Z|\theta)}{Q(Z)} \right]$$

$\log$ 是凸函数 (Jensen 不等式，线性参数的函数大于函数值的线性)

$$\log P(Y|\theta) = \log \sum_Z Q(Z) \frac{P(Y, Z|\theta)}{Q(Z)} \geq \sum_Z Q(Z) \log \frac{P(Y, Z|\theta)}{Q(Z)}$$

得到下界，取等号时下式为常数

$$C \equiv \frac{P(Y, Z|\theta)}{Q(Z)} \Rightarrow Q(Z) = \frac{1}{C} P(Y, Z|\theta)$$

两边对 Z 累加

$$\sum_Z Q(Z) = \sum_Z \frac{1}{C} P(Y, Z|\theta)$$

$$1 = \frac{1}{C} \sum_Z P(Y, Z|\theta)$$

$$C = P(Y|\theta)$$

$$Q(Z) = \frac{P(Y, Z|\theta)}{P(Y|\theta)} = P(Z|Y, \theta)$$

即 $Q(Z)$ 取 Z 的一个后验分布，可以得到 $\log P(Y|\theta)$ 一个下界。

$$\log P(Y|\theta) \geq E_{P(Z|Y, \theta)} \left[\log \frac{P(Y, Z|\theta)}{P(Z|Y, \theta)} \right]$$