

# EM(期望最大) 算法.

$$MLE : P(X|\theta)$$

$$\theta_{MLE} = \arg \max \log P(X|\theta) \quad \text{log-likelihood}$$

MLE 可以直接求得解方程，但若有  
潜变量，解会很难求出。

EM 算法是一种迭代优化策略，每一次迭代都分为 2 步：

① 期望步 E-step (E 步, Expectation)

② 极大步 M-step (M 步, Maximization)

$X$ : observed data

$Z$ : unobserved data (latent variable)

$(X, Z)$ : complete data

$\theta$ : parameter

EM 公式：

$$\theta^{(t+1)} = \arg \max_{\theta} \int \log P(X, Z|\theta) P(Z|X, \theta^{(t)}) dZ$$

$$E\text{-step: } P(Z|X, \theta^{(t)}) \longrightarrow E_{Z|X, \theta^{(t)}} [\log P(X, Z|\theta)]$$

$$M\text{-step: } \theta^{(t+1)} = \arg \max \theta E_{Z|X, \theta^{(t)}} [\log P(X, Z|\theta)]$$

一、公式推导 (ELBO + KL 散度)

$$\log P(X|\theta) = \log P(X, Z|\theta) - \log P(Z|X, \theta)$$

$$= \log \frac{P(X, Z|\theta)}{q(Z)} - \log \frac{P(Z|X, \theta)}{q(Z)} \quad q(Z) \text{ 是关于 } Z \text{ 的分布.}$$

将等式两边看作函数，等式两边同时对  $q(Z)$  求期望。

$$\text{左边} = \int_Z q(Z) \log P(X|\theta) dZ = \log P(X|\theta) \underbrace{\int_Z q(Z) dZ}_{1} = \log P(X|\theta)$$

$$\text{右边} = \underbrace{\int_Z q(Z) \log \frac{P(X, Z|\theta)}{q(Z)} dZ}_{ELBO} - \underbrace{\int_Z q(Z) \log \frac{P(Z|X, \theta)}{q(Z)} dZ}_{KL(q(Z) \| P(Z|X, \theta))}$$

$$21) \log P(X|\theta) = ELBO + KL(q||p)$$

由于KL散度大于等于0，因此， $\log P(X|\theta) \geq ELBO$ ，即ELBO是似然函数 $\log P(X|\theta)$ 的下界。

当且仅当 $q(z) = p(z|X, \theta)$ 时， $KL(q||p) = 0$ ，此时 $\log P(X|\theta) = ELBO$

在每次迭代中取 $q(z) = P(z|X, \theta^t)$ ，就可以保证 $\log P(X|\theta^t)$ 与ELBO相等，也就是：

$$\log P(X|\theta) = \int_z P(z|X, \theta^t) \log \frac{P(X, z|\theta)}{P(z|X, \theta^t)} dz - \int_z P(z|X, \theta^t) \log \frac{P(z|X, \theta)}{P(z|X, \theta^t)} dz$$

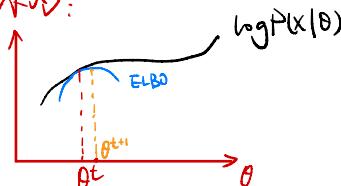
当 $\theta = \theta^t$ 时， $\log P(X|\theta) = ELBO$

ELBO和 $\log P(X|\theta)$ 都是 $\theta$ 的函数，在第七次迭代时模型的参数为 $\theta^t$ 。

此时：

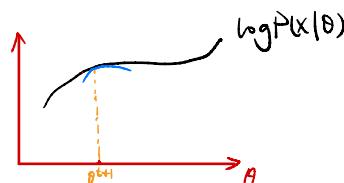
$$ELBO = \int_z P(z|X, \theta^t) \log \frac{P(X, z|\theta)}{P(z|X, \theta^t)} dz.$$

$\log P(X|\theta) = ELBO + KL \geq ELBO$ . 只有当 $\theta = \theta^t$ 时，二者相等  
故函数可表示为：



当 $\theta = \theta^{t+1}$ 时，参数为 $\theta^{t+1}$ 的ELBO取得最大值，如上图。

将参数更新为 $\theta^{t+1}$ ，ELBO图像会随之更新，如图：



此时的ELBO函数只有在 $\theta = \theta^{t+1}$ 时，才会等于 $\log P(X|\theta)$ 。

因为在 $\theta^t$ 时， $\theta^{t+1} = \arg \max_{\theta} ELBO$  且ELBO是 $\log P(X|\theta)$ 的下界，

所以 $\log P(X|\theta^{t+1}) \geq \log P(X|\theta)$ ，依据此规律一直迭代，完成参数估计。

$$\begin{aligned}
\theta^{t+1} &= \arg \max_{\theta} ELBO = \arg \max_{\theta} \int_Z P(Z|X, \theta^t) \log \frac{P(X, Z|\theta)}{P(Z|X, \theta^t)} dZ \\
&= \arg \max_{\theta} \int_Z P(Z|X, \theta^t) \log P(X, Z|\theta) dZ - \underbrace{\arg \max_{\theta} \int_Z P(Z|X, \theta^t) \log P(Z|X, \theta^t) dZ}_{\text{与 } \theta \text{ 无关}} \\
&= \arg \max_{\theta} \int_Z P(Z|X, \theta^t) \log P(X, Z|\theta) dZ \\
&= \arg \max_{\theta} E_{Z|X, \theta^t} [\log P(X, Z|\theta)]
\end{aligned}$$

公式推导 (ELBO + Jensen不等式)

$$\begin{aligned}
\log P(X|\theta) &= \log \int_Z P(X, Z|\theta) dZ \\
&= \log \int_Z \frac{P(X, Z|\theta)}{q(Z)} q(Z) dZ \\
&= \log E_{q(Z)} \left[ \frac{P(X, Z|\theta)}{q(Z)} \right] \\
&\geq \underbrace{E_{q(Z)} \left[ \log \frac{P(X, Z|\theta)}{q(Z)} \right]}_{ELBO} \quad \text{当 } \frac{P(X, Z|\theta)}{q(Z)} = C \text{ 时等号成立}
\end{aligned}$$

$$\begin{aligned}
q(Z) &= \frac{1}{C} P(X, Z|\theta) \\
1 &= \int q(Z) dZ = \int_Z \frac{1}{C} P(X, Z|\theta) dZ \\
1 &= \frac{1}{C} P(X|\theta) \implies C = P(X|\theta)
\end{aligned}$$

$$\therefore q(Z) = \frac{1}{P(X|\theta)} \cdot P(X, Z|\theta) = P(Z|X, \theta)$$

此时 ELBO + kC 效果相同.

三、EM算法的收敛性证明.

证明：当  $\theta^{(t)} \rightarrow \theta^{(t+1)}$  时，有  $\log P(X|\theta^{(t)}) \leq \log P(X|\theta^{(t+1)})$

$$\begin{aligned}
\log P(X|\theta) &= \log P(X, Z|\theta) - \log P(Z|X, \theta) \\
\text{等式两边同时除 } P(Z|X, \theta^{(t)}) \text{ 得得}:
\end{aligned}$$

$$\text{左边} = \int_Z P(Z|X, \theta^{(t)}) \underbrace{\log P(X|\theta)}_{\text{与 } Z \text{ 无关}} dZ = \log P(X|\theta) \underbrace{\int_Z P(Z|X, \theta^{(t)}) dZ}_1 = \log P(X|\theta)$$

$$\text{右边} = \underbrace{\int p(z|x, \theta^{(t)}) \log p(x, z|\theta) dz}_{Q(\theta, \theta^{(t)})} - \underbrace{\int p(z|x, \theta^{(t)}) \log p(z|x, \theta) dz}_{H(\theta, \theta^{(t)})}$$

令：  
 $Q(\theta, \theta^{(t)}) = \arg \max_{\theta} \log p(x, z|\theta) \cdot p(z|x, \theta^{(t)}) dz$ ,

$\Rightarrow Q(\theta^{(t+1)}, \theta^{(t)}) \geq Q(\theta, \theta^{(t)})$ ,  $\therefore Q(\theta^{(t+1)}, \theta^{(t)}) \geq Q(\theta^{(t)}, \theta^{(t)})$

再证  $H(\theta^{(t+1)}, \theta^{(t)}) \leq H(\theta^{(t)}, \theta^{(t)})$  从而得

$$\begin{aligned} & \overline{H(\theta^{(t+1)}, \theta^{(t)}) - H(\theta^{(t)}, \theta^{(t)})} \\ &= \int_z p(z|x, \theta^{(t)}) \cdot \log p(z|x, \theta^{(t+1)}) dz - \int_z p(z|x, \theta^{(t)}) \cdot \log p(z|x, \theta^{(t)}) dz \\ &= \int_z p(z|x, \theta^{(t)}) \underbrace{\log \frac{p(z|x, \theta^{(t+1)})}{p(z|x, \theta^{(t)})}}_{-\text{KL 散度}} dz \leq 0 \end{aligned}$$

故  $H(\theta^{(t+1)}, \theta^{(t)}) - H(\theta^{(t)}, \theta^{(t)}) \leq 0$ .

$$\therefore \log(p(x|\theta^{(t+1)})) \geq \log(p(x|\theta^{(t)}))$$

EM是算法，不是模型，类似于GD.

广义EM

$$\text{目标函数: } \log p(x|\theta) = \text{ELBO} + \text{KL}(q||p)$$

$$\begin{cases} \text{ELBO} = E_{q(z)} \left[ \log \frac{p(x, z|\theta)}{q(z)} \right] \\ \text{KL}(q||p) = \int q(z) \log \frac{q(z)}{p(z|x, \theta)} dz \end{cases}$$

ELBO与 $\theta$ 和 $q$ 有关，可以写为关于 $\theta$ 和 $q$ 的函数，记为 $L(q, \theta)$

在狭义EM中，我们是令 $q(z) = p(z|x, \theta)$ 的，但实际这个后验概率是可能无法求解的。

则当 $\theta$ 固定时， $\log p(x|\theta)$ 固定，我们希望 $\text{KL}(q||p)$ 越小越好，这样ELBO就会更大。

$$\left\{ \begin{array}{l} \hat{q} = \arg \min_q \text{KL}(q||p) = \arg \max_q L(q, \theta) \\ \text{固定 } q, \theta = \arg \max_\theta L(\theta, \theta) \end{array} \right.$$

广义EM：

$$E\text{-step: } q^{(t+1)} = \arg \max_q L(q, \theta^{(t)})$$

$$M\text{-step: } \theta^{(t+1)} = \arg \max_\theta L(q^{(t+1)}, \theta)$$

$$\begin{aligned}
 \text{其中 } ELBO &= L(q, \theta) = E_{q(z)}[\log P(X, z|\theta) - \log Q(z)] \\
 &= E_{q(z)}[\log P(X, z|\theta)] - \underbrace{E_{q(z)}[\log Q(z)]}_{H[q]} \quad q \text{ 的熵}
 \end{aligned}$$

EM 算法类似于坐标上升法 (SMO)，如果在 EM 框架中，无法求解  $z$  的后验概率，则需要采用一些变种的 EM 来估算这个后验：

- ① 基于平均场的变分推断，VBEM / VEM
- ② 基于蒙特卡洛的 EM，MCEM